



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

154

CONTRIBUTIONS

APPLICATIONS OF BAYESIAN COMPUTATIONAL STATISTICS AND MODELING TO LARGE-SCALE GEOSCIENTIFIC PROBLEMS

JOUNI SUSILUOTO



FINNISH METEOROLOGICAL INSTITUTE
CONTRIBUTIONS 154

APPLICATIONS OF BAYESIAN COMPUTATIONAL STATISTICS AND
MODELING TO LARGE-SCALE GEOSCIENTIFIC PROBLEMS

Jouni Susiluoto

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in the hall "Athena" (#107), Siltavuorenpenger 3A, Helsinki, on October 16th 2019 at 12 o'clock noon.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF HELSINKI
HELSINKI, FINLAND

SUPERVISOR

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

PRE-EXAMINERS

Prof. Matti Vihola, University of Jyväskylä, Finland
Prof. Ralph Smith, North Carolina State University, USA

OPPONENT

Prof. Kody Law, University of Manchester, UK

CUSTOS

Prof. Samuli Siltanen, 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 © 2019 Jouni Susiluoto
ISSN 0782-6117
ISBN 978-952-336-080-8 (paperback)
ISBN 978-952-336-081-5 (PDF)
Helsinki 2019
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, 154, FMI-CONT-154
Date 30.9.2019

Author	Title
Jouni Susiluoto	Applications of Bayesian computational statistics and modeling to large-scale geoscientific problems

Abstract

Climate change is one of the most important, pressing, and furthest reaching global challenges that humanity faces in the 21st century. Already affecting daily lives of many directly and everyone indirectly, changes in climate are projected to have many catastrophic consequences. For this reason, researching climate and climate change is needed.

Studying complex geoscientific phenomena such as climate change consists of a patchwork of challenging mathematical, statistical, and computational problems. To solve these problems, local and global process models and statistical models are combined with both small *in situ* observation data sets with only few observations, and equally well with enormous global remote sensing data products containing hundreds of millions of data points. This integration of models and data can be done in a Bayesian inverse modeling setting if the algorithms and computational methods used are chosen and implemented carefully. The methods used in the four publications on which this thesis is based range from high-dimensional Bayesian spatial statistical models and Markov chain Monte Carlo methods to time series modeling and point estimation via optimization.

The particular geoscientific problems considered are: finding the spatio-temporal distribution of atmospheric carbon dioxide based on sparse remote sensing data, quantifying uncertainties in modeling methane emissions from boreal wetlands, analyzing and quantifying the effect of climate change on growing season in the boreal region, and using statistical methods to calibrate a terrestrial ecosystem model.

In addition to analyzing these problems, the research and the results help to understand model performance and how modeling uncertainties in very large computational problems can be approached, also providing algorithm implementations on top of which future efforts may be built.

Publishing unit	Classification (UDC)
Climate System Research	519.676, 551.588.7

Keywords

climate change, computational statistics, Markov chain Monte Carlo, Gaussian processes, Bayesian hierarchical models, carbon cycle, remote sensing, wetlands, methane emissions

ISSN and series title	ISBN
0782-6117	978-952-336-080-8 (paperback)
Finnish Meteorological Institute Contributions	978-952-336-081-5 (pdf)

DOI	Language	Pages
https://doi.org/10.35614/isbn.9789523360815	English	165

SELECTED PUBLICATIONS

- I **J. Susiluoto**, A. Spantini, H. Haario, Y. Marzouk. *Efficient multi-scale Gaussian process regression for massive remote sensing data with satGP v0.1*. Geosci. Model Dev. Discuss. 2019. URL <https://doi.org/10.5194/gmd-2019-156>
J.S. wrote the satGP program, designed most and implemented all of the novel computational ideas, designed and carried out the experiments, analyzed the results, and prepared the manuscript and the figures for publication.
- II **J. Susiluoto**, M. Raivonen, L. Backman, M. Laine, J. Mäkelä, O. Pelto, T. Vesala, and T. Aalto. *Calibrating the sqHIMMELI v1.0 wetland methane emission model with hierarchical modeling and adaptive MCMC*. Geosci. Model Dev., 11, 1199-1228, 2018. doi: 10.5194/gmd-11-1199-2018. URL www.geosci-model-dev.net/11/1199/2018/
J.S. designed the research, implemented the algorithms, carried out the simulations, analyzed the results, and prepared the manuscript and the figures for publication.
- III J. Pulliainen, M. Aurela, T. Laurila, T. Aalto, M. Takala, M. Salminen, M. Kulmala, A. Barr, M. Heimann, A. Lindroth, A. Laaksonen, C. Derksen, A. Mäkelä, T. Markkanen, J. Lemmetyinen, **J. Susiluoto**, S. Dengel, I. Mammarella, J.-P. Tuovinen, and T. Vesala. *Early snowmelt significantly enhances boreal springtime carbon uptake*. Proceedings of the National Academy of Sciences, 114(42):11081–11086, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1707889114. URL <http://www.pnas.org/content/114/42/11081>.
J.S. performed the climate model simulations, including the various auxiliary runs for creating the initial model states, post-processed the model output, and participated in analyzing the data.
- IV J. Mäkelä, **J. Susiluoto**, T. Markkanen, M. Aurela, H. Järvinen, I. Mammarella, S. Hagemann, and T. Aalto. *Constraining ecosystem model with adaptive metropolis algorithm using boreal forest site eddy covariance measurements*. Nonlinear Processes in Geophysics, 23(6):447–465, 2016. doi: 10.5194/npg-23-447-2016. URL <http://www.nonlin-processes-geophys.net/23/447/2016/>.
J.S. developed and tested the sampling framework used for performing the simulations, provided initial modifications for the JSBACH model to efficiently work with that framework, and contributed to preparing the manuscript.

PARTIAL LIST OF SYMBOLS

Symbol	Meaning
\mathbb{N}	$\{0, 1, 2, 3, \dots\}$
\mathbb{N}^+	$\{1, 2, 3, \dots\}$
$[a, b)$	$\{x \in \mathbb{R} : a \leq x < b\}$
\mathbb{R}^+	$[0, \infty)$
\triangleq	Is defined to be
\approx	Is approximately equal to
\propto	Is proportional to
\sim	Is distributed according to
$a \leftarrow b$	a is assigned the value b
$\mathcal{N}(\mu, \Sigma)$	Multivariate normal distribution with mean vector μ and covariance Σ
Λ	Lebesgue measure
$f \circ g$	Composition of functions f and g
$\ x\ _{\Gamma}$	$\sqrt{x^T \Gamma^{-1} x}$
$ K $	Determinant of matrix K
$\Gamma(s)$	Gamma function with argument s
\xrightarrow{d}	Converges in distribution to
$\delta_{x,x'}$	Dirac delta function
$a \wedge b$	$\min(a, b)$
$\mu \ll \nu$	μ is absolutely continuous with respect to ν
$X \perp\!\!\!\perp Y$	X and Y are independent random variables
$1_{[a,b]}$	$1_{[a,b]}(x) = 1$ if $a \leq x \leq b$, otherwise 0.

PARTIAL LIST OF ABBREVIATIONS

Abbreviation	Stands for
ARMA	Autoregressive moving average model (2.56)
DAG	Directed acyclic graph
ECMWF	European Center for Medium-Range Weather Forecasts
GPP	Gross primary production
GSSD	Growing season starting date
IPCC	Intergovernmental Panel on Climate Change
MAP	Maximum <i>a posteriori</i>
MCMC	Markov chain Monte Carlo
MLE	Maximum likelihood estimate
MRF	Markov random field
NLL	Negative log-likelihood
OCO-2	Orbiting Carbon Observatory 2 (NASA satellite)
PDE	Partial differential equation
i.i.d.	Independent and identically distributed
pdf	Probability density function
pmf	Probability mass function
s.t.	Such that

ACKNOWLEDGMENTS

This work would not have been possible without the input of a large number of people: colleagues, family, friends, and others. For that simple reason, the following list is by no means exhaustive.

I would like to thank Profs. Samuli Siltanen from University of Helsinki (UHEL) and Marko Laine from the Finnish Meteorological Institute (FMI) for more and less formally advising me through the long doctoral process. I was lucky to get to work with you.

Thanks are also in order to Prof. Ari Laaksonen and head of the greenhouse gases group Tuomas Laurila, who initially employed me at the FMI greenhouse gases group solely based on a recommendation by Profs. Timo Vesala and Eero Nikinmaa (UHEL), both of who also deserve my gratitude. I sincerely thank Doc. Tuula Aalto, the group leader of the carbon cycle modeling group at the FMI, and Prof. Heikki Järvinen (UHEL), who initially pushed me closer to the applied statistics community at and around FMI.

I would like to thank all my co-authors, and in particular (in alphabetic order) colleague and friend Jarmo Mäkelä, Prof. Jouni Pulliainen, and Dr. Maarit Raivonen. Discussions with Prof. Johanna Tamminen and Dr. Janne Hakkarainen lead me to start looking into remote sensing-related problems, and that has been both a positive challenge and a pleasure.

From the years as a researcher at FMI, I want to thank the companionship and help of Drs. Leif Backman, Yao Gao, Sauli Lindberg (UHEL), Tiina Markkanen, Pirkka Ollinaho, Joni-Pekka Pietikäinen, Tea Thum, Aki Tsuruta, Simo Tukiainen, and others, for all kinds of (non-)peer support. Especially in the very first years, the company of Dr. Antti-Ilari Partanen (FMI) and Dr. Antti Solonen (LUT) was important.

For help with technical problems, which were not few nor far between, I would like to thank my friends Ilmari Karonen, Seppo Varho, and Jussi Virolainen. Veronika Gayler, Christian Reich and Reiner Schnur from Max Planck Institute for Meteorology in Hamburg, and Declan O'Donnell (FMI) helped me with ECHAM/JSBACH modeling issues. Regarding the text, I would like to thank Prof. Lassi Roininen for reading an early version and providing a long list of helpful comments.

The thesis work was improved immensely by the patience and bottomless knowledge of professors Heikki Haario (LUT) and Youssef Marzouk from Massachusetts

Institute of Technology (MIT), who made it possible for me to undertake and complete major parts of this work at MIT. Neither of you were my formal supervisors, but if senior colleagues were to be ranked by supervision hours, both of you would be very high on the list. Equally important was that working with you was a lot of fun.

I tip my imaginary hat to Alessio Spantini (MIT) for all the nice dinners and for all the things I learned from our collaboration, and also to Sebastian Springer (LUT) for the irreplaceable companionship in Boston. The uncertainty quantification group at the AeroAstro department of MIT, and in particular many of its graduate students and post-docs, taught me astonishingly much and it was a privilege to get to experience the extraordinarily supportive atmosphere that you all created. Some of these people are: Ricardo Baptista, Daniele Bigoni, Remy Lam, Alexander Marquez, Andrea Scarinci, Chaitanya Talnikar, Anirban Chaudhuri, Zheng Wang, and Benjamin Zhang.

A few very special mentions of people whose names did not come up yet: thank you to my long time friends Theo Kurten, Juuso Laatikainen, Pirkka-Pekka Laurila, Ari Keränen, Daniel Marszalec, and Esko Oksanen for all the good times.

Markku Koskinen, Matti Lindholm, and Tuukka Susiluoto took me many times to admire and appreciate the nature behind all the modeling, and those experiences always helped to remember why wasting hours in the lab and at the computer matters. Pekka Heinonen helped with various disasters in Finland while I was concentrating on research in Boston and that gave me peace of mind to work – you probably have no idea about how much it meant to me.

My parents Ilkka and Liisa always provided support and practical advice, and without them I would have stumbled many more times in the process. Thank you for always placing us children first in your priorities. My siblings Anne, Elina, Kaisa and Tuukka along with their families taught me irreplaceable life lessons, which emerged useful when navigating the frustrations and the unavoidable setbacks with research.

Last, I want to thank Emilija Rožukaitė for many good things, but most importantly for just being yourself. Without you, much of this work and many other even more exciting adventures in life would never have happened.

Cambridge, MA, May 2019
Jouni Susiluoto

CONTENTS

1	Introduction	1
2	Theory	5
2.1	Probabilistic background and notation	5
2.1.1	Probability and random variables	5
2.1.2	Model calibration via Bayes' theorem	8
2.1.3	Forward models and inverse problems	9
2.1.4	Standard point estimation and cross validation methods	11
2.2	Uncertainty	11
2.2.1	Sources of uncertainty	11
2.2.2	Distributions for uncertainty modeling	12
2.3	Linear regression	14
2.4	Gaussian processes	14
2.4.1	A parametric form for the Gaussian process mean function	16
2.4.2	Gaussian process covariance kernels	16
2.5	Graphical models	19
2.5.1	Directed graphical models	19
2.5.2	Undirected graphical models	20
2.6	Monte Carlo algorithms	21
2.6.1	Markov chain Monte Carlo	23
2.6.2	Gibbs sampling and Metropolis within Gibbs	27
2.6.3	Importance sampling and resampling	28
2.7	Hierarchical Bayesian models	29
2.8	Bayesian modeling with time series data	30
2.8.1	AR, MA, and ARMA models	31
2.8.2	Practical parameter estimation in the ARMA setting	31
3	Applications to geosciences	35
3.1	Overview of scientific contributions	35
3.1.1	Spatio-temporal high resolution CO ₂ distributions	35
3.1.2	Uncertainties in Boreal wetland CH ₄ emission processes	35

3.1.3	Effects of climate change on growing season and gross primary production	36
3.1.4	Monte Carlo estimates of land surface scheme hydrology and gas exchange parameters	37
3.1.5	Other related work	37
3.2	Models, observations, and algorithms control computational cost . .	38
3.2.1	Parallel models and parallel algorithms	38
3.2.2	The role of the observation data	40
3.3	Efficient multi-scale Gaussian processes for massive remote sensing data	42
3.3.1	Gaussian process model algorithm description	43
3.3.2	Obtaining the GP mean function from a Gaussian MRF	44
3.3.3	Identifiability of multi-scale parameters	45
3.3.4	Learning multi-scale kernel parameters from OCO-2 data . . .	47
3.3.5	Posterior XCO ₂ fields	49
3.3.6	Wind-informed kernel	50
3.4	Bayesian inference of physics of a Boreal wetland with hierarchical MCMC	51
3.4.1	The HIMMELI forward model	51
3.4.2	Bayesian Inference	52
3.4.3	Results and discussion	53
3.5	Climate and land surface modeling	56
3.5.1	The ECHAM/JSBACH forward model	57
3.5.2	Paper III – climate change has shifted the growing season . .	57
3.5.3	Paper IV – constraining LSS parameters with flux data with adaptive MCMC	58
4	Conclusions and future work	61
	References	65

1 INTRODUCTION

Climate change is one of the most critical challenges that humankind faces in the twenty-first century. It will potentially cause huge economic, societal and environmental disruptions, which the general public is in many parts of the world slowly starting to realize. In recent years politicians, scientists and news outlets among others have attributed events such as devastating hurricanes, forest fires, giant icebergs splintering away from glaciers, floods, catastrophic losses in biodiversity in pristine rainforests, extreme droughts, crop failures, and so on, to climate change. The research presented in this thesis strives to explain parts of the underlying mechanisms better.

Climate change is caused by heat-trapping gases, most notably carbon dioxide (CO₂) and methane (CH₄), that are released to the atmosphere both naturally and by humans. The radiative forcing potential of atmospheric carbon dioxide compared to the pre-industrial level is currently at 1.68 W/m² whereas that of methane is at 0.97 W/m², according to the latest report by the Intergovernmental Panel on Climate Change (IPCC) (IPCC, 2013). Other gases effect the radiative balance as well, but CO₂ and CH₄ are the most important ones, and by a wide margin.

The major source of CO₂ is the burning of fossil fuels to power factories, cars, power plants, etc. The natural CO₂ sources are dwarfed in comparison to these anthropogenic sources, without which the atmospheric CO₂ concentrations would be stable. For methane, the anthropogenic sources include leaks from oil and natural gas fields, farming, landfills, and coal mining, but wetlands, where peat is anaerobically decomposed by *archaea* (prokaryotic organisms), are also an important component. The inner workings of wetlands differ widely from one to another, depending for instance on temperature, local plant species, soil chemistry, and availability of nutrients.

How carbon circulates in air, land, and water, is complicated and consists of a large number of processes. Much of the carbon dioxide emitted to the atmosphere is dissolved in water, little by little lowering the pH of the oceans. Another part is photosynthesized by plants, adding to the terrestrial carbon pool. The leftover CO₂ stays to increase the atmospheric concentration, which during the last 30 years has risen by almost 20%. The second order mechanisms are complex – for instance changes in terrestrial and marine ecosystems affect their responses to the changing levels of atmospheric carbon dioxide.

An important part of most modern analyses of the carbon cycle is uncertainty

quantification. Uncertainty quantification tries to formally analyze and attribute sources of uncertainty in any estimates to different parts of the estimation process, such as the uncertainty arising from measurement and modeling errors. Evaluating the uncertainties sometimes critically changes outcomes of research, as was shown for instance by an Oxford study from summer 2018: after accounting for uncertainties, it was found to be plausible that there is no alien life in the Milky Way, contrary to the usual opposite conclusion from a non-probabilistic application of the Drake Equation.

For producing actionable climate-related scientific results, sources, sinks, and stocks of carbon need to be estimated, typically with complicated climate models and/or sophisticated statistical techniques. This task is not made easier by the intimate coupling of Earth's water and carbon cycles. In the research presented in this thesis, several such complications that arise from intertwining and interacting processes are looked at. Photosynthesis takes place by the action of plants opening their stomata, which inevitably lets out water vapor. In times of drought, wetland carbon decomposition changes from anaerobic to aerobic, but this behavior is difficult to model due to the nonlinear changes in the microbial populations affecting decomposition of organic matter. Finally, climate change affects the snow clearing date across the Boreal ecosystems, which is reflected in the total gross primary production during the growing season.

This thesis looks into both modeling different aspects of the carbon cycle and evaluating the associated uncertainties. The work utilizes site-level carbon dioxide and methane flux measurement data with time series analysis and Markov chain Monte Carlo (MCMC) (Gelman, 1997) techniques, global climate modeling with large amounts of flux measurement data from all over the world, and remote sensing CO₂ data from the NASA Orbiting Carbon Observatory 2 (OCO-2) satellite (Crisp et al., 2012; O'Dell et al., 2012) with stochastic processes and spatial statistics techniques.

The simplest way to estimate a quantity by modeling is to obtain a prediction by initializing the model according to best expert knowledge and data available and performing a single model simulation. This method is often used when the computer model in question is extremely expensive, as is the case with computational fluid dynamics models, which are used for e.g. climate and weather models and rocket engine or aircraft component performance simulations. An example of such direct simulation is also part of **Paper III**, where the gross primary production response to changing snow clearance date is evaluated and compared against flux measurement data with the ECHAM5/JSBACH/CBALANCE family of models (Roeckner et al., 2003a) from the Max Planck Institute for Meteorology in Hamburg. Due to a single simulation taking several weeks, no uncertainty quantification was possible.

If the model is computationally less demanding, statistical methods utilized for uncertainty quantification can be more sophisticated. **Paper IV** employs an MCMC algorithm to evaluate parameter posteriors of several parameters affecting the carbon and water cycles. Similarly to Paper III, Paper IV uses the JSBACH model, but restricts the spatial domain to measurement sites instead of performing the simu-

lations for a larger region. Pre-computed weather data is used for model forcing, saving remarkably in computation time by refraining from solving the complicated and expensive atmospheric component at each time step.

A wetland methane emission model is utilized in **Paper II** to analyze what parts of the methane production process are constrained by flux measurement data. Since the model is less complex, a more sophisticated modeling approach can be used for modeling uncertainties. An adaptive MCMC algorithm is employed in a Metropolis within Gibbs setting with hierarchical modeling of annually changing environmental parameters with an autoregressive moving average time series model for defining the error model. The results indicate that without further measurement data, it is very challenging to state the importance of the different processes. This is an important result, since there are enormous quantities of peat in the boreal wetlands, which might be eventually released by the thawing of the Siberian permafrost. The full parameter posterior uncertainty of such models has not been extensively evaluated earlier in literature, nor has a hierarchical approach been used.

In contrast to the flux measurement observations used in Papers II-IV, **Paper I** utilizes remote sensing CO₂ measurements from the OCO-2 satellite. Remote sensing of greenhouse gas concentrations has obvious benefits compared to *in situ* measurements in that remote sensing provides almost global coverage and measures similarly everywhere. However, the approach also brings problems: gaps in data due to clouds, unknown biases and errors, and long distances between satellite trajectories. Utilizing remote sensing data for constructing time-dependent high resolution CO₂ distributions is therefore still work in progress, and so far estimates published in the literature show overly smoothed CO₂ fields, not being able to utilize the data to its full potential. The results we present should hence be an important opening: an open source multi-scale Gaussian process software, able to compute the demanding spatial statistics problem with enormous amounts of data (at least hundreds of millions of observations), and retaining the local fine structure. The computation enables calculating the posterior mean and marginal variances, but also drawing samples of random functions and calibrating covariance kernel parameters based on data. We additionally describe several novel ideas for covariance modeling, some of which have not been used either in this or any other context before, such as wind-informed kernels, multi-scale kernels, and periodic kernels. We validate the multi-scale approach with synthetic studies, and show initial applications of the methodology to the OCO-2 v9 data product.

The Papers described above underline the multidisciplinary nature of climate science. This thesis has to deal with all of those disciplines and therefore it contains some more fundamental aspects of mathematics (measure theory, probability, random functions), more applied topics (statistics), computational paradigms (programming, graphical models, inference algorithms), physical modeling (process/forward models), and climate science (analyzing the results). To communicate that full scientific process, most of these technical aspects are presented in Chapters 2 and 3. The

presentation of the mathematical theory in chapter 2 is not always comprehensive, since a full treatment would take up too much space. The topics are well known in literature, however, and the reader is referred to the literature cited below for further details.

For a general introduction to inverse problems, see e.g. Mueller and Siltanen (2012); Tarantola (2005). For general Bayesian and non-Bayesian statistics, see Gelman et al. (2013), Casella and Berger (2002), or Bickel and Doksum (2015, 2016). For the measure-theoretic foundations of probability, (Williams, 1991) is a good starting point. Kalman filter and dynamic linear models are treated in Särkkä (2013) and Durbin and Koopman (2012). For Gaussian processes, good references, and the ones mainly used for this thesis are Rasmussen and Williams (2006) and Santner et al. (2003). For an interesting measure-theoretic exposition of random functions but technically beyond the level required in this work, see e.g. Karatzas and Shreve (1998); Stroock (2018). A solid general treatment of Bayesian inverse problems, emphasizing infinite-dimensional settings, is also given by Stuart (2010), while Gamerman (1997) describes more comprehensively the fundamentals of MCMC.

This thesis is structured in the following way: Chapter 2 will explore theory, starting from a very short review of basic probability, introducing Bayes' theorem and inverse problems. It will then cover uncertainty, linear models, Gaussian random functions, and graphical models, through which the algorithms used in the Papers are explained. This is followed by a presentation of the Monte Carlo techniques used in the Papers, such as MCMC including Gibbs sampling, and importance sampling. The chapter ends with a discussion of hierarchical Bayesian models and autoregressive moving average (ARMA) time series modeling. Chapter 3 discusses the research in the Papers against the theoretical background, concentrating on computational issues and climate science. Chapter 4 contains a short discussion of how the analyses could be further improved, where the current limitations of the presented approaches are, and how some of the most straightforward lines of future work look like.

2 THEORY

Quantifying uncertainty is based on the notion of probability. The uncertainty of predicted CO₂ concentration in June 2050 in Helsinki can be given as a *credible interval*: with a given probability the concentration is between $x - \delta$ and $x + \delta$ ppm. Another way of describing the uncertainty is describing the *distribution* of possible values, for instance by stating that the predicted concentration is a random variable distributed according to, say, normal distribution with mean x and variance $\frac{\delta^2}{4}$.

Uncertainty quantification in geosciences is important, since it affects how to best evaluate risk. This includes for instance how to minimize expected (arbitrary) loss due to climate change, deforestation, particulate emissions, wildfires and natural disasters, among others.

Uncertainties are in this work predominantly quantified using the Bayesian paradigm and the essential theory for doing so is presented in this chapter. The readers who are intimately familiar with Bayesian models, time series, random functions, and associated algorithms, may choose to skip this review and only use it as reference when necessary. Likewise, the reader with very little mathematical background may just want to skip the chapter, since it is rather condensed and not suitable for self-study – for that purpose the references at the end of chapter 1 can serve as starting points. For the reader who is familiar with the problems described in especially Papers I-II, this chapter may provide valuable information about how to in practice go about solving the associated modeling and computational problems.

2.1 PROBABILISTIC BACKGROUND AND NOTATION

2.1.1 PROBABILITY AND RANDOM VARIABLES

Let Ω be the set of possible outcomes of an experiment. A σ -algebra of Ω , \mathcal{F} , is a set of subsets of Ω such that complements and countable unions of any $F \in \mathcal{F}$ are also members of \mathcal{F} , as is the full space Ω . A *probability space* is a triplet $(\Omega, \mathcal{F}, \mu)$, where μ is a probability measure, $\mu : \mathcal{F} \rightarrow [0, 1]$ s.t. $\mu(\emptyset) = 0$ and $\mu(\Omega) = 1$ (Bickel and Doksum, 2015). The sets $F \in \mathcal{F}$ are then called μ -*measurable*. Given spaces (Ω, \mathcal{F}) and (Ω', \mathcal{F}') , a mapping $h : \Omega \rightarrow \Omega'$ is a *measurable function* if $\forall F' \in \mathcal{F}'$ it is true that $h^{-1}(F') \in \mathcal{F}$ (Williams, 1991).

Measure μ is *absolutely continuous* with respect to ν , written $\mu \ll \nu$, if $\forall F \subset \mathcal{F}$,

it holds that $\nu(F) = 0 \Rightarrow \mu(F) = 0$. The measure ν is σ -finite if $\Omega = \cup_{i=1}^{\infty} F_i$ with $F_i \in \mathcal{F}$ and $\forall i, \nu(F_i) < \infty$. The *Radon-Nikodym Theorem* (Williams, 1991) states that given $\mu \ll \nu$ with ν σ -finite, there exists a function $g : \Omega \rightarrow [0, \infty]$ such that

$$\mu(F) = \int_F g(x) d\nu(x). \quad (2.1)$$

The function g is called the *Radon-Nikodym derivative of μ with respect to ν* . Given in the above setting a second measurable space (Ω', \mathcal{F}') and a measurable function $h : \Omega \rightarrow \Omega'$, the *pushforward measure of μ* is denoted by $h_*(\mu) : \Omega' \rightarrow [0, \infty)$ with

$$h_*(\mu)(F') \triangleq \mu(h^{-1}(F')), \quad (2.2)$$

where $F' \in \mathcal{F}'$ (Peyré and Cuturi, 2018).

A *random variable X* is a measurable function from a probability space to a measurable space, $X : (\Omega, \mathcal{F}, \mu) \rightarrow (S, \mathcal{S})$, where \mathcal{S} is a σ -algebra on the nonempty set S (e.g. Williams (1991)). In this work the random variables are generally real-valued, $S = \mathbb{R}$. The set Ω is called the *sample space* (Casella and Berger, 2002).

Given a real-valued random variable X as above, the *law of X* , \mathcal{L}_X , is defined as $\mathcal{L}_X \triangleq X^{-1} \circ \mu$. This can be thought of as the pushforward $X_*(\mu)(U)$ for sets $U \in \mathcal{B}(\mathbb{R})$ via (2.2), where $\mathcal{B}(\mathbb{R})$ denotes the standard Borel σ -algebra over \mathbb{R} . The (cumulative) *distribution function of X* (cdf) is then defined (Williams, 1991) by

$$F_X(a) \triangleq \mathcal{L}_X((-\infty, a]) = \mu(\{\omega \in \Omega : X(\omega) < a\}). \quad (2.3)$$

For all practical purposes, finite-dimensional random variables are often associated with probability density functions (pdf), and discrete with probability mass functions (pmf). The pdf of a real-valued random variable X , $f_X(x)$, if it exists, is defined via $\int_a^b f_X(x) d\Lambda(x) = \Pr(a \leq X \leq b)$, where Λ denotes the standard Lebesgue measure. If $\mathcal{L}_X \ll \Lambda$, the pdf can also be described (Williams, 1991) as the Radon-Nikodym derivative of the law of X with respect to the Lebesgue measure,

$$f_X = \frac{d\mathcal{L}_X}{d\Lambda}. \quad (2.4)$$

A real-valued *random vector* is a random variable, which maps the sample space onto \mathbb{R}^q for some $q \in \mathbb{N}$ (Casella and Berger, 2002). The definitions of pmf, pdf, and cdf generalize trivially. Functions of random variables are random variables.

For random variables X and Y with sample spaces Ω_X and Ω_Y the *joint density* is the pdf/pmf of the random vector (X, Y) , and is denoted by $f_{X,Y}(x, y)$, and it can be *marginalized* over either of the arguments by integration, e.g. $f_X(x) = \int_{\Omega_Y} f(x, y) d\Lambda(y)$ (Williams, 1991). In the case where the sample space is \mathbb{R} this becomes $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$ (Casella and Berger, 2002). Conditioning the random variable X on Y , denoted $X|Y$ defines a new random variable whose density function is called the *conditional density, of X given Y* and it is denoted and defined by

$f_{X|Y}(x) = f_{X,Y}(x, y)/f_Y(y)$. The elementary *chain rule* is the definition of conditional probability written in the form $f_{X,Y}(x, y) = f_{X|Y}(x)f_Y(y)$.

The *expectation* of a function g of a random variable or vector $X : (\Omega, \mathcal{F}, \mu) \rightarrow \mathbb{R}^q$ is given by $\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega))d\mu$ (Casella and Berger, 2002), which, given a density function $f_X(x)$ for X , boils down to $\mathbb{E}[g(X)] = \int_{\Omega} g(x)f_X(x)d\Lambda(x)$. With that the *covariance* of a random variables X and Y is given by $\text{Cov}(X, Y) = \mathbb{E}[X - \mathbb{E}[X]]\mathbb{E}[Y - \mathbb{E}[Y]]$, with *variance* of X defined as $\text{Cov}(X, X)$ and written as $\mathbb{V}[X]$. The covariance matrix C of a random vector X has elements $C_{ij} = \text{Cov}(X_i, X_j)$ (Casella and Berger, 2002). The *correlation* between random variables X and Y is defined as $\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}}$.

For finite samples $x^1 \dots x^N$ from any distribution, the unbiased estimates of the mean and covariance are given by $\bar{x} = \frac{1}{N} \sum_{i=1}^N x^i$ and $S = \frac{1}{N-1} \sum_{i=1}^N (x^i - \bar{x})(x^i - \bar{x})^T$ respectively (Casella and Berger, 2002). In this thesis sample sizes are generally very large, and therefore sample covariances are also often denoted by letter C . The elements C_{kl} describe the covariance between the k^{th} and l^{th} dimension of vectors x . With pairs of vector data $(x^1 \dots x^n, y^1 \dots y^n)$, correlation refers to a matrix with the *Pearson correlation coefficients* as elements, as in

$$\text{Corr}(x^1 \dots x^n, y^1 \dots y^n)_{kl} = \frac{1}{N-1} \sum_{i=1}^N \frac{(x_k^i - \bar{x}_k)(y_l^i - \bar{y}_l)^T}{\sqrt{\mathbb{V}[X_k]\mathbb{V}[X_l]}}. \quad (2.5)$$

Two sub- σ -algebras $\mathcal{F}_1, \mathcal{F}_2$ of \mathcal{F} – that is, they are σ -algebras and $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ – are independent if $\Pr(x \in F_1 \cap F_2) = \Pr(x \in F_1)\Pr(x \in F_2)$ for all $F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2$. Two random variables *independent*, if their σ -algebras are independent, written $X \perp\!\!\!\perp Y$ (Williams, 1991). In practice for distributions with well-behaving density functions this translates to that two random variables X and Y are independent if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ (Casella and Berger, 2002). In addition, they are *conditionally independent* given a third random variable Z , written $X \perp\!\!\!\perp Y|Z$, if $(X|Z) \perp\!\!\!\perp (Y|Z)$ (Bickel and Doksum, 2015).

In the Bayesian formulation of probability (e.g. Gelman et al. (2013)), which is followed in this work, it is conventional to write $p(x)$ instead of $f_X(x)$ for a random vector X and from here on that notation is adopted, except for where reference to the particular random vector is explicitly desired. While traditionally in the *frequentist* view, any model parameters are treated as unknown constants, in the *Bayesian* setting they are modeled as random variables and the object of interest then is how those parameters are distributed.

The famous *Bayes' theorem*, on which Bayesian probability theory and statistics are based (Gamerman, 1997), states that for arbitrary random variables X and Y ,

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \quad (2.6)$$

and is directly proven by noting that $p(y|x)p(x) = p(y, x) = p(x|y)p(y)$. The term on the left hand side is the *posterior distribution* of X , the term $p(y|x)$ is called the

likelihood, and despite the notation is considered to be a function of x , $p(x)$ is the *prior distribution* that codifies all our *a priori* knowledge of the parameters X , and $p(y)$ is the *marginal likelihood* of observations y , or sometimes *evidence* (Gelman et al., 2013; Tarantola, 2005), that usually cannot be computed in closed form.

Bayes' formula presents a way to update our knowledge regarding a random variable by updating the prior distribution with new data. Let μ^y denote the *posterior measure*, the measure corresponding to $p(x|y)$ in (2.6) as its pdf, and let μ^0 similarly denote the *prior measure* with pdf $p(x)$. With fixed data y , using (2.1) and (2.6), and given some μ^y -measurable set F ,

$$\begin{aligned}\mu^y(F) &= \int_F p(x|y)dx \propto \int_F p(y|x)p(x)dx = \int_F p(y|x)d\mu^0 \\ &\Rightarrow p(y|x) \propto \frac{d\mu^y}{d\mu^0}\end{aligned}\quad (2.7)$$

meaning that the likelihood is proportional to the Radon-Nikodym derivative of the posterior with respect to the prior. The benefits of this approach is discussed in detail by e.g. Stuart (2010).

2.1.2 MODEL CALIBRATION VIA BAYES' THEOREM

Bayesian calibration of a nontrivial model M , where the evidence term cannot be evaluated in closed form (e.g. a weather model or some other partial differential equation (PDE) model) is carried out by evaluating the nominator of the right hand side of the Bayes' theorem (denominator can be dealt with algorithmically, see section 2.6.1).

Models used in geophysics and in this work are often discretized in time but the dynamics evolve continuously in time, meaning that the time parameter is in some continuous space, $t \in \mathbb{R}$. Let

$$x \triangleq M(\theta, x_0) \quad (2.8)$$

be the output of a discretization of such a dynamical model for the time-evolution of some initial state vector x_0 governed by parameters $\theta \in \Theta$, where Θ is some set, typically \mathbb{R}^q with some $q \in \mathbb{N}^+$. The observations are denoted by $y \in \mathcal{Y}$, and a function, $\phi \in \mathcal{Y}^{\mathcal{X}}$, called the *observation operator*, is used for mapping the space of model paths \mathcal{X} to the space of observables, \mathcal{Y} (Stuart, 2010). These are in principle some Banach spaces (normed complete linear spaces, see e.g. Rudin (1987)) – for example L^p -spaces – but for discussing a finite number of states and observations, finite-dimensional vector spaces suffice. In practice, in this work the mapping ϕ is the identity map, since the models are (unrealistically) thought to represent real physical quantities.

For Bayesian estimation of parameters in the context of such a system, the *observation equation* or *model equation* (Stuart, 2010) can in case of additive error – perhaps the most common situation – be written as

$$y = \phi(x) + \epsilon, \quad (2.9)$$

where $\epsilon \sim \nu$, where ν is the density function of some probability measure. This density ν , according to which the model-observation mismatch is modeled, is in this work sometimes referred to as an *error model*.

Equation (2.9) defines the likelihood term in (2.6),

$$p(y|\theta) = \nu(y - \phi(x)), \quad (2.10)$$

where explicit dependence on the initial state has been suppressed. Time-discretized versions of (2.8) and (2.9) can then be written as

$$x_i \triangleq M(t_i; \theta, x_0) \quad \text{and} \quad (2.11)$$

$$y_i \triangleq \phi(x)_i + \epsilon_i, \quad (2.12)$$

with states x_i and observations $y_i \triangleq y(t_i)$ taken at times $t_{i=1\dots N}$ and with the discretization of the continuous states mapped by the observation operator defined as $\phi(x)_i \triangleq \phi(x)|_{t=t_i}$. Another common model for the observations y substitutes a multiplicative error for the additive one in (2.9).

The choice of ν dictates how the model-observation mismatch is expected to be distributed and the particular form of ν is a modeling choice, which can be justified by making sure that the residuals obtained by sampling the posterior $p(\theta|y)$ are distributed according to ν . This is a difficult step: first, the residuals may change unexpectedly with θ , especially with models with chaotic dynamics, and second, changing the values of any auxiliary model parameters or how the autocorrelation structures are modeled also affect how ν should be picked. Since in reality the model-observation mismatch may be a result of several different processes with different statistical characteristics (e.g. Tarantola (2005), Ex. 1.22), the final choice of ν is often a well-justified compromise.

2.1.3 FORWARD MODELS AND INVERSE PROBLEMS

The computer implementation of a mathematical model $M(\theta, x_0)$ as described in (2.8) is called a *forward model*, and it is used to solve a *forward problem* (Mueller and Siltanen, 2012; Tarantola, 2005), yielding a discretization of the continuous model trajectory x given initial conditions and any required parameters. If M is computationally extremely demanding, solving the forward problem only once may be the best available approach. Executing the forward model alone does not normally, however, provide information about the model parameter uncertainties, nor does it produce new information about the values of the model parameters.

The *inverse problem* (Mueller and Siltanen, 2012; Tarantola, 2005) associated with the forward problem can be solved to provide estimates of θ and x_0 , either with uncertainties or without. For obtaining point estimates, the problem can take any of

the forms

$$\hat{\theta} = \arg \min_{\theta} \mathcal{L}(\theta) \quad (2.13)$$

$$\hat{x}_0 = \arg \min_{x_0} \mathcal{L}(x_0) \quad (2.14)$$

$$\hat{\theta}, \hat{x}_0 = \arg \min_{\theta, x_0} \mathcal{L}(\theta; x_0), \quad (2.15)$$

where \mathcal{L} is a suitable loss function, for instance a negative logarithm of a likelihood (NLL) given some observations and an observation model. Variations of this particular form, $\mathcal{L}(\theta) = -\log p(y|\theta)$, are used widely in this work. They are treated in more detail in section 2.1.4.

The famous *Hadamard conditions* (Mueller and Siltanen, 2012) state, that a problem is *ill-posed*, if it does not have a solution, if the solution is not unique, or if the solution is not a continuous function of the initial conditions. If none of these conditions (i.e. remove the word not) hold, the problem is *well-posed*. In geophysics all three conditions are often true, meaning that problems are strongly ill-posed, and the practical implications of this often is that the inversion presented in (2.13 - 2.15) gets stuck in local minima since the optimization problems are very rarely convex.

For linear problems with Gaussian errors, (2.13) has a closed-form solution. However, adding noise often quickly shatters the usability of the naive inversion – inverting the forward model – in many systems. This can be avoided by perturbing the setup and adding a regularization term, the most commonly used one of which is the *ridge regression* or *Tikhonov regularization* (Mueller and Siltanen, 2012), which in the Bayesian setting with log-likelihood as the loss function amounts to incorporating a Gaussian prior to the optimization problem in (2.13 - 2.15),

$$\hat{\theta} = \arg \min_{\theta} \mathcal{L}(\theta) + \alpha \|\theta\|_2^2, \quad (2.16)$$

where α is a regularization parameter controlling the prior weight. By using different forms of the regularization term such as $\|\theta\|_{\Gamma}^2$ for some positive definite matrix Γ , different types of prior formulations can be prescribed.

In the context of geophysical models, closed-form solutions for the inversion are not available and optimization algorithms need to be used. The present work utilizes L-BFGS (Nocedal, 1980), BOBYQUA (Powell, 2009), and Nelder-Mead (Nelder and Mead, 1965) algorithms for solving for point estimates in various inverse problems. For state estimation, the 4DVAR algorithm (Dimet and Talagrand, 1986) is commonly used in numerical weather prediction, and the Kalman filter family of algorithms can be utilized for both state and parameter estimation.

While prescribing a prior alone may work, several other approaches are available to work around the problem of local minima in the response surface of the loss function. For obtaining point estimates, stochastic optimization algorithms such as stochastic gradient descent have become popular recently, especially in the machine learning

community. While this is a viable approach, it is not feasible when the gradients are not available. For moderate parameter dimension, scaling down \mathcal{L} to ensure mixing and using a Markov chain Monte Carlo algorithm to find $\mathbb{E}[\theta]$ can work reliably and sometimes be faster than using global optimization algorithms such as ISRES (Runarsson and Xin Yao, 2005).

2.1.4 STANDARD POINT ESTIMATION AND CROSS VALIDATION METHODS

If $\mathcal{L}(\theta) = -\log p(y|\theta)$ in (2.13), the corresponding $\hat{\theta}$ is called a *maximum likelihood estimate* of θ . By setting $\mathcal{L}(\theta) = -\log p(\theta|y)$, the *maximum a posterior estimate* (MAP) is obtained instead (Casella and Berger, 2002; Stuart, 2010). Since log is a monotonous function, its presence above is not necessary, but often convenient. The MAP estimate corresponds to $\hat{\theta}$ in (2.16) when \mathcal{L} in that equation is the NLL. While maximum likelihood estimation is useful and often used, it may overestimate the confidence in the predictive performance of the model. If observation/model error covariance is not fully known, it is relatively common practice in geosciences to use a diagonal covariance model with a Gaussian likelihood as a best guess (for an example, see e.g. Mäkelä et al. (2016)).

From a Bayesian perspective, overconfidence with predictive performance is a general issue for point estimation, since any predictive quantities obtained by using a point estimate for model parameters do not reflect the uncertainty that should be carried over by the propagation of uncertainty in those model parameters. This may be overcome by using *cross-validation* (Gelman et al., 2013), where the cross-validation prediction error for a set of observations A_i is estimated by excluding that set from the training set, obtaining an estimate for the model parameters θ denoted by $\hat{\theta}_{XV}^i$ using that training set, and then predicting the observations in A_i using the parameters $\hat{\theta}_{XV}^i$ and finally comparing to the true observed quantities. Usually, $\cup_{i=1}^M A_i = A$ and $A_i \cap A_j = \emptyset$ when $i \neq j$. A much-used special case is when, for all i , $A_i = \{y_i\}$. This is called *leave one out cross validation* (LOOCV) (Gelman et al., 2013). Cross validation is used in a regression modeling setting in Paper II to evaluate what independent variables best explain annual model parameter variations produced by the hierarchical model used.

2.2 UNCERTAINTY

2.2.1 SOURCES OF UNCERTAINTY

The term ϵ in (2.9) describes the total uncertainty in the model-observation mismatch $y - \phi(x)$. In reality it needs to describe errors from various sources. If characteristics of separate sources of model-observation mismatch are known, ϵ can and should be split into several different components (Stuart, 2010). In Paper II, where the model-data mismatch is known to change in time, ARMA modeling is used to describe

the correlation structure in the time series while additional modeling accounts for heteroscedasticity.

The most straightforward error source to describe is often the *measurement error*, which describes the error contribution from sensor noise of the instrument making the measurement. This error component is typically assumed to be independent and identically distributed (i.i.d.) Gaussian. However, for instance in the case of CH₄ flux measurements in Paper II, it is better described by the Laplace distribution (Richardson et al., 2006).

For discretized dynamical models, *representation error* describes how averaging over a finite domain (e.g. time-space hypercube of a grid point from one model time step to the next) to compare with localized observations induces error (e.g. Ganesan et al. (2014)). This source is controlled by the exact form of ϕ .

Other sources are *random model error* and *model bias* due to for instance rare or unmodeled events or incomplete information about the initial conditions, *autocorrelation* of the observation errors, and *numerical error* from finite machine precision.

While problems arising from machine precision can be a nuisance, e.g. when calculating a Cholesky factorization, $C = L^T L$ (Trefethen and Bau, 1997; Boyd and Vandenberghe, 2004) of a covariance matrix with a large condition number using single precision, a greater difficulty with geophysical models (and many other models as well) is caused by model bias and unmodeled events. An example of such an event is extreme drought in Finland, where photosynthesis is normally not limited by the availability of water, and models typically have not needed to take that to account.

2.2.2 DISTRIBUTIONS FOR UNCERTAINTY MODELING

In the context of any specific problem, the form of ϵ in the observation equation (2.9) needs to be prescribed. The choices utilized in the various problems tackled in this thesis are presented in this section.

A random vector X following the *normal* or *Gaussian distribution* (Casella and Berger, 2002) with mean μ and covariance matrix C has the probability density function

$$\mathcal{N}(\mu, C) \triangleq f_X(x) = (2\pi)^{-\frac{n}{2}} |C|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \|x - \mu\|_C^2\right), \quad (2.17)$$

where $\|x - \mu\|_C$ stands for $\sqrt{(x - \mu)^T C^{-1} (x - \mu)}$. If the random vector X is split into two parts of sizes p and q , i.e. $X = (X_1, X_2)^T$, then the joint distribution can be written as

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} C(X_1, X_1) & C(X_1, X_2) \\ C(X_2, X_1) & C(X_2, X_2) \end{bmatrix}\right) \quad (2.18)$$

and the conditional distribution $f_{X_1|X_2}$ is Gaussian with its moments given by

$$\mathbb{E}[X_1|X_2] = \mu_1 + C(X_1, X_2)C(X_2, X_2)^{-1}(X_2 - \mu_2) \quad (2.19)$$

$$\text{Cov}(X_1|X_2) = C(X_1, X_1) - C(X_1, X_2)C(X_2, X_2)^{-1}C(X_2, X_1). \quad (2.20)$$

The right hand side in (2.20) is known as the *Schur complement* of $C(X_2, X_2)$ (Boyd and Vandenberghe, 2004), and in the setting of (2.18), the marginal distribution $\int_{\mathbb{R}^q} f_{X_1, X_2}(x_1, x_2) dx_2$ is also Gaussian.

For any random variable Z , with mean μ_z and finite variance σ_z^2 , the *central limit theorem* (CLT) states, that at the limit when $N \rightarrow \infty$, $\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{Z_i - \mu_z}{\sigma_z} \xrightarrow{d} \mathcal{N}(0, 1)$ (e.g. Williams (1991); Vershynin (2018)). In practice this means that large-sample averages are well-behaved in that their tails are controlled by the squared exponent in (2.17). The CLT does not, however, state how fast the tail probabilities vanish – this depends on what kind of a random variable Z is. For instance for sub-Gaussian random variables (tails probabilities decaying at least at squared exponential speed) Hoeffding’s inequality gives the exact tail bounds (Vershynin, 2018).

The χ^2 -distribution with $k \in \mathbb{N}^+$ degrees of freedom describes the distribution of the sum of squares of k standard normal $\mathcal{N}(0, 1)$ random variables and hence is supported on $x > 0$. The weighted sum of squares from the quadratic form in the log-likelihood of a normal observation model, (2.17), is χ^2 -distributed, given that in that equation the Cholesky factor of C whitens the residuals $x_i - \mu_i$.

The *scaled inverse χ^2 -distribution* (Gelman et al., 2013) adds a scale parameter $s > 0$, and has the pdf

$$f_X(x) = \left(\frac{k}{2}\right)^{\frac{k}{2}} e^{-\left(\frac{2x}{ks^2}\right)} \frac{s^k x^{-\left(\frac{k}{2}+1\right)}}{\Gamma(k/2)} \quad (2.21)$$

with $\mathbb{E}[X] = \frac{s^2 k}{k-2}$ and $\mathbb{V}[X] = \frac{2k^2 s^4}{(k-2)^2 (k-4)}$. It can be used for e.g. prescribing priors for variance parameters.

Often in finite sample sizes the tails of the normal distribution are too thin. A heavier-tailed version to be used in finite-sample settings would be the *Student’s t -distribution*, but we utilize the *two-sided exponential* or *Laplace distribution* (Casella and Berger, 2002) instead, with pdf

$$f_X(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \mu|}{\sigma}\right), \quad (2.22)$$

where μ and $\sigma > 0$ are the location and scale parameters, respectively. Additionally, $\mathbb{E}[X] = \mu$ and $\mathbb{V}[X] = 2\sigma^2$. Flux measurements done with instruments designed to be used for measuring trace gas fluxes at the biosphere-atmosphere boundary have been reported to follow the Laplace distribution instead of the normal distribution.

The *uniform distribution* is denoted by $\mathcal{U}_{[0,1]}$ and has the probability density function $f_X(x) = 1_{[0,1]}$. If X follows the discrete *Bernoulli distribution* with parameter p , denoted $X \sim \text{Ber}(p)$, it has the probability mass function $f_X(x) \in \{0, 1\}$ s.t. $\text{Pr}(X=1) = p$ (Casella and Berger, 2002).

All of the aforementioned continuous distributions belong to or are closely related to the *Gamma family* of distributions, an exponential family (Bickel and Doksum, 2015). This is convenient and by design, since using distributions from the same

family results in conjugacy that can be exploited when used in e.g. hierarchical models, as described in section 2.7.

2.3 LINEAR REGRESSION

One of the most commonly used tools in any context to statistically analyze data is *linear regression* (Casella and Berger, 2002), which amounts to fitting parameters controlling the orientation of a hyperplane to minimize squared error between that plane and some data. Let $A \in \mathbb{R}^{p \times n}$ be a data matrix containing n vector-valued measurements, called *independent variables*, of length p in the columns, let $y \in \mathbb{R}^n$ be a vector of *dependent variables*, and let $\beta \in \mathbb{R}^p$ be a vector of *regression coefficients* with prior covariance Σ . The regression problem is written as

$$A^T \beta = y + \epsilon, \quad (2.23)$$

where $\epsilon \sim \mathcal{N}(0, \Gamma)$ is the measurement error associated with y . For an exactly determined or overdetermined system, $\text{rank}(A) \geq p$, the (Tikhonov-) *regularized least squares solution* of (2.23) and its covariance are given by (Tarantola, 2005)

$$\mathbb{E}[\beta|y] = \hat{\beta} = \arg \min_{\beta} \left\{ \|A^T \beta - y\|_{\Gamma}^2 + \|\beta\|_{\Sigma}^2 \right\} \quad (2.24)$$

$$= (A\Gamma^{-1}A^T + \Sigma^{-1})^{-1}A\Gamma^{-1}y, \text{ and}$$

$$\text{Cov}(\beta|y) = (A\Gamma^{-1}A^T + \Sigma), \quad (2.25)$$

where the notation $\|\cdot\|_{\Sigma}$ was defined in the context of (2.17). If in this equation $\Sigma^{-1} = 0$, $\hat{\beta}$ is the *ordinary least squares solution*. As an alternative, the use of sparsity-inducing ℓ^1 -norm for regularization is customary in big data applications, but this comes at the cost of needing to use an algorithm such as the *least absolute shrinkage and selection operator* (LASSO) for obtaining $\hat{\beta}$ (Tibshirani, 1996). In this work only ridge regression-type regularization and Gaussian priors are used, however. For further details, see e.g. (Lassas and Siltanen, 2004).

2.4 GAUSSIAN PROCESSES

Given an index set \mathcal{D} , a *stochastic process* or *random function* is an indexed set of random variables $X_d : (\Omega, \mathcal{F}, \mu) \rightarrow (S, \mathcal{S})$ for all $d \in \mathcal{D}$ (Williams, 1991), and the space S is usually taken to be \mathbb{R}^d with the Borel σ -algebra $\mathcal{B}(\mathbb{R}^d)$. A classical example of a stochastic process is the Wiener process (random walk), for which $\mathcal{D} = \mathbb{R}^+$, and it holds that

$$d_1 < d_2 < d \Rightarrow (X_{d_1} \perp\!\!\!\perp X_d) | X_{d_2}, \text{ and} \quad (2.26)$$

$$X_d | X_{d_2} \sim \mathcal{N}(x_{d_2}, d - d_2). \quad (2.27)$$

A *Gaussian process*, or *Gaussian random function* is a stochastic process indexed over an index set \mathcal{D} defined by a *mean function* $m(d)$ and a *covariance function* $k(d, d')$ in that for any finite collection of size N of indices $D \subset \mathcal{D}$, the joint distribution of random variables $\{X_d : d \in D\}$ is multivariate Gaussian (Rasmussen and Williams, 2006) with

$$X_D \sim \mathcal{N}(m, K), \quad (2.28)$$

where $X_D = (X_{d_1}, \dots, X_{d_N})^T$, $m = (m(d_1), \dots, m(d_N))^T$ and K is a matrix with elements $K_{ij} = k(d_i, d_j)$. The requirement that K is a covariance matrix implies that k is symmetric in its arguments. While the measure-theoretic treatment of stochastic processes leads to many interesting results, this level of mathematical detail is not needed here; see e.g. (Karatzas and Shreve, 1998; Stroock, 2018; Rozanov, 1998; Øksendal, 2010) for further details.

If the index set \mathcal{D} is two-dimensional, the term *Gaussian field* is often used in the literature. For a time-dependent process, the index set is usually taken to be \mathbb{R}^+ and, non-surprisingly, the letter t is used. For the Gaussian processes in Paper I, a spatio-temporal index set is used and the index set elements are denoted by x . A Gaussian process is *stationary* if its unconditional mean and covariance functions do not change under translation.

That a quantity of interest Ψ is modeled as a Gaussian process with mean and covariance functions $m(d)$ and $k(d, d'; \theta)$, is written $\Psi \sim \text{GP}(m(d), k(d, d'; \theta))$, following Rasmussen and Williams (2006) and Gelman et al. (2013). Given a set of observations $\psi_D \in \mathbb{R}^n$ of the quantity of interest Ψ indexed by some index set $D \subset \mathcal{D}$, the log marginal likelihood for a given set of covariance kernel parameters θ can be directly evaluated (Rasmussen and Williams, 2006) via (2.17) as

$$\log p(\psi_D | \theta) = -\frac{1}{2} \|\psi_D - m\|_K^2 - \frac{1}{2} \log |K| - \frac{n}{2} \log(2\pi). \quad (2.29)$$

The mean function selection affects the maximum likelihood estimate of the covariance parameters given observed data, and the decision of what to include in the mean function and what to leave for the covariance function is a modeling choice.

For calculating the marginal distribution of Ψ at some *test input* $d^* \notin D$, $d^* \in \mathcal{D}$ conditioned on observations ψ_D , equations (2.18 - 2.20) can be directly employed, with $X_1 = \Psi^*$ and $X_2 = \psi_D$. Here Ψ^* denotes the marginal distribution of random field Ψ at test input d^* . The covariance $K(\psi_D, \psi_D)$ then has the elements $k(d, d')$ with $d, d' \in D$.

An alternative way to model the evolution of randomness in a dynamical system is a *dynamic linear model* (DLM), in which a state space model is used in conjunction with the Kalman filter or Kalman smoother algorithms for parameter estimation (Durbin and Koopman, 2012; Gamerman, 1997). Given a linear model M and a Gaussian observation model in (2.9), the Kalman filter first predicts $x_t | x_{t-1}$, the state at time t given the state at time $t - 1$ and its covariance, and then updates those estimates using any available observations at time t . In Paper I the DLM approach could

have been utilized for state estimation, much like was done in Laine et al. (2014), even though due to the size of the problem and the nature of the data this would have been challenging.

2.4.1 A PARAMETRIC FORM FOR THE GAUSSIAN PROCESS MEAN FUNCTION

In case the mean of a Gaussian process prior is not zero, a convenient way of prescribing it is via the parametrization (Santner et al., 2003)

$$m(d) = \sum_{i=1}^p \zeta_i(d) \beta_i, \quad (2.30)$$

where ζ_i are some functions of index d that are expected to capture the dynamics of the variation, and β_i are coefficients that can be determined for best fit.

Let F be a matrix with elements $F_{ij} = \zeta_i(d_j)$ and θ be the covariance function parameters. The least-squares solution for the β -parameters are once again given by (2.23 - 2.25), with $A \leftarrow F$, $\Gamma \leftarrow K$, and $y \leftarrow \psi$, yielding

$$\beta | \psi, \theta \sim \mathcal{N} \left((FK^{-1}F^T)^{-1}FK^{-1}\psi, (FK^{-1}F^T)^{-1} \right). \quad (2.31)$$

While this form is useful in that knowing the covariance model it allows one to get closed-form point estimates of the β -factors and their uncertainties, it does not cover non-linear cases such as parameters appearing in the arguments of a non-linear ζ_i .

2.4.2 GAUSSIAN PROCESS COVARIANCE KERNELS

There are several standard parameterized forms for describing covariance kernels, and of those the *exponential*, *Matérn*, and *periodic kernels* are utilized in Paper I. The notation presented here is from that paper, and it is reused in chapter 3.

Let θ be the set of parameters controlling a covariance kernel, often containing at least a *scale parameter* ℓ , and a *maximum covariance parameter* τ^2 . As a shorthand, let

$$\xi_{\ell_I}^\gamma(d, d') = \sum_{c \in I} \frac{|d_c - d'_c|^\gamma}{\ell_c^\gamma}, \quad (2.32)$$

where $I \ni c$ is the set of dimensions c of the members of the index set \mathcal{D} . For instance for a Gaussian process indexed with both a time and space dimension s.t. $d = (d_x, d_t)^T \in \mathbb{R}^2$, it is natural that the time and space axes can have different covariance scale parameters ℓ_x and ℓ_t .

The γ -*exponential family* of covariance kernels (Rasmussen and Williams, 2006) with $\theta = (\gamma, \ell, \tau^2)$, is defined by the covariance function

$$k_{\text{exp}}(d, d'; \theta, I) = \tau^2 \exp \left(-\xi_{\ell_I}^\gamma(d, d') \right), \quad (2.33)$$

which, with $\gamma = 2$, yields infinitely differentiable realizations of the random process that look very smooth.

The *Matérn family* of covariance kernels (Rasmussen and Williams, 2006), with $\theta = (\nu, \ell, \tau^2)$, is given by

$$k_M(d, d'; \theta) = \frac{\tau^2 s^\nu}{\Gamma(\nu) 2^{\nu-1}} K_\nu(s), \quad (2.34)$$

where $s = 2\sqrt{\nu} \xi_{\ell_i}^1(d, d')$, $\nu = \alpha - \frac{q}{2}$, where α is a smoothness parameter and q is the dimensionality of s , and K_ν is the modified Bessel function of the second kind of order ν . The value $\alpha = \infty$ corresponds to the squared exponential kernel and $\alpha = 1$ corresponds to the exponential kernel with $\gamma = 1$. Despite this similarity between the Matérn and exponential kernels, the realizations of the random function from the processes with values $1 < \alpha < \infty$ do not correspond to those from the kernel k_{exp} with any values of γ . The smoothness parameter ν is not estimated in this work, but that can also be done, see e.g. (Roininen et al., 2018).

The Matérn kernel is expensive to evaluate for any ν that is not a half-integer, since K_ν is an infinite series that truncates only for the half-integer values. Figure 2.1 gives a visual example of how realizations from exponential and Matérn can look like.

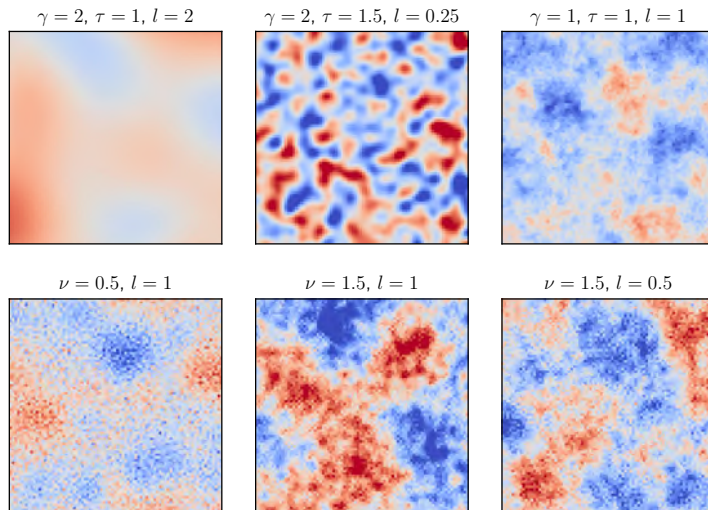


Figure 2.1: Example draws from Gaussian processes with exponential (first row) and Matérn (second row) covariance kernels show how the smoothness and scale change when the covariance kernel parameters θ are varied. These draws were generated by explicitly calculating the covariance matrix K by evaluating the covariance function in question between all pixel pairs, and drawing from that covariance by multiplying an i.i.d. standard normal vector by the Cholesky factor of K .

A *periodic kernel* with $\theta = (\tau^2, \ell_{\text{per}}, \theta_{\text{exp}})$ is defined in Paper I based on (Gelman

et al., 2013) by

$$k_{\text{per}}(d, d'; \theta, I) = \tau^2 \exp \left(-2\ell_{\text{per}}^{-2} \sin^2 \left(\frac{\pi(t - t')}{\Delta_t} \right) - \xi_{\ell_{I \setminus \{t\}}}^\gamma(d, d') \right), \quad (2.35)$$

in which the term θ_{exp} defines the scale parameters for the exponential functions ξ controlling the spatial component, and ℓ_{per} gives the periodic (e.g. inter-annual) covariance width for the temporal dimension. Normally the exponential spatial dependence, the last term in the exponent (2.35) is not present, but in the context of this work the periodicity is wanted to be restricted to the time dimension only.

The periodic kernel can be used to describe situations, where the dynamics of the data is expected to be periodic. For instance carbon dioxide fluxes do have an annual cycle due to the seasons repeating every year. If the mean function has periodic bias, this also can be caught by the periodic kernel. The periodic kernel is particular in that covariance over large temporal distances is possible, and, as in the context of Paper I, it can therefore be thought of having predictive capabilities even outside the temporal domain of the available observations.

A symmetric matrix $C \in \mathbb{R}^{n \times n}$ is *positive semi-definite* (PSD) if $\|x\|_C \geq 0$ for all $x \in \mathbb{R}^n$ (e.g. (Gruber, 2013)). In this work PSD matrices are always symmetric, even though sometimes the notion is taken to more generally refer to the situation where $\frac{1}{2}(C^T + C)$ is PSD. Since sums of PSD matrices are PSD, linear combinations of covariance functions are also valid covariance functions. This allows for lots of flexibility in describing combined effects of covariances of different scale, roughness, and amplitude.

A *multi-scale covariance kernel*, as defined in Paper I, captures covariances at various length scales. Given observation error variances of σ_x^2 for each observation at x , the multi-scale covariance function may then have for instance the form

$$k(x, x'; \theta) = \delta_{x, x'} \sigma_x^2 + k_{\text{per}}(x, x'; \theta, I_S) + k_M(x, x'; \theta) + k_{\text{exp}}(x, x'; \theta, I_{ST}). \quad (2.36)$$

These kernel components are called *subkernels* in Paper I. What complexity and how many scale levels or kernel components are needed depend on the data. The identifiability of the parameters given data sampled using a multi-scale kernel is looked at in section 3.3.3.

The Gaussian process prediction problem can be solved locally using covariance tapering, as done in Paper I. Such *Vecchia approximations* (Vecchia, 1988) have been recently studied also by others, e.g. with a satellite remote sensing application to chlorophyll fluorescence data presented by Katzfuss et al. (2018). There are also various additional types of approximations to Gaussian processes to make them tractable with large data sets. A recent comparison of these methods is presented in Heaton et al. (2017).

2.5 GRAPHICAL MODELS

A *graphical model* (Lauritzen, 1996; Wainwright and Jordan, 2008) is a model on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where \mathcal{V} is the set of *vertices* or *nodes*, and \mathcal{E} is the set of *edges*. The vertices correspond to random variables and the edges describe how those random variables depend on each other. Graphical models facilitate describing the conditional dependence structure, such as Markov structure, in Bayesian models. They are often used in situations where these dependency structures are complex and approximate inference algorithms are used to make the inference task tractable. The objective of the inference task is typically find out marginal distributions of the nodes, or the joint MAP estimate.

The graphical model framework can be also seen as an approach to looking at (typically high-dimensional) statistical inference problems. For different classes of graphical models there exist standard algorithms for performing inference (Wainwright and Jordan, 2008). This usually amounts to calculating expectations or point estimates or sampling from posterior distributions, conditionals, and marginals.

2.5.1 DIRECTED GRAPHICAL MODELS

A *directed graphical model* (Wainwright and Jordan, 2008) or a *Bayesian network* describes the conditional dependence structure of a *directed acyclic graph* (DAG), where edges $e_{i \rightarrow j} \in \mathcal{E}$ have a direction and where there are no loops. In that case, the meaning of the DAG is, that the joint distribution of all the vertices factorizes according to

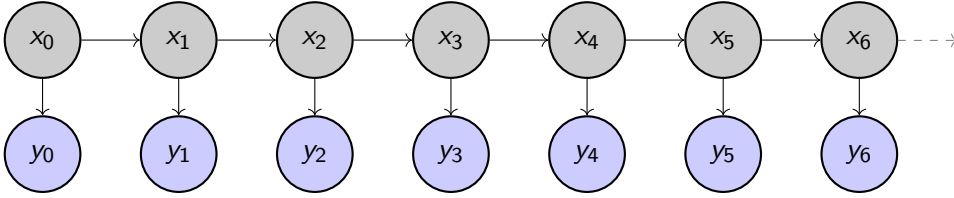
$$p(\{\nu \in \mathcal{V}\}) = \prod_{\nu \in \mathcal{V}} p(\nu | \nu_{\text{parents}}). \quad (2.37)$$

The *Kalman filter* (KF) (Kalman (1960), for a modern exposition see e.g. Särkkä (2013) or Law et al. (2015)) can be used as example of such a model, as is shown in figure 2.2, where the conditional dependence structure described by the arrows implies that the joint distribution $p(x_0 \dots x_N, y_0 \dots y_N)$ may be computed as in (2.37) as $p(y_0 | x_0) p(x_0) \prod_{i=1}^N p(y_i | x_i) p(x_i | x_{i-1})$. This decomposition is a modeling choice, without which use of the KF algorithm would not be justified. The KF can also be described as an algorithm for solving the *hidden Markov model* (HMM) represented by the graph and its decomposition – hidden in that the states x_i are not directly observed, and Markov since $Y_i | X_0 \dots X_i = Y_i | X_i$. Alternatively, a *state space model* could be used, by specifying e.g. $x_{i+1} = f(x_i) + \xi_i$ and $y_i = x_i + \epsilon_i$ with $\xi_i \sim \mathcal{N}(0, \Gamma_i)$ and $\epsilon_i \sim \mathcal{N}(0, \Sigma_i)$ for some covariance matrices Σ_i and Γ_i . In the case of the standard KF, the function f would be linear. The state space approach is developed thoroughly for time series data by e.g. Durbin and Koopman (2012). Due to not using the KF in the Papers, the KF update formulas are not presented here.

In addition to the Kalman filter, for instance Markov chain Monte Carlo algorithms and hierarchical Bayesian models, both of which are described later, can be described as directed graphical models. Such models can be thought of as being *generative* in

that given parents ν_{parents} , realizations of ν can be generated directly. This paradigm is widely used in machine learning with for example generative adversarial networks and other models, see e.g. Goodfellow et al. (2014).

Figure 2.2: In the Kalman filter algorithm, the mean and the covariance of state x are updated at each time step i whenever observations $y_i \sim Y_i$ become available, as is represented by this DAG.



2.5.2 UNDIRECTED GRAPHICAL MODELS

An *undirected graphical model* or *Markov random field* (MRF) is a graphical model whose edges are not directed. These undirected edges determine if the global, local, or pairwise *Markov properties* hold (Lauritzen, 1996). These properties are equal if $p(\nu_{\mathcal{V}})$ is always strictly positive. For the algorithms in this work, the global Markov property is assumed, stating that any two different vertices ν_i and ν_j of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which are separated by a set of nodes \mathcal{A} (in other words $e_{\nu_i, \nu_j} \notin \mathcal{E}$) are conditionally independent given \mathcal{A} .

With the condition $p(\nu_{\mathcal{V}}) > 0$, the joint distribution of the graph can be written as a *maximal clique factorization*,

$$p(\{\nu \in \mathcal{V}\}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi(\nu_c), \quad (2.38)$$

where Z is the normalizing *partition function*, and the set of maximal cliques \mathcal{C} , with $\cup \mathcal{C} = \mathcal{V}$, contains maximal sets of nodes c such that if $\nu_i, \nu_j \in c$, then $e_{\nu_i, \nu_j} \in \mathcal{E}$. The functions ψ are called *potential* or *compatibility functions*. For a lattice graph the maximal cliques are the adjacent pairs of random variables.

According to Hammersley and Clifford (1971), the maximal clique factorization and the conditional independence structure given by the graph are essentially identical. This suggests that efficient algorithms can be derived by working with maximal cliques of a graph.

The mean function modeling in Paper I is an example of an undirected graphical model, where the spatial dependence of the mean function parameters of a Gaussian process is modeled according to (3.3) and these β -parameters are allowed to change from one location to another based on spatially local observations. The important difference between the undirected and (acyclic) directed graphical models is that the

interdependence of the nodes is bi-directional ruling out straightforward sequential inference by just following the arrows of a DAG. Connections between Gaussian MRFs and Matérn class Gaussian processes is discussed e.g. by Lindgren et al. (2011).

An example of approximate inference in an undirected graph is given in figure 2.3, where marginalization in a lattice graph can be carried out effectively by diagonally calculating marginals corner-to-corner and back and carefully accounting for propagating beliefs (calculated marginals). In the variable elimination algorithm, the reconstituted graphs after elimination would have diagonal edges. In Paper I those are not considered for performance reasons, since the nodes diagonal to each other can then be computed in parallel due to absence of diagonal edges in the graph. Since solving the β coefficients in (2.31) involves inverting the covariance matrix K , and since this inversion is an $\mathcal{O}(n^3)$ process in the number of observations, computing the marginals in parallel separately is for large grids around 100 times faster even with a 12-threaded standard desktop workstation.

Not using the exact variable elimination algorithm does introduce an approximation error, but in the application of Paper I it is for several reasons typically either small or very small. First, with remote sensing data there are generally a large number of observations available for computing the β for each vertex ν . This means that when the spatial resolution of the grid is not excessively fine, the covariance with the observations selected for the other vertices of the reconstituted graph, referring to matrix K of the joint system in (2.31), would be much smaller. Second, when there are not that many observations available for fitting the parameters at each vertex the different vertices will share observations leading to similar β coefficients due to shared data. Third, at present only the modes are actually used, and therefore the joint and marginal variances of the β factors are not of paramount importance. Implementation of the exact variable elimination algorithm is planned to be added to the software tool presented in Paper I in the future.

2.6 MONTE CARLO ALGORITHMS

In this section X denotes a real-valued random variable or vector, i.e. $X : (\Omega, \mathcal{F}, \mu) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \Lambda)$, and y denotes the observed data. Formulations with more general state spaces are common, but in this work \mathbb{R}^d with Lebesgue measure Λ suffices.

Given any distribution $f_X(x)$ that samples need to be generated from, if its cumulative distribution function $F_X(x)$ is available, then independent samples can be trivially generated with the *inverse cumulative distribution function sampling* or *inversion sampling* (Tarantola, 2005): If $u \sim \mathcal{U}_{[0,1]}$, then obviously $F_X^{-1}(u) \sim f_X(x)$. A closed form of the cumulative distribution function is, however, rarely available.

In the Bayesian inverse problem setting Bayes' theorem (equation (2.6)) is used for finding the posterior distribution $p(x|y)$ of x by evaluating the likelihood function, $p(y|x)$. When $p(y|x)$ is computationally demanding to evaluate, usually also the evidence term $p(y) = \int_{\mathbb{R}^d} p(y|x)p(x)dx$ is intractable, and for finding the pos-

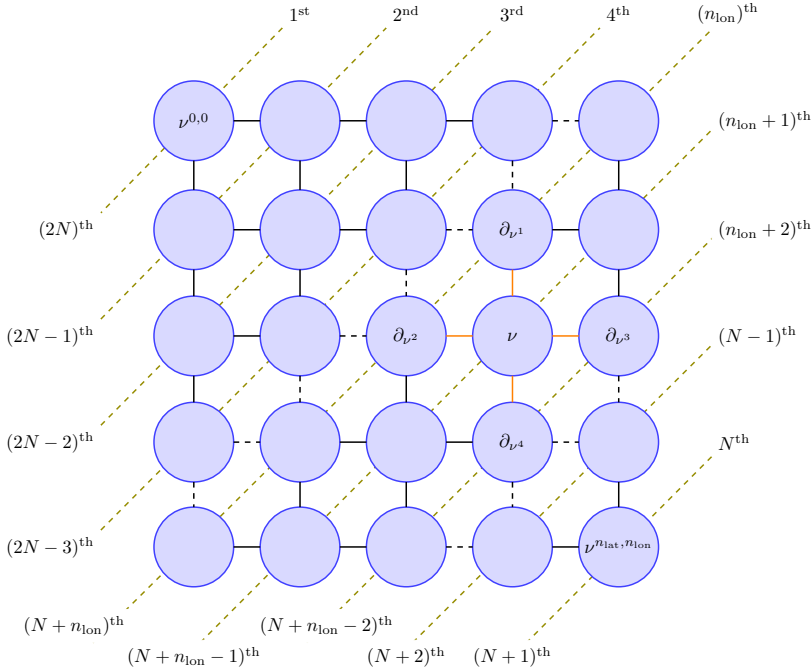


Figure 2.3: Figure from Paper I with ρ_i denoted by ∂_{ν^i} . The marginal distribution of vertex ν , $p(\nu)$, is conditional only on the neighbors $\partial_{\nu^1} \dots \partial_{\nu^4}$ (red edges) due to the Markov structure in the pictured lattice graph. This graph is used for solving for mean function coefficients in Paper I. Each connected pair is a maximal clique in this particular case. For effective solving, the vertices on the diagonal dashed lines are computed simultaneously. The order numbers labeling the diagonal lines represent an ordering in which the diagonals can be computed in parallel to get all the marginals in $\mathcal{O}(N)$ wall time, where $N = n_{\text{lat}} + n_{\text{lon}} - 1$. The $(N + 1)^{\text{th}}$ computation in the corner is not conditioned on already-computed neighbors to avoid double counting data.

terior distribution clever algorithms are needed. *Monte Carlo algorithms*, nowadays discussed in a multitude of standard references such as Gamerman (1997); Gelman et al. (2013); Bickel and Doksum (2016); Tarantola (2005), are algorithms that utilize randomness for calculating expectations or drawing samples from a distribution. They are useful and necessary when a closed-form expression of the likelihood function is not available, which is always the case with complex geophysical process models.

One of the simplest such methods is the *rejection sampling* (Bickel and Doksum, 2016) algorithm: given an unknown unnormalized distribution $f(x)$, a constant M , and a known distribution $g_X(x)$, such that i.i.d. samples can be generated from $g_X(x)$, and given that for all $x \in \text{support}(f)$ it holds that $f(x) \leq Mg(x)$, samples x_i drawn from g are accepted as samples from f if $Y_i = 1$, where $Y_i \sim \text{Ber}\left(\frac{f(x_i)}{Mg(x_i)}\right)$. While this method works well for very low-dimensional targets if a good guess at M and g are available, the curse of dimensionality quickly destroys its performance. For this reason more sophisticated algorithms are utilized, which despite often producing correlated samples offer far superior performance. The Monte Carlo methods used in this work and described in this section are *Gibbs sampling*, *Adaptive Metropolis MCMC*, and *sampling-importance resampling* (SIR).

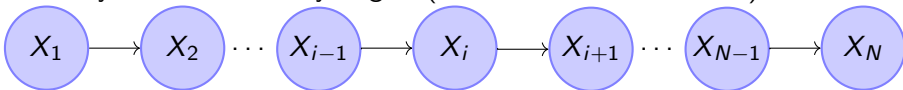
2.6.1 MARKOV CHAIN MONTE CARLO

A *Markov chain* is a sequence of random variables $X_1, X_2 \dots X_N$ such that for all i , the Markov property

$$p(x_i | x_1 \dots x_{i-1}) = p(x_i | x_{i-1}) \quad (2.39)$$

is satisfied (Gamerman, 1997). This extremely simple generative model is described in figure 2.4. As a conceptual bridge to section 2.4, Markov chains can be characterized as random functions with a discrete index set $\mathcal{D} = \mathbb{N}^+$ (Williams, 1991) and as with stochastic processes in section 2.4, it is useful to think about X_i as states of a dynamical system and of the indexes i as time. An obvious difference to the rejection sampling and inverse cdf sampling methods is that the samples generated by MCMC are not independent.

Figure 2.4: The random variables in a Markov chain depend only on the value of the preceding member. In MCMC algorithms this is exploited for efficiently generating correlated samples from a desired target distribution $p(x)$, since in theory $X_i \sim p(x)$ approximately for all sufficiently large i (Bickel and Doksum, 2016).



A Markov chain is *homogeneous* (Bickel and Doksum, 2015) if (2.39) is satisfied and $f_{X_i | X_{i-1}}(x_i | x_{i-1}) = f_{X_2 | X_1}(x_2 | x_1) \forall i \in \mathbb{N}, i \geq 2$, in other words the conditionals are not dependent on the index i . The evolution of a homogeneous chain is determined

by the *Markov transition kernel*,

$$q(x_1, x_2) = p(x_2|x_1), \quad (2.40)$$

which is a function giving the probability of transitioning from x_1 to x_2 . If the state space is finite, then q is a matrix, whose elements q_{ij} give the transition probabilities from x_i to x_j . Given an MCMC chain with state space \mathcal{X} and with transition kernel $q(\cdot, \cdot)$, the *stationary distribution* of the chain, π , if it exists and is unique, is given by

$$\pi(y) = \int_{\mathbb{R}^d} \pi(x)q(x, y)dx \quad (2.41)$$

with some states $x, y \in \mathbb{R}^d$. If a chain can be constructed in such a way that it can be thought of as integrating over the state space as in the formula above, then the MCMC chain, after discarding some *burn-in* (or *warm-up* according to Gelman et al. (2013)) period to forget the starting point of the chain, can be seen as representing correlated draws from π . For if $X_i \sim \pi$, then $X_j \sim \pi$ for all $j > i$. Not any q will do, however, and the conditions for allowing this interpretation are clarified below. The following definitions are presented e.g. in Gamerman (1997) and Bickel and Doksum (2015).

A Markov chain is *stationary* if $\forall k \in \mathbb{N}$ and $\forall m \in \mathbb{N}^+$,

$$f_{X_1, \dots, X_m}(x_1, \dots, x_m) = f_{X_{k+1}, \dots, X_{k+m}}(x_{k+1}, \dots, x_{k+m}). \quad (2.42)$$

A finite state Markov chain is *aperiodic* if it does not revisit the same state at fixed intervals, and for continuous state spaces such as \mathbb{R}^n , aperiodic chains do not visit any sets $F \subset \mathbb{R}^n$ s.t. $\Lambda(F) > 0$ at fixed intervals.

A finite state Markov chain is *positive recurrent*, if the expected visit time to any state is finite. For continuous states, the concept of *Harris recurrence* is used instead: a chain is Harris recurrent if the probability of revisiting any set $F \subseteq \mathbb{R}^q$ s.t. $\Lambda(F) > 0$ in a finite number of steps is one.

A finite state Markov chain is *irreducible* if any state x in the state space is reachable from any other state x' in a finite number of steps. For continuous state spaces, let ν be a measure on some state space (S, \mathcal{S}) , typically $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The chain is ν -*irreducible* if for any $x \in S$ and any $F \subset S$ with $\nu(F) > 0$ there is an i_0 such that $\Pr(X_{i+i_0} \in F | X_i = x) > 0$. The chain is irreducible if any such ν exists.

An irreducible and aperiodic finite-state Markov chain is called *ergodic*. In the continuous case a Markov chain is ergodic if it is irreducible and Harris-recurrent. Furthermore, if the chain satisfies *detailed balance*,

$$\pi(x)q(x, y) = \pi(y)q(y, x), \quad (2.43)$$

then the chain is said to be *reversible*. This can be stated in that the *probability flow* from x to y is the same as from y to x . For an ergodic MCMC chain, (2.43) can be trivially manipulated with

$$\int_{\mathbb{R}^d} \pi(x)q(x, y)dx = \int_{\mathbb{R}^d} \pi(y)q(y, x)dx = \pi(y). \quad (2.44)$$

Equation (2.44) is actually just (2.41), meaning that transitioning from x , which is a random draw from the stationary distribution π , with the kernel q , yields another draw from π .

This development leads to the conclusion that devising transition kernels which generate ergodic reversible chains is desirable, since such chains ultimately automatically produce samples from the target distribution. The effectiveness, however, depends on the *mixing time* – how fast the random state variables of a Markov chain initialized at random end up being distributed according to π – and how correlated the samples are.

The *Metropolis-Hastings (MH) algorithm* is by far the most famous MCMC algorithm, and it satisfies the detailed balance condition (Geman, 1997). Let $t(y; x)$ be the *proposal density*, which is a probability density on \mathbb{R}^d evaluated at y .

The MH transition kernel $q(x, y)$ is defined with the help of the *acceptance probability*

$$\alpha(x, y) \triangleq \left(1 \wedge \frac{\pi(y)t(x; y)}{\pi(x)t(y; x)} \right), \quad (2.45)$$

using which it can be written in the form

$$q(x, y) = t(y; x)\alpha(x, y) \quad x \neq y. \quad (2.46)$$

From this it follows that the probability the chain stays put is

$$\Pr(X_{i+1} = X_i) = 1 - \int_{z \in \mathbb{R}^d \setminus \{x_i\}} q(x_i, z) dz, \quad (2.47)$$

as demonstrated by Geman (1997) in the discrete state space case. The function t is most often parametrized by the location x of the chain at the previous iteration. A notable exception to this rule are independent proposals, which do not depend on the current chain location x .

While ergodicity depends on the proposal distribution, detailed balance for the MH algorithm follows from a direct calculation: for $\pi(y)t(x; y) > \pi(x)t(y; x)$ and $\pi(x)t(y; x) > \pi(y)t(x; y)$, respectively,

$$\begin{aligned} \pi(x)q(x, y) &= \pi(x)t(y; x)\alpha(x, y) = \pi(x)t(y; x) = \alpha(y, x)t(x; y)\pi(y) = \pi(y)q(y, x), \\ \pi(y)q(y, x) &= \pi(y)t(x; y)\alpha(y, x) = \pi(y)t(x; y) = \alpha(x, y)t(y; x)\pi(x) = \pi(x)q(x, y), \end{aligned} \quad (2.48)$$

and the case $\pi(y)t(x; y) = \pi(x)t(y; x)$ is trivial. When the proposal is symmetric, $\frac{t(x; y)}{t(y; x)} = 1$ and the MH-algorithm reduces to the *Metropolis algorithm* (Tarantola, 2005).

The power of the MH algorithm (and MCMC algorithms in general) is in how the density π in (2.46) is evaluated. When sampling a posterior distribution as given by Bayes' theorem (2.6), the observed data is fixed and whether a proposed point is

accepted or not depends only on the outcome of a Bernoulli trial whose parameter is the ratio of the posterior density evaluated at the proposed and current points. This makes evaluating the evidence term unnecessary.

Since the samples generated with MCMC are correlated, calculating the *effective sample size* (ESS) is useful. The (1-d) ESS is defined by (Gamerman, 1997)

$$\eta = N \left(1 + 2 \sum_{i=1}^{\infty} \rho_i(x_n)_{n=1}^{\infty} \right)^{-1}, \quad (2.49)$$

where N is the length of the Markov chain $(x_n)_{n=1}^N$, $\rho_i \triangleq \text{Corr}(x_1 \dots x_{N-i}, x_{i+1} \dots x_N)$ is its lag- i autocorrelation coefficient, and the series is in practice truncated due to finite chain length and due to that after the initial decay in autocorrelation the terms tend to only contribute noise. There are many options for computing essential sample sizes for multivariate chains, but a canonical version of the ESS does not exist. One common way is to compute the ESS for each coordinate projection separately.

Draws generated with MCMC algorithms should be seen as draws from the posterior only after the chain has mixed well (Gelman et al., 2013), since only after some i_0 it is true that $X_{i_0+i} \sim \pi$ for all $i \in \mathbb{N}$. A practical way to find such an i_0 is to run multiple chains initialized at random points and to include as posterior samples from each chain the tails s.t. the inter-chain statistics agree with the within-chain statistics. If a single chain is used for e.g. computational reasons, whether the chain finds the target distribution or not can be usually also seen by looking at the chain for each state variable separately. In the rare case when π is multi-modal, comparing 2-d pairwise marginals with varying degrees of burn-in may be more revealing. While MCMC is used in Papers I, II, and IV, only the experiments with real-world data in Paper I exhibited multi-modal features (not shown).

As for the other Monte Carlo estimates, central limit theorems apply, implying that the variance of the estimator for the mean of a scalar target density π behaves according to $|\mathbb{E}[\pi] - \frac{1}{N-i_0} \sum_{i_0}^N x_i| \sim \mathcal{N}(0, \frac{\sigma^2}{\eta})$, where η is given by (2.49) and iterations before i_0 have been discarded as burn-in.

The applications presented in Papers I, II and IV use a variation of the MH algorithm, the *Adaptive Metropolis* (AM) algorithm (Haario et al., 2001), which produces non-homogeneous chains and therefore is not reversible. The chains are, however, ergodic and converge to the target distribution when the chain length tends to infinity. The sampling procedure with AM is identical to that of standard MH, except for that the covariance of the Gaussian proposal density is recalculated every once in a while¹ to match the sample covariance, scaled by the factor $\frac{2.4^2}{d}$, where d is the parameter dimension. This choice yields an optimal acceptance ratio for Gaussian targets (Roberts et al., 1997).

¹How often the adaptation is done is implementation-dependent. It is known, however, that adapting at every iteration may lead to the algorithm misbehaving. As an example, in Paper II the adaptation was done whenever the iteration number was the square of an integer.

If the target is non-Gaussian, AM will still work, but not quite as effectively. In order to decrease the correlatedness of the samples, the *Delayed Rejection* (DR) algorithm (Tierney and Mira, 1999) may be implemented on top of AM, resulting in the *Delayed rejection adaptive Metropolis* (DRAM) algorithm (Haario et al., 2006). When a proposed state is rejected for the first time, the DR algorithm, instead of immediately repeating the previous value in the chain, proposes other points from scaled proposal densities. These later proposals are accepted with a modified acceptance probability that takes care of that the chain remains reversible. Practical experience, for instance from preliminary simulations for Paper II, showed that while DRAM works it will in many cases not improve nor deteriorate the performance of the sample generation. With multi-modal targets, the performance can, however, be dramatically improved with DR (see comment SC1 by Laine, Susiluoto, Tamminen, and Haario in the discussion of Lu et al. (2017)).

There are many alternative MCMC algorithms and new ones are, such as (Titsias and Papaspiliopoulos, 2016; Bouchard-Côté et al., 2015), are continuously being developed. Many of these more modern algorithms as well as the older Metropolis-adjusted Langevin (Grenander and Miller, 1994) and Hamiltonian Monte Carlo (Duane et al., 1987), utilize gradient information of the posterior density to improve the quality of the samples. Without gradients or good guesses at the covariances in the posterior, the AM method in practice performs well, as is shown in the Papers.

2.6.2 GIBBS SAMPLING AND METROPOLIS WITHIN GIBBS

Gibbs sampling is a Markov chain Monte Carlo method, where a multivariate target π of state $X = (X^1, \dots, X^d)^T$ with random variable elements X^i is sequentially sampled component by component with the Markov transition kernel

$$q(x, y) = \prod_{i=1}^d p(y^i | z^{-i}), \quad (2.50)$$

where $z^{-i} = (y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_d)$ (Gamerman, 1997). The resulting chain is homogeneous, and Gibbs sampling has been shown to have the full joint posterior distribution as the stationary distribution (Tierney, 1994). Gelman et al. (2013) presents Gibbs sampling as a special case of the MH algorithm.

Given the form of the proposal, it is natural to use Gibbs sampling in situations where analytic forms of the conditionals are available. This situation arises when a hierarchical statistical model is constructed utilizing conjugate priors, as outlined in section 2.7. When some of the conditionals are not available in closed form, other forms of sampling may be employed, such as rejection sampling or the Metropolis algorithm. In the latter case, the algorithm is then called Metropolis within Gibbs (Gamerman, 1997). It is used in Paper II.

Generally the number of samples needed in Monte Carlo sampling scales very poorly with parameter dimension even with best methods. Limiting the dimension

of the Metropolis-sampled part may help somewhat due to that in Gibbs sampling proposed parameters are always accepted (Gamerman, 1997). In the presence of parameter correlations, Monte Carlo algorithms proposing the correlated parameters together are generally superior. If an approximation of the joint posterior density is available, this can be used to rotate the parameter axes for achieving better mixing. Often-cited methods improving sampling efficiency based on this idea, applicable in generic Monte Carlo sampling settings, include for instance Active subspaces (Constantine et al., 2015), Likelihood-informed subspaces (Cui et al., 2014), and truncating the standard singular value decomposition (Mueller and Siltanen, 2012; Gruber, 2013).

2.6.3 IMPORTANCE SAMPLING AND RESAMPLING

Let random variable $X \sim \pi$ take values $x \in \mathcal{X}$, and let $\pi_b(x)$ be a distribution from where samples can be generated, called the *biasing distribution*, with the corresponding measure denoted by μ_b . *Importance sampling* is a method for estimating expectations of a function $g(x)$ by evaluating it at samples drawn from π_b and re-weighting those samples according to the ratio of g and π_b . More formally,

$$\mathbb{E}[g(X)] = \int_{\mathcal{X}} \frac{g(x)\pi(x)}{\pi_b(x)} d\mu_b(x), \quad (2.51)$$

which with a finite sample of size N becomes

$$\mathbb{E}[g(X)] \approx \hat{g} \triangleq \sum_{x \sim \pi_b} \frac{g(x)}{\pi_b(x)/\pi(x)}, \quad (2.52)$$

where \hat{g} is called the *importance sampling estimate of $\mathbb{E}[g(X)]$* . Gelman et al. (2013) treat $\pi(x)$ as an unscaled posterior density, but while this may be a useful depiction, conditioning on data is not generally necessary for describing importance sampling.

Importance sampling is particularly useful for rare event simulation with computationally demanding models; in case only tails of a parameter distribution trigger a rare event such as a catastrophic drought, flood, nuclear reactor meltdown or a flu pandemic, compared to naive Monte Carlo sampling the accuracy of the calculated expectation can be increased dramatically by using a biasing distribution with most of the mass in this rare event triggering region. The condition $\text{support}(g) \subseteq \text{support}(\pi_b)$ needs to be satisfied for importance sampling to work – otherwise it could happen that no samples are used from an area in \mathcal{X} where g is large, and this would introduce bias to \hat{g} . The optimal choice for the biasing distribution that minimizes the variance of \hat{g} is $\pi_b(x) \propto |g(x)|\pi(x)$ (Casella and Berger, 2002).

If the samples for computing the sum over the biasing distribution in (2.52) are taken from a previous Monte Carlo sample, the procedure of computing \hat{g} can also be used to generate samples from g by re-weighting those Monte Carlo samples with $\frac{\pi(x)g(x)}{\pi_b(x)}$ and drawing independently according to the obtained weights. With $g(x) \equiv 1$

this is called *Sampling-Importance-Resampling* (SIR) (Gelman et al., 2013). The method can be useful e.g. if after conducting a Monte Carlo experiment there is need for an adjustment of the likelihood function, or if adequate data exists for repurposing output of model evaluations for calculation of additional statistics. The SIR method is utilized in Paper II.

2.7 HIERARCHICAL BAYESIAN MODELS

A *hierarchical Bayesian model* (Gamerman, 1997; Gelman et al., 2013) is a modeling approach for situations where parameters of a distribution need to be modeled as random variables. A typical example and the one utilized in this work involves creating a model for an ensemble of related experiments indexed with $i \in \{1, \dots, n\}$. These experiments are conducted in such a way that the dependency structure of the data on any associated random variables is shared but observations y_i differ for each ensemble member.

In the hierarchical model described in figure 2.5 the parameters θ_i , possibly associated with each ensemble member, share a common prior distribution $p(\theta_i|\nu)$ parameterized with parameters ν , but the parameters θ_i can have different posterior marginal distributions $p(\theta_i|y)$ depending on the data y . The ν are called *hyperparameters* and their priors are called *hyperpriors* (Gelman et al., 2013). It is possible to have the data also depend on auxiliary parameters τ that are prescribed a fixed prior. In such a setting the full joint posterior distribution can be written as

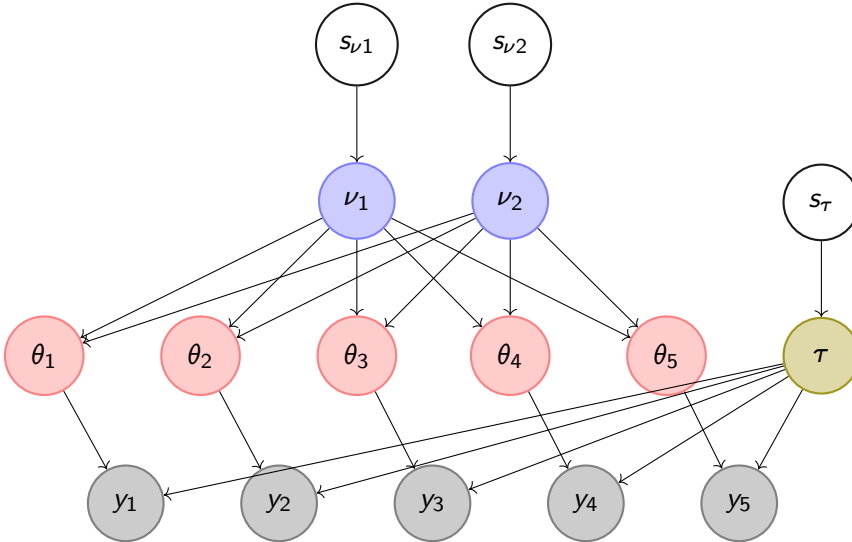
$$p(\theta, \tau, \nu|y) \propto p(\nu)p(\tau) \prod p(y_i|\theta_i, \tau)p(\theta_i|\nu). \quad (2.53)$$

Figure 2.5 expresses the model described above as a directed graphical model, as is customary for hierarchical models (Wainwright and Jordan, 2008). The graphical description intuitively reveals the conditional independence structure in (2.53).

More complex hierarchical models describing multiple levels of shared dependence structure may also be constructed. For instance the (hyper)parameters ν could again depend on other random variables ζ with assigned hyperprior distributions $p(\zeta)$, instead of just depending on the fixed parameters s .

Sampling from posterior distributions of these models often is facilitated by using *conjugate priors*, meaning that the families of the prior distributions are chosen to be such that the conditional distributions have closed forms and are easy to sample from (Gelman et al., 2013; Gamerman, 1997). This is for instance the case in the example of (2.53) if $\tau \sim \mathcal{N}(\mu_\tau, \Sigma_\tau)$ with some vector τ and covariance Σ_τ , $\nu = (\mu_\theta, \sigma_\theta^2)$, $\theta_i \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, $\mu_\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$, and $\frac{1}{\sigma_\mu^2} \sim \text{Scale-inv-}\chi^2(s)$ with some hyperprior parameters μ_0, σ_0^2 , and s . If the θ_i parameters are not mutually correlated, and especially if furthermore the τ parameters are not correlated with the θ_i , then the θ_i parameters may be effectively Gibbs-sampled and the dimension of Metropolis-sampling the τ parameters remains smaller as mentioned earlier in section 2.6.2.

Figure 2.5: An example of the simple hierarchical model in (2.53) described as a DAG. Observations y_i are generated by parameters θ_i , all of whose priors depend on hyperparameters ν_1 and ν_2 , and parameters τ which have a fixed prior with parameters s_τ . The priors of the hyperparameters ν are the hyperprior distributions with fixed parameters s_{ν_1} and s_{ν_2} .



This is called *Metropolis within Gibbs* or *Gibbs within Metropolis*, depending on the source.

2.8 BAYESIAN MODELING WITH TIME SERIES DATA

In geosciences, trace gas flux measurements are often done at flux measurement sites with fixed instruments producing time series data. Typical time series measurement data y_t with time index $t \in \{1 \dots T\}$ is evenly distributed in time, even though more often than not there are gaps in the data due to various reasons such as instrument malfunction, power outages, or weather. In Paper II time series flux measurement data is used for Bayesian model calibration. As with other Bayesian modeling, also in that setting the posterior shape depends strongly on how the residual autocorrelations in (2.10) are modeled.

The model M , designed to produce states x related to observations y as in (2.8) and (2.9) has unknown biases and random errors which may be time-dependent, as is the case in e.g. Paper II. The probability model for the model-observation mismatch needs to account for any such structure generated by any error source, see also section 2.1.2.

2.8.1 AR, MA, AND ARMA MODELS

Let the model-data residuals be denoted by r_t with $t \in \{1 \dots T\}$. A possible model for the residual autocorrelations is the *autoregressive model* (Harvey, 1990; Durbin and Koopman, 2012; Chatfield, 1989) of order n , $AR(n)$, which models the time series r with

$$r_i = \sum_{j=1}^n \phi_j r_{i-j} + \epsilon_i, \quad (2.54)$$

where ϕ_j is the *lag- j autocorrelation coefficient* and ϵ_i are some random variables, which are assumed to be independent. However, if after fitting the ϕ_j parameters with a reasonable choice of order n the ϵ_i still *de facto* end up being autocorrelated, other models may be tried. A second much used model for time series data is the *moving average model* of order m , $MA(m)$ (*ibid.*), given by

$$r_i = \sum_{k=1}^m \xi_k \epsilon_{i-k} + \epsilon_i, \quad (2.55)$$

where the difference to the AR models is that while AR models add random error on top of a weighted sum of the previous data values, the MA model adds the random error on top of a weighted sum of previous random errors. These models can be combined to form an *autoregressive moving average model of order (m, n)* , denoted $ARMA(m, n)$ (*ibid.*), with

$$r_i = \sum_{j=1}^n \phi_j r_{i-j} + \sum_{k=1}^m \xi_k \epsilon_{i-k} + \epsilon_i. \quad (2.56)$$

There are several variations to these models such as introducing nonlinearities or exogenous inputs (Durbin and Koopman, 2012). However, since model complexity is a liability when interpreting the results, these were not used in Paper II.

2.8.2 PRACTICAL PARAMETER ESTIMATION IN THE ARMA SETTING

This section follows the presentation in Paper II, where the r_i in (2.56) are generated by a non-trivial dynamical model via (2.9). To perform Bayesian inference with Monte Carlo methods, the parameters ξ and ϕ add another layer of difficulty since the likelihood given any model parameters θ , $p(y|\theta)$, depends on both the ARMA parameters ϕ and ξ in addition to θ . While the fully Bayesian way of doing this would be finding the full joint posterior distribution $p(\theta, \phi, \xi|y)$ via evaluating $p(y|\theta, \phi, \xi)p(\theta, \phi, \xi)$, where the prior parameters would usually be independent, this optimization problem is not generally convex and both minimization algorithms and MCMC algorithm may get stuck in local minima and/or drift to nonphysical parameter regions. It may also happen that the model parameters are not constrained by the data under the statistical model used.

An alternative to finding the full joint parameter posterior distribution is to find a point estimate $\hat{\theta}$ of the model parameters by minimizing some statistic of the data (residuals), e.g. sum of absolute values, running mean, or sum of squares of the residuals or their subset, and then find point estimates for the error model parameters ξ and ϕ . Monte Carlo simulations to find the posterior distribution of the model parameters can then be performed given these estimates.

Given $\hat{\theta}$, finding the order (m, n) of the model and point estimates of the parameters ϕ and ξ can be done by minimizing the *Bayesian information criterion* (BIC) (Bickel and Doksum, 2016) – a standard method for model selection – giving

$$(\hat{\phi}, \hat{\xi}, m, n) = \arg \min_{(\phi, \xi, m, n)} \text{BIC} = \arg \min_{(\phi, \xi, m, n)} \{n_{\text{par}} \log(n_{\text{obs}}) - 2 \log(p(r|\phi, \xi))\}. \quad (2.57)$$

In the above expression, n_{obs} is the number of observations, which in the absence of gaps in data equals T , and $n_{\text{par}} = m + n$ is the number of parameters. Other popular criteria for model selection, such as the Akaike information criterion (AIC) (Bickel and Doksum, 2016), which uses the penalty $2n_{\text{par}}$ instead of $n_{\text{par}} \log(n_{\text{obs}})$, often produce similar (but not identical) results. For finding the ARMA(2,1) parameters used in Paper II, residuals r were simulated by random sampling fifty parameter vectors θ from an approximate posterior, and the vast majority of those residual time series resulted in optimal ϕ and ξ parameters close to each other.

The resulting time series of error terms ϵ in (2.56) can be checked to not to be autocorrelated by calculating the *Durbin-Watson statistic* (Durbin and Watson, 1950, 1951),

$$T(\epsilon) = \frac{\sum_{i=2}^T (\epsilon_i - \epsilon_{i-1})^2}{\epsilon^T \epsilon}, \quad (2.58)$$

where a gapless observation series has been assumed, but gaps can be taken care of if needed by discarding any indexes with no observations. If $T(\epsilon)$ is close to 2, the time series has no substantial lag-1 autocorrelation (see p. 26).

Model residuals as in (2.10) are usually expected to be zero-mean, since any constant term could be simply added to the definition of model M in (2.8). To make sure the error model is correct and the obtained posterior shape is accurate, the appropriate scale parameters for the distribution of the ϵ_i in (2.56) need to be found. If the magnitude of the error terms varies in time as is often the case in geophysical applications, the error is called *heteroscedastic* (Harvey, 1990). To utilize such time series for the likelihood formulation, an easy way to proceed is to preprocess the series using a parametric model γ with parameters α , resulting in a new homoscedastic time series $\epsilon^* = \gamma(\epsilon; \alpha)$.

These α can for uncorrelated zero-mean residuals be found by minimizing some distributional distance such as the *Kullback-Leibler divergence* (KL-divergence) (discussed e.g. in Peyré and Cuturi (2018)) between an empirical distribution (histogram) $\eta_{\gamma(\epsilon; \alpha)}(x)$ of the ϵ^* -terms, and the actual error model $\nu(x)$. The KL-divergence for

two continuous real-valued distributions p and q is given by

$$D_{\text{KL}}(p(x) \| q(x)) = \int_{\mathbb{R}^d} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx \quad (2.59)$$

and the appropriate parameters for the error model can be expressed with it as

$$\hat{\alpha} = \arg \min_{\alpha} D_{\text{KL}}(\eta_{\gamma(\epsilon; \alpha)}(x) \| \nu(x)). \quad (2.60)$$

The process described in this section outlines one simple approach to do *covariance estimation* in a time series context, which is an important and often also a difficult part of Bayesian parameter estimation studies. This difficulty reflects the complexities of the error structures arising from combining real-world data and complicated computer models. The outline of the parameter estimation procedure presented here can be compared with the parameter estimation procedure of Gaussian process covariance kernels in (2.29) and mean functions parameter fields in section 2.5.2, which however omit the model selection and heteroscedasticity considerations. For the GP work in Paper I, these considerations still remain to be fully addressed.

3 APPLICATIONS TO GEOSCIENCES

3.1 OVERVIEW OF SCIENTIFIC CONTRIBUTIONS

The *large-scale geoscientific problems* in the title of this thesis refers to the topics presented in Papers I-IV and this section provides motivational context for that work while the details are discussed later. More emphasis is given to Papers I and II than to Papers III and IV.

3.1.1 SPATIO-TEMPORAL HIGH RESOLUTION CO₂ DISTRIBUTIONS

The research carried out in **Paper I** tries to answer the following question: *Where is the carbon dioxide in the atmosphere?* This question is important in its own right since the general public has shown much interest in it, but the answer could be applied to atmospheric flux inversion, or statistical emission models could be developed based on the results. Since Gaussian processes provide uncertainty information via the theory in section 2.4, the results can also be applied to validation schemes and hypothesis testing.

Current scientific literature of Gaussian processes or kriging applied to atmospheric remote sensing of global CO₂ does not contain any high resolution studies that the authors of Paper I would be aware of. The high number of observations leads to computational compromises which often result in overly smoothed posterior fields. However, the multi-scale approach presented in Paper I is able to produce *arbitrarily high resolution CO₂ maps* with both fine and coarse scale features.

The spatial statistics software presented is not constrained to CO₂ or the OCO-2, but *can be used with any remote sensing data* in principle. The software is able to *learn kernel and mean function parameters*, and is able to *sample from extremely high-dimensionally discretized posterior or prior distributions* as defined by the multi-scale kernel description.

3.1.2 UNCERTAINTIES IN BOREAL WETLAND CH₄ EMISSION PROCESSES

Out of all methane emissions, those from natural wetlands have the highest uncertainty. While this in itself is more than enough reason to study uncertainties in process-based wetland emission models, there is another important reason as well:

changing climate increases uncertainty regarding future emissions. **Paper II** studies a Finnish boreal wetland site with a model that was developed in tandem with writing Paper II (Raivonen et al., 2017). The central questions that we try to answer are *how much uncertainty is there in the model parameters controlling the physical processes, and do the model parameters and hence the wetland's behavior react to environmental changes.*

Many wetland methane emission models have been written, but their systematic calibration has in general not been a research priority in the community. More specifically, at the time of writing we were unaware of any Bayesian calibration studies of wetland emission models. For this reason it is valuable that the work answers questions such as, *given flux measurements, model, and input data, how are the model parameters correlated in the posterior distribution, and how much interchangeability is there between the methane production and transportation processes.*

One difficulty that this study does not yet tackle arises from that the many different types of wetlands all over the boreal region all behave differently. Understanding the functioning of these different environments would require a calibration process for all these types and to accomplish that a spatial statistics or regression/classification study of boreal wetland distributions would be needed. Despite this opportunity for future research, Paper II can be seen as *groundwork for future larger-scale studies of uncertainties related to boreal wetland CH₄ emissions.*

3.1.3 EFFECTS OF CLIMATE CHANGE ON GROWING SEASON AND GROSS PRIMARY PRODUCTION

Carbon emissions to the atmosphere are the main driving force of climate change and while the overall mechanisms have been known for a long time, how climate changes is actually a complex process. The emissions are balanced partly by the uptake of carbon from the atmosphere by plants, and the magnitude of this uptake is controlled by many factors. In the boreal region, the date of snow clearance regulates when the growing season starting date (GSSD). **Paper III** answers the questions *how many days earlier does the growing season start than in the 1970s, and how much additional carbon is getting photosynthesized in this process?*

To answer these questions, Paper III utilizes a wide latitude of flux measurements from all over the boreal region, and compares that with global climate model simulations forced with data from the European Center for Medium-Range Weather Forecasts (ECMWF).

While boreal ecosystems have been studied widely, the connection between snow clearance date and gross primary production (GPP) has not been studied previously in this fashion. This study is a pure simulation study in the sense that, unlike in the other Papers, uncertainties are not quantified (except for providing the *p*-values of regression estimates).

3.1.4 MONTE CARLO ESTIMATES OF LAND SURFACE SCHEME HYDROLOGY AND GAS EXCHANGE PARAMETERS

In land surface schemes (LSS) of climate models the model hydrology description often poses difficulties since changes in hydrological conditions may produce nonlinear effects on other modeled variables such as GPP. In models the hydrology-related sub-routines are intimately connected to the carbon exchange via stomata, since stomata control both CO₂ and water transport in the gaseous phase. Since models utilize discrete plant functional types¹ for land cover description, the parameter values controlling model behavior for each type are generic, average best guesses. Therefore, in case of a rare event such as a major drought, the model may perform worse than it could with calibrated parameter values. Furthermore, the generic parameter values are generally not the best ones available for a particular measurement site.

Against this background, **Paper IV** looks at *how the hydrology-related parameters of the land surface scheme correlate in the posterior distribution*, and asks *whether the model is able to capture a rare event (drought) with calibrated parameters*. In addition, the MCMC calibration is done with different temporal poolings of the data, allowing to look at *how the uncertainty estimates and model performance change depending on the data averaging performed*.

3.1.5 OTHER RELATED WORK

Two additional publications co-authored by J.S., Raivonen et al. (2017) and Mäkelä et al. (2019), are intimately connected to Papers II and IV, respectively. Even though they are not a part of this thesis work, they are briefly mentioned here to give context to Papers II and IV.

In Raivonen et al. (2017), the HIMMELI wetland methane emission model and the physical processes are described in detail, and this article provides motivation behind the modeling choices and more clarity regarding the underlying biology than Paper II does. In its approach, it is purely a model development manuscript and does not explicitly employ the techniques described in Chapter 2.

As a continuation of Paper IV, Mäkelä et al. (2019) evaluates how different stomatal conductance formulations in the JSBACH LSS are or are not able to explain measured fluxes under different environmental conditions. It looks at a wider variety of flux measurement sites (10 as opposed to two in Paper IV), uses an adaptive population importance sampler (APIS) for carrying out the Monte Carlo sampling, and has, due to the different conductance models, a model selection flavor. The more comprehensive and methodical approach than the one taken in Paper IV brings the findings closer to upstream integration to improve the performance of JSBACH in the boreal region.

¹Plant functional types are collections of parameters with which the model distinguishes ecosystem types from each other. These parameters contain values for e.g. maximum leaf area index, biomass, nitrogen deposition rate, etc.

3.2 MODELS, OBSERVATIONS, AND ALGORITHMS CONTROL COMPUTATIONAL COST

3.2.1 PARALLEL MODELS AND PARALLEL ALGORITHMS

Geoscience is a versatile field of applied science that often serves as a testing ground for novel computational methods. Despite this versatility, a large variety of these problems, especially when it comes to uncertainty quantification, are computationally constrained, as is easily seen from the descriptions of the Monte Carlo algorithms in section 2.6. While parallel resources for computation are nowadays readily available, creating efficiently scalable code is a challenge, often due to data and memory bandwidth limitations, but also due to the sequential nature of many sampling techniques.

Figure 3.1 describes the computational cost of the problems tackled in this thesis and helps to explain why the inference algorithms and modeling paradigms were chosen in the very way they were. In **Paper I**, (light blue arrow, bottom right in figure 3.1) the Gaussian process software is able to compute marginals globally in a half-degree grid for every day for four and a half years with OCO-2 data (with reasonable settings) in ten months' time on one CPU core. The inbuilt OpenMP parallelization brings this down to a few days on a modern supercomputer node, but since utilizing several nodes would require architectural changes, the maximum size of the problems is currently limited by available single-node resources. The current implementation requires keeping all observations in memory, and therefore problems with the largest numbers of observations can not be computed on a modern laptop. On a supercomputer node, computing with billions of observations is possible. This also applies to generating gridded draws from the GP.

In **Paper II**, (red and orange arrows in the middle), the forward model – a wetland methane emission model – runs parallelized (downward component of arrows) to yield a speed-up in computation. The experiment was designed so that the forward model simulations of the different years for any given parameter in the MCMC chain were run on different cores simultaneously, meaning that the temporal domain was split into several parts. This guided the inference algorithm choice towards the Metropolis within Gibbs MCMC paradigm, which is well suited for hierarchical modeling, see sections 2.7 and 2.6.1. In the first preliminary experiments (orange arrow), available in the discussion paper of Paper II, several experiments were performed with different model discretizations. While this aspect was dropped from the final version, the parallelization scheme employed decreased the amount of simultaneous model evaluations by the number of different discretizations (leftward component). Together these choices took the core hour requirement down from several years to less than a month. In the final simulations a single MCMC chain was computed and therefore no algorithmic parallelization was possible. However, the speed-up from the time domain decomposition remained.

Paper IV employed several parallel MCMC chains to generate posterior estimates

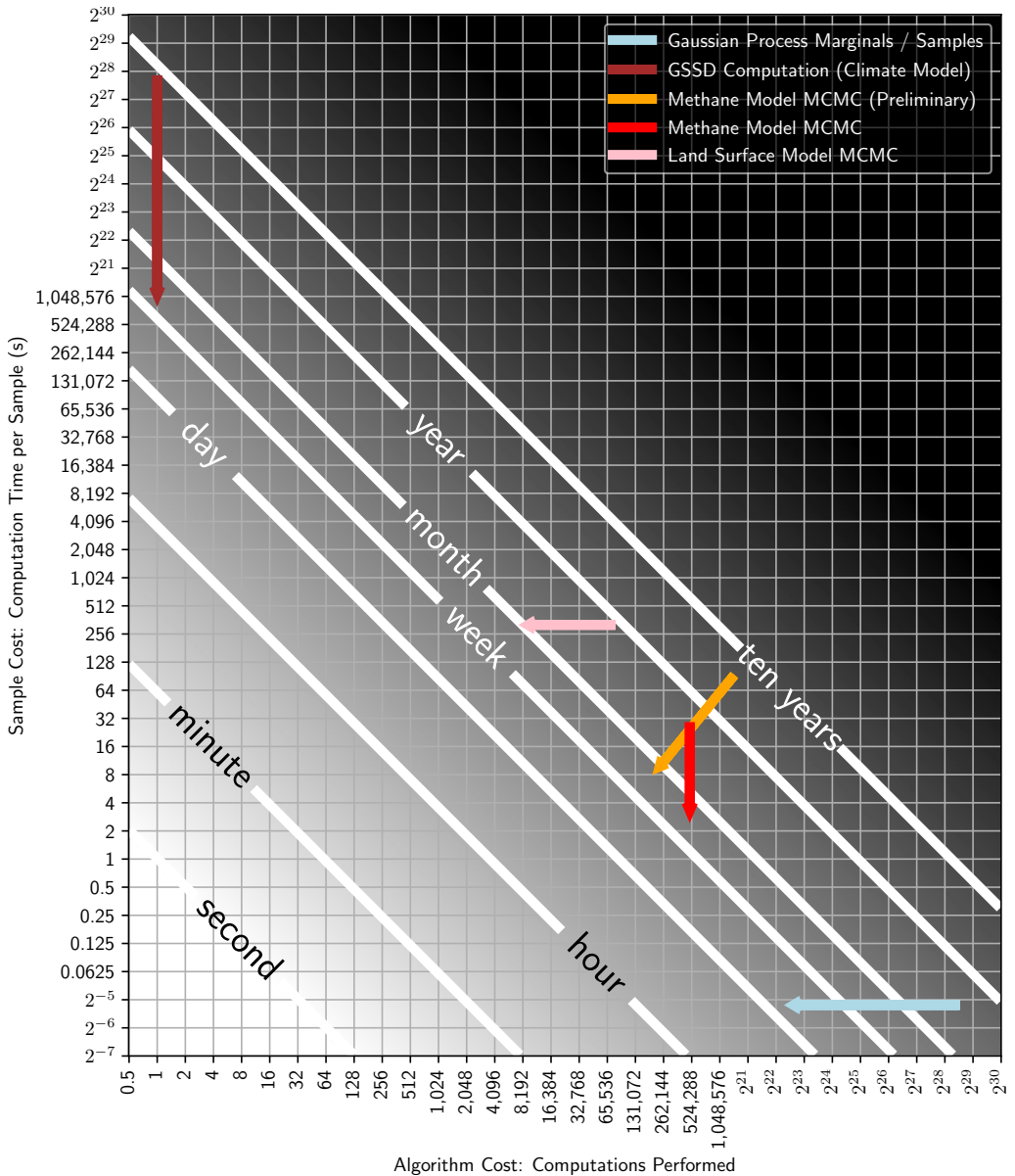


Figure 3.1: This figure shows how the computational cost is divided between sampling schemes and sample generation. The x-axis shows the number of samples, and the y-axis shows the cost per sample. The diagonal white lines show the contours for constant 1-core computation time. Arrows start at the total computational cost of a problem, and end at the computational cost that takes to account both model and algorithm parallelization. All the experiments conducted are constrained by available CPU time and the logarithmic scales along both axes provide perspective to how expensive the most demanding simulations in Papers I-IV were. The black area in the very upper right is unfeasible without massive parallelization or with dedicated accelerators. Increasing Gaussian process model or climate model resolution would easily extend the light blue and the brown arrows to that area. The initialism GSSD in the legend stands for growing season starting date.

of a set of land surface model parameters (pink arrow in the middle). The uncoupled simulations were run on a fast laptop with four hyperthreaded cores. More effective parallelization possibilities could have been utilized on a supercomputer, but that was avoided here to facilitate code development and avoid code porting.

The very opposite to Paper I in terms of parallelization is **Paper III** (brown arrow in the top left corner), where no algorithm was used – just a single climate model simulation with reanalyzed ECMWF forcing data and with the objective of producing data for a regression analysis to back up and quantify other scientific reasoning based on multiple sources of *in situ* measurement data. The simulation was carried out on ten supercomputer nodes (parallelized using the Message Passing Interface library) and in several segments due to model instability. Performing the final simulation required generating initial carbon and water pools, which doubled the amount of computation.

The common denominator of these computational challenges is that only algorithms which *in practice* yield results in a month's walltime are feasible, and the experiments were designed accordingly. While the arrows in figure 3.1 are not normalized in that they represent true computation times in different computing environments, they still reveal where the practical computational constraints are in the research reported in this thesis.

Even though a month is a reasonable amount of time to be spent on computer simulations, for all the above experiments that time is only the tip of an iceberg. Different model configurations needed to be tested and tried before the final product-yielding simulations could be performed, and many of the computational problems in the Papers contained smaller but still important computational sub-problems, such as calculating the MLE of model parameters, creating initial conditions, etc. Those are not pictured in figure 3.1.

3.2.2 THE ROLE OF THE OBSERVATION DATA

Observations are used in the Papers for four primary purposes; (1) forward model forcing, (2) forward model calibration, (3) error model calibration, and (4) forward model validation. The term forward model is reserved here for dynamical models – the statistical models describing the model-observation mismatch are called error models as was discussed in section 2.1.2.

The number of observations and the way they are utilized in the Papers varies wildly and therefore a summary of observation usage is given in Table 3.1. In the table an observation vector may contain several variables, which are detailed in the Papers themselves.

Paper I uses OCO-2 satellite observations, of which there are 249 million for the time period considered, and even after selecting data according to quality-flagging, 116 million observations remain. There is no dynamical forward model, only a statistical model describing the data. Since the model parameters are fitted to all the data, separate validation is not needed. The calculated set of marginals can be effectively

Table 3.1: Number of observation vectors used for each Paper, and the ways in which observations are utilized. Here *forcing* refers to model forcing (setting model states based on data), *learn M* refers to learning parameters θ controlling the model M , as in (2.8), *learn ϵ* refers to learning error model parameters as in (2.9), and *validate* refers to whether observations are used in a direct validation scheme, such as cross validation.

	#obs vectors	Forcing	Learn M	Learn ϵ	Validate	Comments
I	~120,000,000	N/A	N/A	yes	no	Details in section 3.3.1
II	~2,500	yes	yes	yes	yes	
III	~100,000	yes	N/A	N/A	no	Full reanalysis fields
IV	~100,000	yes	yes	no	yes	Data is aggregated

regarded as a statistic $T(y)$ that adequately (to the modeler) summarizes the huge number of observations. The approximate GP algorithm, presented in section 3.3.1, describes how the large number of inputs is handled.

Papers II and IV utilize time series flux observations for constraining the models, and while Paper II also does cross validation for the hierarchical model, in Paper IV a straightforward simple validation is performed on an alternate site. The difference between number of observations is explained by that Paper IV uses half-hourly data, whereas Paper II uses daily means, since the model used in II does not realistically describe the diurnal cycle and therefore using the half-hourly observations would amount to fitting noise. Both of these Papers utilize measurement data to force the forward model, but the error model calibration in Paper IV is not rigorous, while Paper II actually uses draws from an approximate posterior predictive distribution to calibrate the error model parameters before the final Bayesian model calibration is performed.

Paper III does not contain a calibration step, and therefore parameter finding is not applicable. The 100,000 forcing fields are full T63-resolution reanalysis fields from ECMWF – either ERA Interim or ERA-40, depending on the year. Paper III utilizes also flux measurement data from 10 sites in Finland, Sweden, Russia, and Canada, but these are not directly tied to the modeling – only via aggregate statistics in Table 1 of Paper III – and therefore they are not reported here.

With a large number of data, a common complication with Bayesian model calibration in geosciences is that the posterior density may in practice contract towards a point estimate that is sometimes not realistic. This behavior is aggravated by any (often unavoidable) model misspecification, but it also takes place without it in the small observation error limit of overdetermined Bayesian inverse problems, as described e.g. by Stuart (2010). With MCMC the practical implication is that the observation error variance in the observation model may need to be inflated to allow posterior exploration. As a result, the size of the posterior is in the end not necessarily reliable. Paper IV solves this problem by building the statistical model

for data averages instead of individual points, and Paper II utilizes an exploratory MCMC based on which an error model used by the SIR algorithm is calibrated. Despite the calibration and due to model misspecification, the choices the modeler has to make are apparent in how the posterior looks like. Parameter correlations in the posterior are more resistant to log-posterior scaling than e.g. marginal variances, since $\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$. For this reason physics-based interpretation and analysis of results of the Bayesian model calibration is justified to be based on analyzing the posterior correlation structures. Among the Papers, this is most emphasized in Paper II.

3.3 EFFICIENT MULTI-SCALE GAUSSIAN PROCESSES FOR MASSIVE REMOTE SENSING DATA

This and the following sections present the research of each individual Paper in more detail than was done in sections 3.1 and 3.2. For even further details, please consult the Papers themselves.

The first Paper of the thesis, Susiluoto et al. (2019), deals with a computational spatial statistics approach to regularize a sparse set of satellite observations into a spatio-temporal grid with arbitrary resolution. The method used is Gaussian process regression, and both marginals and samples from both the prior and the posterior are obtained. The space-dependent mean function of the Gaussian process is learned utilizing an approximate elimination algorithm on a regular lattice graph to learn the modes of the marginal distributions over a Markov Random Field.

The Gaussian process theory is described in section 2.4, the MRF and the elimination algorithm are described in section 2.5.2, the objectives and highlights of this part were briefly stated in section 3.1.1, and an outline of computational cost and size of the problem was given in section 3.2. While these will be slightly expanded here, the main focus is on additional key details, computation, and discussion.

Several kriging/GP studies such as Zeng et al. (2013, 2017); Nguyen et al. (2014); Hammerling et al. (2012b,a); Tadić et al. (2017); Zammit-Mangion et al. (2015), and Zammit-Mangion et al. (2018) have been conducted with remote sensing CO₂ data over the years. The majority of those have used data from the GOSAT satellite, while a handful of exploratory publications related to the OCO-2 satellite have been published. For details, see the introduction section in Paper I.

Compared to other CO₂ measuring instruments the sun-synchronous OCO-2 satellite is particularly interesting, since it provides high resolution column-integrated dry air CO₂ mole fraction (XCO₂) measurements. It does so by applying an algorithm to retrieved absorption spectra of reflected sunlight. The footprint of a single measurement is only 1.29 by 2.25 kilometers in size, with eight measurements abreast. Clouds and aerosols often result in quality-flagged and missing measurements. The approximate revisiting time to any particular location is 16 days, but obviously not

all area between two trajectories is covered during one 16-day period, and the closer to the equator the satellite is, the larger the uncharted area.

Despite the high spatial resolution of the satellite measurements, there are at the moment, as far as we know, no published CO₂ maps based on only data and showing any of that finer structure. The central problem is computation: in order to calculate the Gaussian process posterior, the covariance matrix of the observations needs to be inverted. This is lots of work with hundreds of millions of observations. How the calculations are performed algorithmically is described next.

3.3.1 GAUSSIAN PROCESS MODEL ALGORITHM DESCRIPTION

The random field Ψ , in Paper I the spatio-temporal XCO₂-field, was defined in section 2.4 to be a Gaussian process, denoted $\Psi \sim \text{GP}(m(x), k(x, x'))$, if the joint distribution of the process at any finite set of points was multivariate normal. The function m had a parametric form given below in (3.3) and exponential, Matérn, and periodic covariance kernels were supported by the software. An additional non-stationary kernel, the wind-informed kernel, is proposed and discussed below in section 3.3.6.

The Gaussian process model computation in practice comes down to computing conditional expectations and variances of the multivariate Gaussian distribution given in (2.19) and (2.20). These distributions are enormous - in the largest simulation in Paper I the dimension n equals 116489343 and storing or solving this size of a linear system, which is an $\mathcal{O}(n^3)$ operation, is not directly possible. For this reason an efficient algorithm and its implementation are needed. The satGP program, consisting of roughly 4000 lines of highly optimized C code and presented in Paper I, is able to approximately compute (level of approximation is controllable with input parameters) the desired spatio-temporal grid of marginals in

$$\text{cost} = \mathcal{O} \left(\frac{An_{\text{times}}}{\omega^2} \left[(n_{\text{ker}}\kappa)^3 + \sum_{l=1}^{n_{\text{ker}}} (r_l \log(r_l) + \kappa \log \kappa) \right] \right) \quad (3.1)$$

time. In this equation, A is the grid area, n_{times} is the number of time steps, ω the grid resolution, n_{ker} the number of subkernels as in (2.36), κ the maximum subkernel size, and $r_l \propto \prod_{i=1}^q \ell_i^l$ is a factor determining the size of the hyper-ellipse outside which covariance with the test input is less than the prescribed covariance threshold σ_{min}^2 . The values of r_l also depend on the maximum covariance parameters τ^2 . This scaling is linear in number of marginals, and the parts in the brackets — first term for inverting the constructed full multi-scale covariance and second for finding observations that are informative for each test input where the marginal is computed — is highly optimized. For additional details regarding observation selection and multiple other computational aspects, please see Paper I and the satGP source code.

The downside of obtaining the linear scaling with the number of test inputs of course is that the full posterior covariance will not be retrieved, only the marginal variances. The posterior covariances can still in principle be calculated from posterior

sample covariances. Another natural possibility is constructing a multi-grid or multi-fidelity Gaussian process (Peherstorfer et al., 2018; Kennedy and O’Hagan, 2000), and this extension would not be impossible to implement in satGP.

The satGP program can draw from the random process by conditioning on previous predictions. Computing the Gaussian field roughly amounts to interpolation by solving a linear system of equations locally at the test input using the observations that are within a desired radius. If the ordering for generating the field is chosen so that instead of interpolation, extrapolation is performed (for instance if in the 1-d case the sampled points would reside at $x_1 = 0$, $x_2 = 0.1$, $x_3 = 0.2 \dots$), such ordering may in practice lead to oscillations in the generated data. For this reason a sparse ordering is used, both in space and time: if the number of inputs where the field is generated is $n_{\text{tot}} = n_{\text{lat}} n_{\text{lon}} n_{\text{times}}$, then the m^{th} computed point is number $(mp \bmod(n_{\text{tot}}))$ in the linear ordering along axes (time, latitude, longitude), with the last of these changing fastest. In the above, p is taken to be the largest prime number under $0.9n_{\text{tot}}$.

The satGP software also contains routines to learn the maximum marginal likelihood estimates (marginalized over the Gaussian process realizations) of the covariance function parameters θ using an approximate random-sampling based method

$$\hat{\theta}_{\text{MLE}} = \arg \min_{\theta} \sum_{x_i \in E_{\text{ref}}} \left\{ \|\psi_i^{\text{obs}}\|_{\tilde{K}_i} + \log |\tilde{K}_i| \right\}, \quad (3.2)$$

where E_{ref} is a set of randomly sampled points from the specified spatio-temporal domain. The vector $\psi_i^{\text{obs}} \in \mathbb{R}^{d_i}$ contains at most $n_{\text{ker}} \kappa$ observations closest in covariance to x_i from which the mean function value at x_i has been subtracted, and \tilde{K}_i is the corresponding covariance matrix determined by the covariance function with parameters θ and the observations ψ_i^{obs} . Due to randomly selecting E_{ref} , this procedure results in the log-likelihood including an unknown multiplicative coefficient and hence an unknown multiplier of covariance in the exponent. Therefore, while posterior mean estimates (for unimodal symmetric), posterior medians, and MAP estimates remain valid, the true size of e.g. credible regions is not known.

The most important input parameters needed by satGP together with the algorithm description illustrate how the software works, and they are shown in table 1 and figure 4 in Paper I.

3.3.2 OBTAINING THE GP MEAN FUNCTION FROM A GAUSSIAN MRF

For describing the XCO₂ field observed by the OCO-2, the mean function (2.30) is assigned the explicit form

$$m(x, t; \beta, \delta) = f(t, \delta)^T \beta = \beta_0 \sin \left(\frac{2\pi t}{\Delta_t} + \delta \right) + \beta_1 \cos \left(\frac{4\pi t}{\Delta_t} + \delta \right) + \beta_2 + Ct, \quad (3.3)$$

where Δ_t is the length of the period, that is, one year, and where the spatial dependence denoted by the argument x comes from the selection of observations for fitting

the coefficients of the mean function. The particular form of (3.3) was chosen for its ability to represent the increase in the CO₂ concentration as a global trend, and also because with this form it is possible to describe the seasons both in the tropics and closer to the poles. The resulting mean function coefficients are shown in figure 3.2.

For spatial smoothness, a Gaussian MRF utilizing the setting presented in section 2.5.2 is used. Since in addition to the β -parameters also the δ -parameter varies from place to place, (2.31) cannot be used directly due to the δ -parameter not conforming to its form. Instead, a first pass calibration is performed utilizing the BFGS gradient-based optimization algorithm to find the mode of all parameters for each vertex, by minimizing

$$l_\nu(\beta, \delta) = \frac{1}{n_{\text{obs}}} \sum_{i=1}^{n_{\text{obs}}} (m(x, t; \beta, \delta) - y_i)^2 + \sum_{\nu' \in \partial\nu} \psi(\nu, \nu'), \quad (3.4)$$

where the latter sum is over the edge potentials corresponding to Gaussian priors defined by the modes of the neighbors. The scaling is arbitrary since the objective is to merely fit the δ -parameter to produce fields that look smooth to enable computing $p(\beta|\delta, y)$.

In propagating the posterior marginals (beliefs) when computing the β -factors, the precision of the neighboring points is scaled according to the distance to those points on the latitude-longitude grid, since close to the poles the grid points are closer to each other than on the equator. For fitting the parameters with (2.31) at each grid point, observations that are nearest in spatial covariance (disregarding the time component) are chosen, and the marginals are computed conditioning on the optimized δ . The uncertainties of the β -factors are given by (2.31), but they are also approximated by the BFGS-algorithm, which therefore in principle could also be used. However, in Paper I the exact computation via (2.31) was utilized.

While only a minor part of Paper I, Fig. 3.2 shows several intriguing features. The constant term β_2 has high values where emission hotspots are known to be. The parameter controlling slow oscillations, β_0 , shows the reversed seasons between the northern and southern hemispheres, and β_1 shows a semiannual signal of higher amplitude in the Congo area. The phase shift parameter δ appears noisy in areas where the β_0 and β_1 parameters are close to zero, which is exactly when the δ -parameter plays very little role. To conclude it is worth remembering that since the parameters act together to describe the CO₂ field, drawing far-reaching conclusions from individual maps should be avoided.

3.3.3 IDENTIFIABILITY OF MULTI-SCALE PARAMETERS

The justification for using the multi-scale covariance kernel formulation, (2.36), is not obvious — it could be that the parameters of the multi-scale model would not be identifiable in practice. In Paper I, multi-scale kernel parameters are recovered from synthetic data generated by drawing a sample from the GP prior. While this

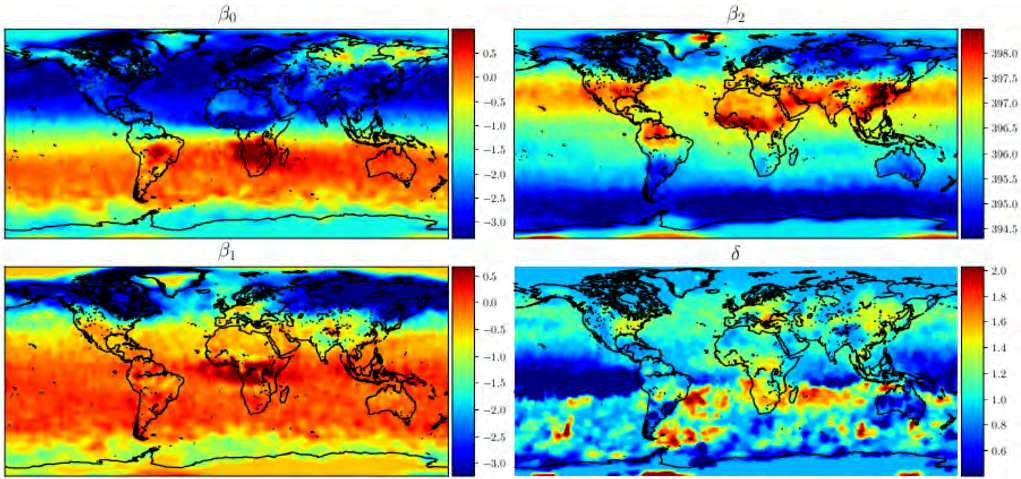


Figure 3.2: Local β factors, and the δ phase-shift of the mean function as in (3.3). The trend component, C , has been fitted to global values, and does not vary spatially.

parameter inversion fails with optimization algorithms, MCMC can be successfully used as a stochastic optimizer. The posterior mean value is a good estimate for parameters as shown in figure 3.3 in a synthetic study with two subkernels. Notice that while the true values are not in the very centers, the scales of the axes reveal that the true values are within a small distance from the center in the parameter space.

This synthetic study validates the multi-scale approach in that since the parameters of the different subkernels are recoverable, the different kernels may indeed be needed for describing the field. In Paper I, a three-component kernel is shown, and while there the length-scale parameters of the smaller-size kernels are slightly overestimated, the ability to approximately find the true parameter values remains.

Table 3.2: Covariance function parameter values learned from OCO-2 data. First column shows the Matérn subkernel parameters, and the second column the parameters of the exponential subkernel.

	$(\cdot) = \text{mat}$	$(\cdot) = \text{exp}$
$\tau^{(\cdot)}$	0.899	2.72
$\ell_{lat}^{(\cdot)}$	0.00513	0.0418
$\ell_{lon}^{(\cdot)}$	0.0363	0.397
$\ell_t^{(\cdot)}$	20h 22min	16d 20h 12min

After validating the parameter estimation process for the multi-scale kernel, the

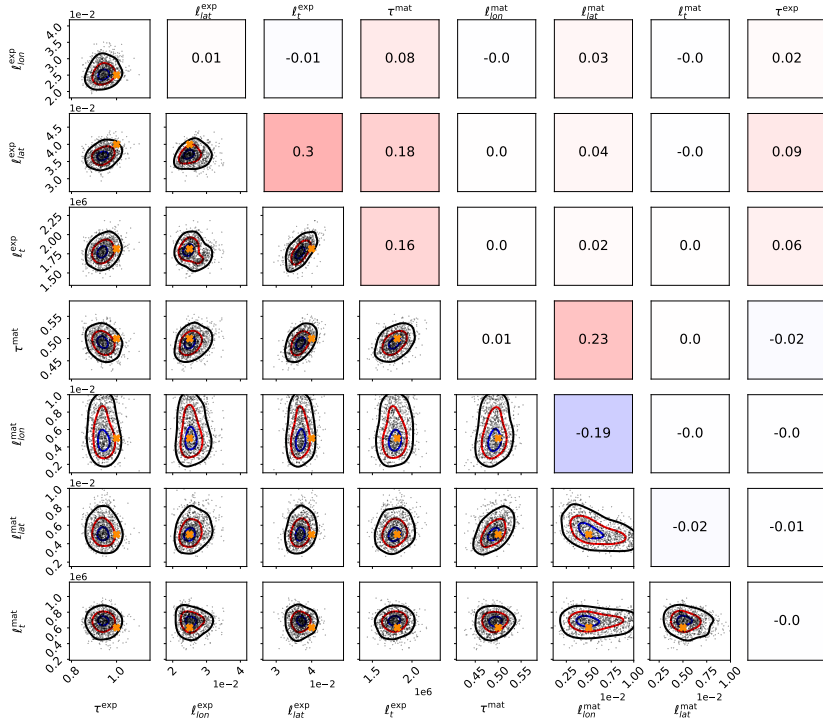


Figure 3.3: Approximate posterior with unknown scaling of the log-posterior from a synthetic study with two subkernels. The first subkernel is of Matérn type and the second an exponential one with smaller and larger length scale parameters respectively. The data was sampled using a random spatial pattern from the prior and 1% noise was added, after which the parameters were learned.

parameters corresponding on the OCO-2 data were learned. No data thinning was applied, and the number of reference points in E_{ref} was set to be 12 with $\kappa = 256$. The resulting parameter values are shown in Table 3.2. The notable aspect of the parameter values is the elongation of covariance ellipses of both kernels in the more informative zonal direction.

3.3.4 LEARNING MULTI-SCALE KERNEL PARAMETERS FROM OCO-2 DATA

The multi-scale kernel allows larger scale features to be combined with local enhancements. In figure 3.4 a covariance kernel consisting of a single subkernel alone with large length scale parameters was compared with that same subkernel combined with a subkernel with shorter length-scale parameters. Observations from the OCO-2 v9 data product were used.

The parameters of the kernels are given by Table 3.2. The total kernel size was kept at 1024 ($\kappa = 512$ for (a-b) and $\kappa = 1024$ for (c-d)) in both experiments. Random data thinning with $\zeta_{\text{train}} = 5$ was applied: the parameter determines how

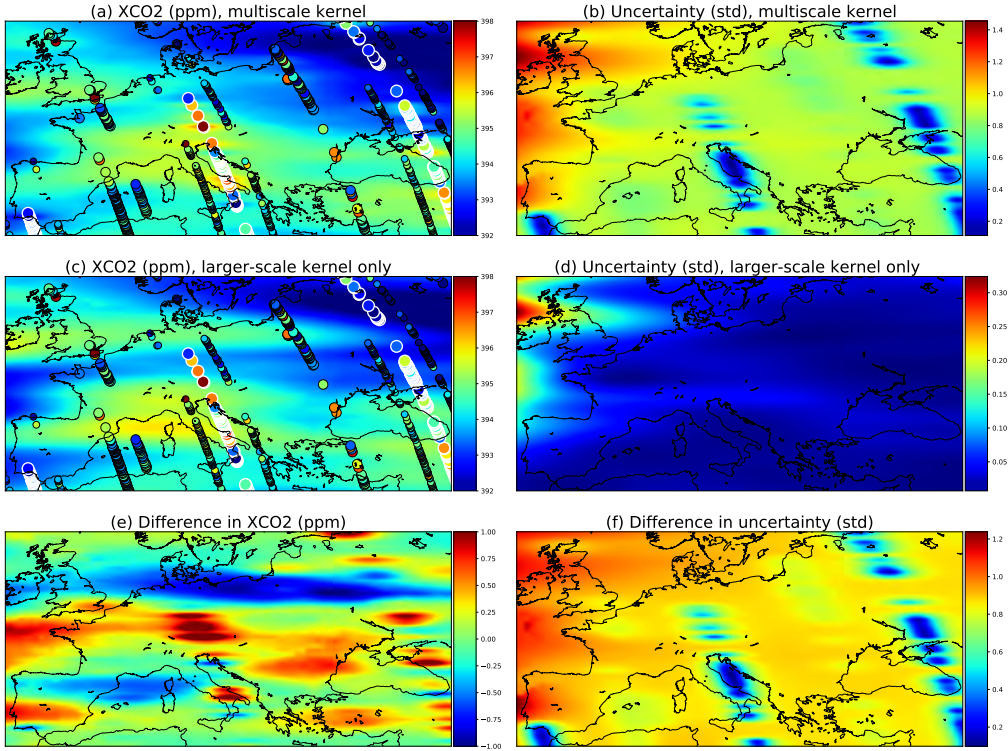


Figure 3.4: Comparison of a multi-scale kernel with two components with the parameters shown in Table 3.2, and a kernel containing only the exponential subkernel in Table 3.2. The observations used are shown in panels (a) and (c) as circles. The large ones with white borders are observations from the present day, September 15th 2014, medium circles are observations from 14th and 16th, and small circles from 13th and 17th.

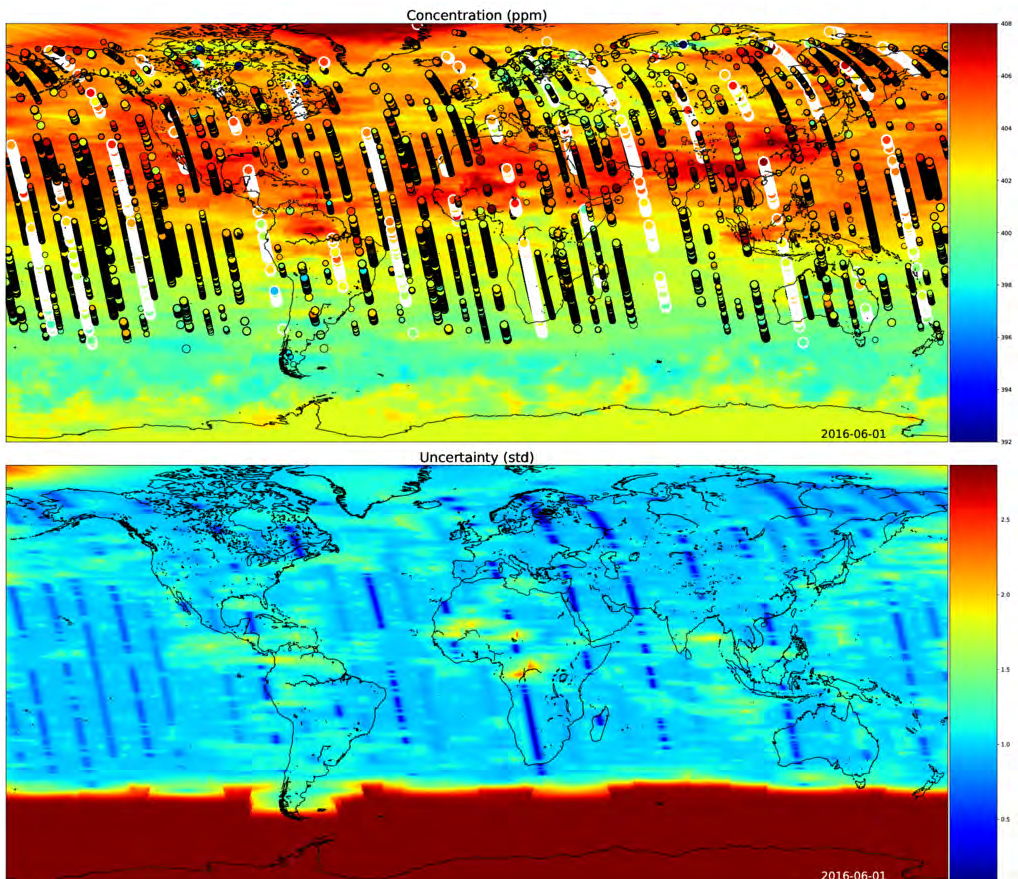
likely including the next observation is, and this probability depends on the distance to the previously added observation with $\Pr(\text{add } y|y_{\text{prev}}) = \frac{\|x - x_{\text{prev}}\|_2}{\omega \zeta_{\text{train}}}$, where ω is the grid resolution and x and x_{prev} denote, as earlier, the locations where observations y and y_{prev} were made. Such thinning discourages observations very close to each other from being included; for further details, see Paper I. Earlier, Tadić et al. (2017) has also used a distance-based probabilistic approach for observation selection, even though the inclusion probability is different. In both of the experiments $\omega = \frac{1}{2}$ was used and the exact same set of observations was utilized for calculating the marginals.

The figure clearly shows where present day observations are found as local enhancements. With the single subkernel with the larger length-scale parameters, the uncertainties are unreasonably low.

3.3.5 POSTERIOR XCO₂ FIELDS

A central reason for creating the Gaussian process software for remote sensing data is to be able to get better estimates of the spatio-temporal distributions of the quantity of interest with uncertainties. Figure 3.5 shows the means and uncertainties of the Gaussian process posterior calculated via (2.19) and (2.20) in a grid. The slight edginess far from observations, especially visible where the uncertain portion starts on the bottom of the lower part, is due to capping the search radius at 1100 km (10 equatorial degrees) in order to facilitate computation. In total 351 million marginals were computed with $\kappa = 256$ and using no data thinning, with parameter values from Table 3.2. The total number of observations used was 116 million.

Figure 3.5: Global GP posterior marginals with uncertainties on first of June 2016. In the summer months, the coverage of the satellite does not reach the South Pole due to lack of sunlight. The circles with the white edges are the current-day observations, the medium circles are observations from one day away, and the smallest circles are observations from two days away. Notice how the uncertainty increases from day to day due to the smaller kernel reducing local uncertainty less and less.



3.3.6 WIND-INFORMED KERNEL

One of the novel ideas in Paper I is the *wind-informed covariance kernel*, which rotates the covariance ellipse according to the wind axes. Given zonal and meridional wind vectors u and v , it is defined by parameters $\theta = (\tau, \ell, \rho, w^*)$. The kernel itself is an exponential kernel whose length-scale component to the direction of the wind, ℓ^\parallel , is scaled by $\sqrt{1 + |w^*| \rho}$, where w^* is the wind velocity vector at the test input x^* (listed above as a parameter since it does not depend on individual inputs x and x'). The additional parameter ρ determines how large a role the wind speed should play. The length scale parameter perpendicular to wind, ℓ^\perp , is not scaled, i.e. $\ell^\perp \leftarrow \ell$. Figure 3.6 shows equicovariance contours for various combinations of ρ and w^* .

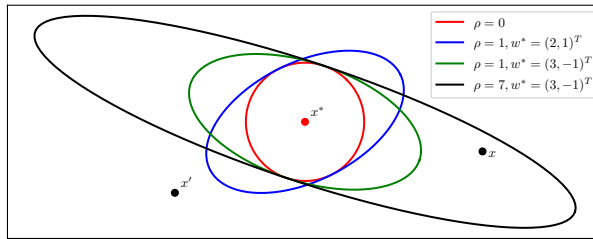


Figure 3.6: Equicovariance ellipses from the wind-informed kernel with various wind vectors w^* and values of ρ . The wind velocities are taken at the test input x^* but the covariance function k is of course evaluated also for each pair of observations x and x' .

The rationale behind the formulation of the wind-informed kernel is that e.g. trace gases are spread by winds and therefore the covariance direction should change according to wind direction. This subkernel type may also be combined with others in a multi-scale kernel.

The wind kernel parameters were calibrated by finding the medians of the approximate posterior calculated with the approximate marginal maximum likelihood method given by (3.2). The parameters found were $\tau = 2.07$, $\ell = 0.038$, and $\rho = 56.7$, and the values of $\kappa = 1024$ and grid resolution of 0.7° were used with the thinning parameter $\zeta_{\text{train}} = 1$ introducing some thinning. An example of the results is shown in figure 3.7, and as expected the uncertainty is clearly reduced where wind is blowing directly towards or away from the observations. The predicted mean of the concentration field is also spread due to the winds. The posterior marginal mean field looks less monotonous than the fields from fixed-direction kernels.

The wind-informed covariance kernel could be formulated in various ways and which formulation works best with what data still needs to be studied further. The winds used in figure 3.7 were processed from the local winds that come with the OCO-2 data. Obviously, winds derived from an actual wind data product would provide better accuracy, especially when the test inputs x^* are far from any observations.

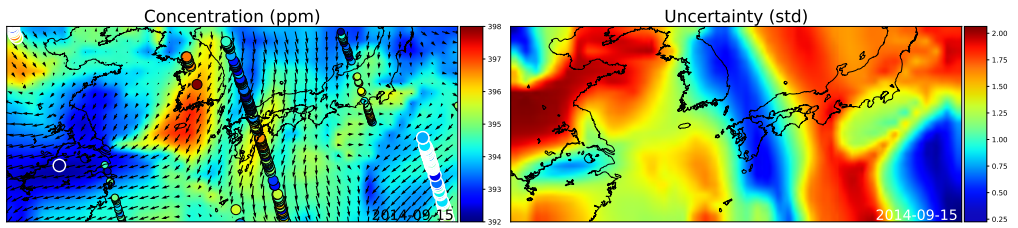


Figure 3.7: Japan, Korea, China: GP posterior marginal mean field of XCO₂ and the corresponding uncertainties produced with the wind-informed kernel. As before, circles with the white edges are present-day observations, medium ones are from adjacent days, and the smallest ones are from two days away. Wind direction and magnitude are given by the arrows.

3.4 BAYESIAN INFERENCE OF PHYSICS OF A BOREAL WETLAND WITH HIERARCHICAL MCMC

Boreal wetlands and peatlands are a major source of CH₄ emissions to the atmosphere, and it is likely that the magnitude of these emissions will grow as climate change progresses. In addition to CH₄, wetlands – in particular drained and managed wetlands – release and/or have the potential to release substantial amounts of CO₂. How substantial these emissions are and will be is not fully known, since peatland carbon emission estimates currently have high uncertainties (or uncertainties are not reported) and Bayesian analysis in the field of wetland emission modeling remains rare.

The research in Paper II and its objectives and results were briefly introduced in section 1 and 3.1.2, related work was mentioned in section 3.1.5, and the computational cost was discussed in section 3.2. The published literature pertaining to Bayesian modeling or model calibration in the context of wetland CH₄ emission models is covered in the introduction section of Paper II. This section describes the computational problem referencing section 2 and discusses some of the main results.

3.4.1 THE HIMMELI FORWARD MODEL

The wetland methane emission model HIMMELI², developed in collaboration between University of Helsinki and Finnish Meteorological Institute (Raivonen et al., 2017), is a 1-d partial differential equation model discretized by soil layers of variable thickness. In addition to CH₄, explicit formulations of CO₂ and O₂ are also included. The model contains processes for CH₄ production from root exudate decomposition and anaerobic peat decay. The transportation of the gases to the surface takes place in three ways: diffusion, transport via stems of aerenchymatous plants, and transportation

²HIMMELI stands for *Helsinki Model of MEthane buiLd-up and emIssion for peatlands*.

due to bubble formation, called *ebullition*.

Methane is produced predominantly when oxygen is not available and this is in HIMMELI controlled by the water table depth (WTD). The exudate input is provided as pre-calculated net primary production (NPP), fraction of which is passed on to the roots. The root depth distribution determines at which depth the exudates are deposited. If the water table level is above that deposition depth, methane may be produced.

The model version in Paper II, called sqHIMMELI, contains also the processes dealing with root exudates and peat decay, whereas in Raivonen et al. (2017) those processes are described as external functions for generating input. The 21 equations defining much of the sqHIMMELI model and the role of the model parameters are described in sections 3.4 and 3.5 of Paper II.

In addition to NPP and WTD, the model takes in soil temperature profiles and leaf area index (LAI) data, which broadly speaking tells how many layers of leaves in the canopy intercept solar radiation. The simulations and the study were performed utilizing measurement time series of the inputs and CO₂ and CH₄ fluxes from a research station in Hyytiälä, Southern Finland. Data from years 2005-2014 was used. For some input variables, filling gaps or other additional modeling were needed, see section 2 in Paper II.

3.4.2 BAYESIAN INFERENCE

The posterior distribution of the parameters controlling most parts of the model physics was computed with Monte Carlo methods. The posterior is a joint distribution of 14 parameters, which are presented in Table 3.3. The parameters partly affect the same processes, and all of the processes are coupled in the model code. For this reason, using samples from the posterior some correlations are to be expected in both the parameters and also between predicted quantities.

The Bayesian calibration was conducted via a hierarchical model described in section 2.7 and shown in figure 2.5. The parameters were divided into two sets: one where the parameters have a changing hyperprior, whose parameters have a fixed prior, and another where the parameters only have a fixed prior. The former are called here (and in Paper II) “hierarchical” and the latter “non-hierarchical” parameters, even though this terminology is not universal. The hierarchical parameters ζ_{exu} and Q_{10} varied from year to year, and their normal priors shared common hyperparameters, with $\frac{1}{\sigma^2} \sim \text{Scale-inv-}\chi^2(k, s)$ and $\mu \sim (\mu_0, \sigma_0^2)$, with fixed k , s , μ_0 and σ_0^2 . These parameters were sampled with Gibbs sampling (section 2.6.2), and the non-hierarchical parameters (see third column in table 3.3) were sampled with an Adaptive Metropolis step (section 2.6.1).

The sqHIMMELI model calculates CH₄ fluxes from the wetland given the model initial state, input data, and parameters. The observation operator is not well known, since even the footprint area of the measurements depends on time-varying external factors such as wind at the surface. Partly for this reason, a heavier tailed Laplace-

Table 3.3: Parameters examined in Paper II. The first column contains parameter symbols, second lists the primary process to which the parameter contributes, and the third lists whether the parameter was modeled in a hierarchical fashion or not. A short functional description of the parameters is given in the last column. The symbol “→” reads “decomposition into” and T stands for temperature. See also Table 3 in Paper II, which gives the prior limits, units, and references.

	Relevant to	Hier.	Parameter controls. . .
τ_{exu}	CH4 prod.	no	decay rate of exudates
ζ_{exu}	CH4 prod.	yes	fraction of NPP converted to exudates
τ_{cato}	CH4 prod.	no	rate of peat→CH4
Q_{10}	CH4 prod.	yes	dependence on T of peat→CH4
$f_{\text{exu}}^{\text{CH}_4}$	CH4 prod.	no	fraction of anaerobic peat→CH4
V_{R0}	Resp.	no	heterotrophic respiration rate
ΔE_R	Resp.	no	dependence of heterotrophic respiration on T
V_{O0}	CH4 oxid.	no	base rate of CH4 oxidation
ΔE_{oxid}	CH4 oxid.	no	dependence of CH4 oxidation on T
λ_{root}	Gas transport	no	root depth
ρ	Gas transport	no	root ending area per biomass
τ	Gas transport	no	root tortuosity parameter
$f_{D,a}$	Gas transport	no	diffusion rate in air-filled peat
$f_{D,d}$	Gas transport	no	diffusion rate in water-filled peat

distributed error model was used with the scaling of the error depending on the day of year, and for this heteroscedasticity model two additional parameters were fitted (see Appendix A of Paper II). The residuals were assumed to be correlated and their covariance structure was described with an ARMA(2,1) model, see section 2.8.1. The ARMA(2,1) parameters were learned as described in section 2.8.2 by minimizing the KL-divergence between the formal error model and the empirical distribution of the residuals. This was done after an initial, exploratory MCMC experiment was conducted to find an approximate posterior mean. The final posterior distribution was estimated using importance resampling, see section 2.6.3, with the exploratory posterior used as a biasing distribution.

3.4.3 RESULTS AND DISCUSSION

The setting presented in the previous section allows for lots of analysis. Figure 3.8 shows the output fluxes from the posterior mean parameter values, including credible intervals as shaded areas generated by random sampling the error model. The figure visually verifies that the calibrated model is able to produce fluxes that look realistic. The exudate pool sizes and the CH4 emissions closely follow the NPP input and the predictive credible intervals look reasonable.

The parameter posterior distribution shown in figure 3.9 contains various correlations reflecting interchangeability between the processes given the likelihood function

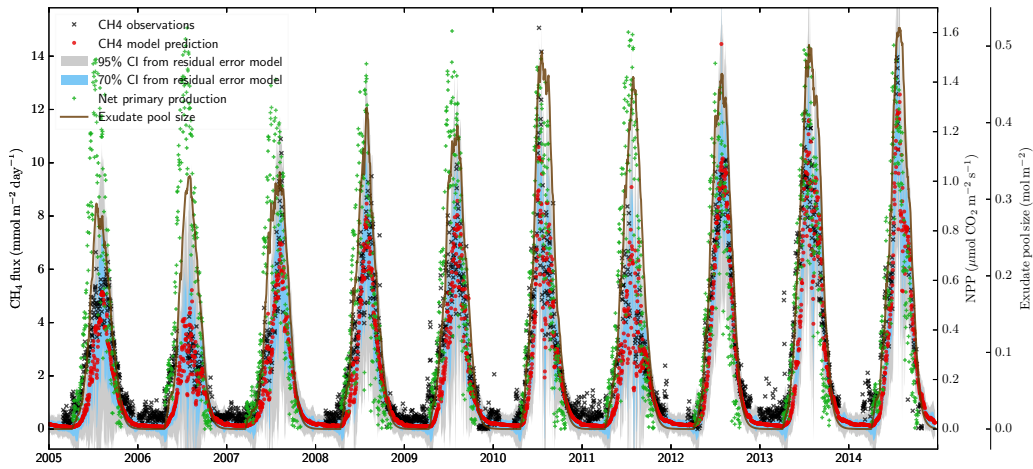


Figure 3.8: Output from the model with posterior mean parameter values. While the fit is good, the calibration is performed with the same observation dataset and therefore the residuals are only relevant for training error.

and the observed flux data. When the model is run with random samples drawn from the posterior distribution, correlations between the processes can be evaluated, as shown in figure 5 in Paper II for year 2012. That figure reveals that plant transport of CH₄ (via hollow stems) is driven by exudate decomposition, and that ebullition is in practice perfectly correlated with diffusion, raising the question of whether modeling ebullition is actually an unnecessary complication. With additional data, such as soil gas profiles, the processes might become better separated. Some of the correlations shown in figure 3.9 are strong, and they are rooted in the model equations, but often indirectly. These correlations are thoroughly discussed in sections 5.3 and 5.4 of Paper II.

For prediction, the hierarchical parameter calibration is of course not possible, and therefore other methods needed to be used for obtaining the Q_{10} and ζ_{exu} parameters for predictive purposes. Two schemes were used in Paper II: simply using the mean of the hierarchical parameters, and constructing a regression model for the ζ_{exu} and the Q_{10} parameters. The latter was performed by taking the posterior mean estimates for all the annually changing Q_{10} and ζ_{exu} parameters and then regressing those values against the mean soil temperature at 35 cm depth of the first 10 weeks of each year for Q_{10} , and against the NPP of 130 first days of each year for ζ_{exu} . The annual errors are shown in figure 3.10. In the figure plant transport is missing since it is the complement of diffusion. The term “all ebullition” refers to any ebullition that is released from the underwater part of the peat layer to air, and since water table is most of the time at least slightly under the surface, this is not a real flux, since the gas will be emitted to the atmosphere ultimately via diffusion in the air. On the right, the regression-based predictions are shown to not produce better annual predictions than

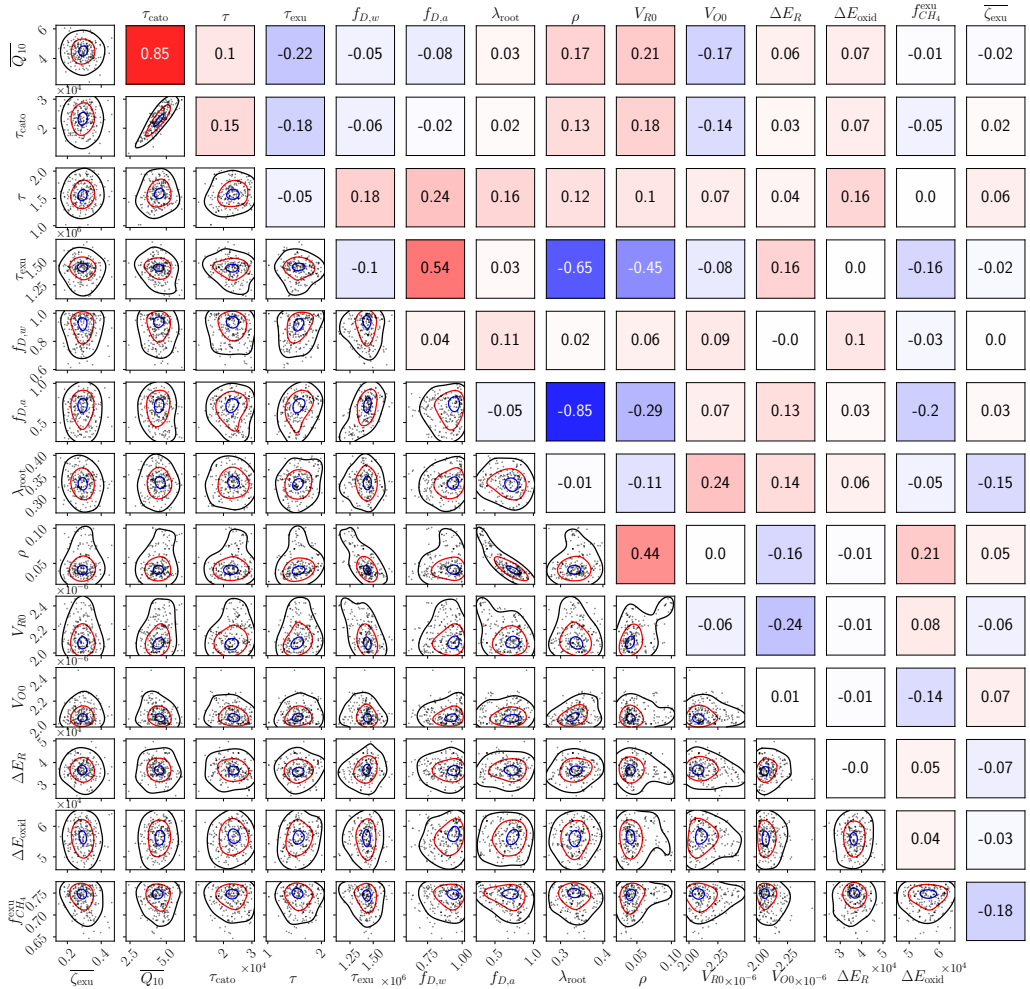


Figure 3.9: Lower triangle on the left: pairwise posterior marginal distributions between parameters, with labels on the left and bottom. The 10%, 50%, and 90% contours, calculated from a kernel density estimate, are shown. The upper right triangle shows pairwise correlations of the parameters with labels on the top and left. For the hierarchically modeled Q_{10} and ζ_{exu} parameters the distributions of the prior means are shown.

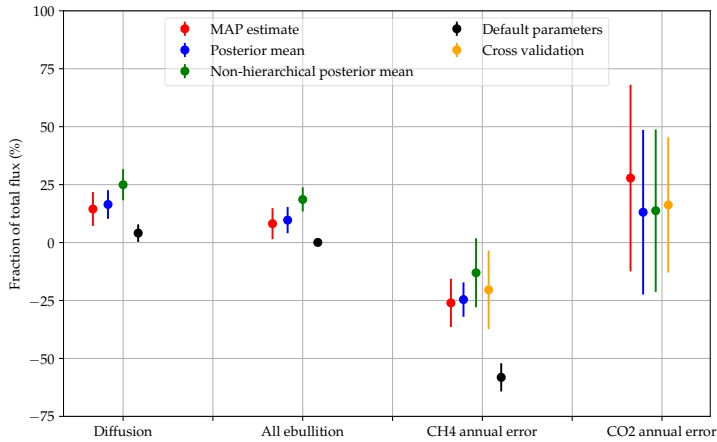


Figure 3.10: Transport component breakdown and annual errors. Plant transport is the complement of diffusion, and ebullition is effectively zero in all cases. For “All ebullition”, see text. Predictive results from cross validation are shown in orange in the error descriptions.

the non-hierarchically modeled parameters, implying that either assessing auxiliary performance metrics – such as using time intervals shorter than a year – is needed, or more complicated parametric models need to be constructed for modeling the time dependence of the parameters.

3.5 CLIMATE AND LAND SURFACE MODELING

Papers III and IV are different from Papers I and II in many respects, but most importantly they both utilize a significantly more complicated forward model, the JSBACH land surface scheme from the Max Planck Institute for Meteorology (MPI-M) in Hamburg. That model is a part of the ECHAM³ climate model and it describes processes interfacing the biosphere and the atmosphere. JSBACH makes independent predictions at each grid point based on external forcing. In Paper III that forcing comes from the atmospheric component of the ECHAM6 climate model, and in Paper IV from measured meteorological conditions at flux measurement sites.

This section describes briefly the research findings of Papers III and IV. The objectives and background for them were discussed in section 1, 3.1.3, and 3.1.4. The computational cost and observations used were described briefly in section 3.2.

³The JSBACH name creatively stands for Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg. Inspiration for such naming came from a previous model called MOZART. Furthermore, the EC in ECHAM stands for the European Center, from where initial code was adapted, while the HAM part of that name once again refers to Hamburg. For code availability, as of 2019, see mpimet.mpg.de/en/science/models/mpi-esm/jsbach.html.

3.5.1 THE ECHAM/JSBACH FORWARD MODEL

The JSBACH model is a complicated PDE model described partly in Roeckner et al. (2003a), more fully in the official model documentation available with the source code, and for relevant parts also in Appendix A of Mäkelä et al. (2016). The ECHAM model, which is used to produce the forcing data for JSBACH in Paper III, solves the atmospheric part including transport of species such as water vapor and trace gases, among everything else. It is a very complicated and heavily parametrized model, and its version 5 is described in the technical reports (Roeckner et al., 2003a,b). The performance of version 6 is further described in Stevens et al. (2013).

The JSBACH model describes different terrain types with *plant functional types*, which summarize the average physical functions of the different terrain types from glaciers to tropical rain forests. In Paper III, particularly changes in areas with the plant functional type *extratropical coniferous forest* were evaluated, and while that same type was used in Paper IV, there the associated parameters were adapted to the local conditions. The most important output variables for the purposes of the Papers were gross primary production, net primary production, evapotranspiration (ET), and snow coverage. All of these variables have to do with the carbon, water, or energy balance of the biosphere-atmosphere boundary.

3.5.2 PAPER III – CLIMATE CHANGE HAS SHIFTED THE GROWING SEASON

Paper III by Pulliainen et al. (2017) utilizes flux measurement data from the Boreal region, passive microwave retrievals of snow clearance date (SCD), modeling, and meteorological reanalysis data to evaluate how much earlier the starting date of spring recovery (SR) has shifted due to climate change, and how much that shift has affected the carbon balance in the first 180 days of the year. The result is that the onset of spring has become 0.23 days earlier each year, translating into an increase in the uptake of carbon of 52 megatons per decade.

The inference process to produce these estimates was the following: the passive microwave remote sensing data was used to retrieve snow clearance dates, and those data were used with in-situ flux measurements of CO₂ to learn the parameters of a regression model for predicting the timing of SR based on SCD. The ECHAM6-JSBACH model was used to calculate the GPP, and earlier SR was found to be weakly correlated with higher springtime GPP.

To produce reliable quantities with modeling, carbon pools in the model were spun up with a 2000-year initial simulation with a lightweight model, CBALANCE, after which a hydrology spin-up was performed using ECHAM with no outside forcing from year 1870 up until 1958. From 1959 onwards the ERA-interim reanalysis dataset was used to nudge the model to keep the meteorology close to the observed, and starting 1979 the ERA-40 dataset was used for that same purpose. The GPP/SR trends were calculated for each grid point from the 36-year period of 1979-2014.

In addition to the global results quoted above, the combination of modeling with

flux measurements allowed looking at the changes regionally. It appears that in Eurasia the change in springtime GPP per decade was proportionally higher (6.8%) than in North America (5.5%). Similarly, the shift of the starting date of spring recovery is also larger in Eurasia, where this figure is a remarkable 3.0 days per decade, while in North America the shift is smaller but still sizable at 1.3 days per decade.

3.5.3 PAPER IV – CONSTRAINING LSS PARAMETERS WITH FLUX DATA WITH ADAPTIVE MCMC

In the last included work, Paper IV, parameters of the JSBACH land surface model were calibrated using the Adaptive Metropolis MCMC algorithm. This work has been introduced in section 1, 3.1.4, and 3.2. Markov chain Monte Carlo was described in section 2.6.1.

The work in Mäkelä et al. (2016) utilizes flux data from two measurement sites. The first of these is in Hyytiälä (61°51'N, 24°17'E), and the second one is in Sodankylä (67°22'N, 26°38'E). These sites are long-running measurement sites where the predominant tree species is the Scots pine (*Pinus Sylvesteris*). For Hyytiälä, half-hourly measurements of CO₂ and H₂O fluxes were used from 1999-2008, while for Sodankylä, the time period was 2000-2008. The JSBACH model calibration used the Hyytiälä data from 2000-2004, whereas for generating the initial conditions for the model the year 1999 was used. For Sodankylä, this spin-up was done with data from all the years and no calibration was performed, instead reusing the data for validation. The aim of the spin-up process was to stabilize the fast carbon pools and the water pools so that local conditions would be represented in initial states of the model.

Since the objective of the study was to improve and better understand how the gas exchange processes in the model are able to describe conditions at these particular sites, the parameters chosen for the calibration were related to gas exchange. These 15 parameters are described in Table 1 of Paper IV. The parameters were calibrated using three different loss functions: one with seasonally averaged data, another one with daily averaged data, and the third one with the original half-hourly data. Three of the parameters were only calibrated with the first one of these.

Even though MCMC usually gives a statistically meaningful posterior distribution, in this work rigorous uncertainty quantification was not attempted as the distributions of the model-observation residuals were not carefully analyzed. The cost functions used were of the standard quadratic form corresponding to a Gaussian observation model

$$\mathcal{L}(\theta) = \sum_i (x - y)^T \Gamma^{-1} (x - y), \quad (3.5)$$

where θ is the model parameter vector, the model output x depends on θ , and the sum is over the (potentially averaged) observations. For the calibration with seasonally averaged data, the vectors x and y contained residuals of mean GPP, mean ET, and maximum LAI, and the diagonal Γ matrix contained, for each period, means

of the observed GPP and ET squared and maximum of the observed LAI squared. For daily and half-hourly calibration LAI was not used, and the elements of Γ were further multiplied with the square root of the number of corresponding observations, inflating the size of the posterior.

A principal component analysis of the MCMC chains revealed that estimates of two parameters controlling bare-soil evaporation – soil dryness-based relative humidity and skin reservoir field capacity (how much water can be held at the very top of the vegetation in a layer of some millimeters) – are in this calibration the least reliable ones. Using the posterior mean values from the MCMC run of the calibration period for Hyytiälä, model performance as measured by (3.5) improved for all the validation runs with the exception that the seasonal calibration in Hyytiälä lead to degraded performance as measured by the daily and half-hourly cost function values. For the Sodankylä site, performance improved with all calibration methods and all metrics when compared to the default parameter values, implying that parameter calibration is generalizable from one site to a similar site at a different location.

The calibrated model was not able to describe a rare drought event in 2006 in Hyytiälä (GPP drop in August 2006 in figure 2 of Paper IV). However, since there were no extended dry periods in the calibration data, the failure of the calibrated model to accommodate for this anomaly was not unexpected.

4 CONCLUSIONS AND FUTURE WORK

The methods presented in section 2 represent a small and relatively simple subset of the very large number of techniques nowadays used for uncertainty quantification and data science. Similarly, the context provided by climate change, and more generally geosciences, is huge, and therefore this work scratches only a corner or two of an immense problem space. In this sense it is fortunate that the mathematical theory is agnostic to the applications and the methods and algorithms can easily be reused.

Each of the Papers presented contained three building blocks: models, data, and algorithms. These building blocks were together used to answer specific climate change-related research questions: statistical models marry process models and observational data, and carefully analyzing the different aspects of the model-observation mismatch enabled the utilization of Bayes' theorem for solving inverse problems, either with Monte Carlo methods or via point estimation.

While models and data were used in all the Papers, only the first two utilized non-trivial statistical estimation techniques to try to understand the statistical properties of the data. Still, even in those two publications, much room was left for further analysis, and in Papers III and IV the price for omitting Bayesian uncertainty quantification was that the posterior and posterior predictive uncertainties remained unknown. On one hand this lack of uncertainty quantification adversely affects how actionable the results are, but on the other hand when expensive computational models are used, conducting Bayesian analysis is often impossible. This was in particular the case with Paper III.

Certain themes recur when evaluating how the research could have been improved. When model-generated input data – for instance wind data in Paper I or leaf area index, net primary production, and water table depth data in Paper II – were used, the propagation of uncertainties pertaining to those quantities were overlooked. While disregarding uncertainties in input data is often necessary, the implications of that are that uncertainty estimates from settings involving modeled input data and complex models need to be approached with caution. The flip side of the coin is that even when all modeling is perfect, the results of any inference are only as good as the data that is used. This was most evident in Paper I, where the quality of the uncertainty information provided with the XCO₂ observations was not always reliable.

The work in the Papers may be critiqued in more specific ways to guide future research endeavors. In Paper I the covariance between measurement errors of the

individual measurements are not known, and neither are the various biases that are known to exist in the data. Regarding satGP, there is room for development in how observations for each subkernel are selected, and the effects of this still need to be analyzed and minimized. The β coefficient fields with their uncertainties may provide further useful information that can be used to devise better formulations of the mean function. Other possible next steps include applying the satGP software to other problems, combining multiple data products, performing model selection to select the best combinations of subkernels for the multi-scale kernels, and general code development and usability enhancements.

The most pressing issue in Paper II is the lack of uncertainty quantification for input data generation. Following that, the error modeling can be further enhanced by treating the instrument error and other error sources separately in the observation equation, potentially yielding improved models for describing the data. Cross validation at other measurement sites and computing regional fluxes with uncertainties would be valuable, both in terms of the actual results and in terms of learning how well the modeled processes actually describe what they are intended to describe.

The regression plots in Paper III show large deviations, which tend to disproportionately affect the trends when Gaussian errors are assumed (e.g. figure 4 in Paper III). Furthermore, while Paper III includes uncertainties in the presentation of the springtime GPP increase due to changes in the spring recovery date (Table 1 in Paper III), those trends were calculated using data from only two measurement stations in both Eurasia and North America, and this may lead to increased representation error.

The *ad hoc* nature of the cost function formulations in Paper IV rules out proper uncertainty quantification, and with it e.g. the possibility to compute Monte Carlo estimates of future carbon balance based on parameter posteriors. The differences in the optimal parameter values between the different loss function formulations shows how important data selection and averaging are, and points out that the design of any model calibration exercise must be based on future modeling needs. The incapability of the model to describe the dry event in the summer of 2006 suggests that process modifications need to be carried out. That this work was undertaken in Mäkelä et al. (2019) (see section 3.1.5) serves as an example of how process models may and should be improved based on statistical analyses.

When research is constrained by the availability and quality of observations, collecting more data and refining the analyses little by little provides more and more confidence in the conclusions. This is what the IPCC reports describe, with each new version having more weight and urgency in both the details and the overall message.

The research presented in this thesis consists of technical results related to climate change, carbon cycle, models, and data. These technicalities, however, hide an important aspect of the work, which is to underline that climate change has already advanced very far (Papers I,III), and this results in unpredictable and difficult-to-model phenomena (Papers II-IV). For these reasons, action needs to be taken to address the problems reported by the scientific community in addition to performing

and funding more research. While a scientist can use Bayesian analysis to improve the models, that same analysis can also be used by policy makers and voters as a small ingredient in cooking up a way to save the world from the most catastrophic climate change scenarios.

REFERENCES

- P. Bickel and K. Doksum. *Mathematical Statistics 2e*, volume 1. CRC Press, 1st edition, 2015. ISBN 9781498723800.
- P. Bickel and K. Doksum. *Mathematical Statistics 2e*, volume 1. CRC Press, 2nd edition, 2016. ISBN 9781498722681.
- A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. The Bouncy Particle Sampler: A Non-Reversible Rejection-Free Markov Chain Monte Carlo Method. *arXiv e-prints*, art. arXiv:1510.02451, Oct 2015.
- S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004. ISBN 0521833787.
- G. Casella and R. Berger. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002. ISBN 9780534243128.
- C. Chatfield. *The analysis of time series: an introduction*. Chapman and Hall, 4th edition, 1989.
- P. G. Constantine, C. Kent, and T. Bui-Thanh. Accelerating MCMC with active subspaces. *arXiv e-prints*, art. arXiv:1510.00024, Sep 2015.
- D. Crisp, B. M. Fisher, C. O'Dell, C. Frankenberg, R. Basilio, H. Bösch, 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. Natraj, 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 – ndash; Part II: Global XCO₂ data characterization. *Atmospheric Measurement Techniques*, 5(4):687–707, 2012. doi: 10.5194/amt-5-687-2012. URL <https://www.atmos-meas-tech.net/5/687/2012/>.
- 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, oct 2014. doi: 10.1088/0266-5611/30/11/114015. URL <https://doi.org/10.1088%2F0266-5611%2F30%2F11%2F114015>.

- F.-X. L. Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, 38A(2):97–110, 1986. doi: 10.1111/j.1600-0870.1986.tb00459.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.1986.tb00459.x>.
- S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216 – 222, 1987. ISSN 0370-2693. doi: [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X). URL <http://www.sciencedirect.com/science/article/pii/037026938791197X>.
- J. Durbin and S. Koopman. *Time Series Analysis by State Space Methods: second Edition*. Oxford Statistical Science Series. OUP Oxford, 2012. ISBN 9780199641178.
- J. Durbin and G. Watson. Testing for serial correlation in least-squares regression, I. *Biometrika*, 37:409–428, 1950.
- J. Durbin and G. Watson. Testing for serial correlation in least-squares regression, II. *Biometrika*, 38:159–178, 1951.
- D. Gamerman. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 1997. ISBN 9780412818202.
- A. L. Ganesan, M. Rigby, A. Zammit-Mangion, A. J. Manning, R. G. Prinn, P. J. Fraser, C. M. Harth, K.-R. Kim, P. B. Krummel, S. Li, J. Mühle, S. J. O’Doherty, S. Park, P. K. Salameh, L. P. Steele, and R. F. Weiss. Characterization of uncertainties in atmospheric trace gas inversions using hierarchical bayesian methods. *Atmospheric Chemistry and Physics*, 14(8):3855–3864, 2014. doi: 10.5194/acp-14-3855-2014. URL <https://www.atmos-chem-phys.net/14/3855/2014/>.
- A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition, 2013.
- I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. *arXiv e-prints*, art. arXiv:1406.2661, Jun 2014.
- U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(4):549–603, 1994. ISSN 00359246. URL <http://www.jstor.org/stable/2346184>.
- M. Gruber. *Matrix Algebra for Linear Models*. Wiley, 2013. ISBN 9781118608814.
- H. Haario, E. Saksman, and J. Tamminen. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

- H. Haario, M. Laine, A. Mira, and E. Saksman. Dram: Efficient adaptive mcmc. *Statistics and Computing*, 16(4):339–354, 2006. ISSN 1573-1375. doi: 10.1007/s11222-006-9438-0. URL <http://dx.doi.org/10.1007/s11222-006-9438-0>.
- D. M. Hammerling, A. M. Michalak, and S. R. Kawa. Mapping of CO₂ at high spatiotemporal resolution using satellite observations: Global distributions from OCO-2. *Journal of Geophysical Research: Atmospheres*, 117(D6), 2012a. doi: 10.1029/2011JD017015. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD017015>.
- D. M. Hammerling, A. M. Michalak, C. O'Dell, and S. R. Kawa. Global CO₂ distributions over land from the Greenhouse Gases Observing Satellite (GOSAT). *Geophysical Research Letters*, 39(8), 2012b. doi: 10.1029/2012GL051203. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL051203>.
- J. Hammersley and P. Clifford. Markov random fields on finite graphs and lattices. 1971.
- A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1990. doi: 10.1017/CBO9781107049994.
- M. J. Heaton, A. Datta, A. Finley, R. Furrer, R. Guhaniyogi, F. Gerber, R. B. Gramacy, D. Hammerling, M. Katzfuss, F. Lindgren, D. W. Nychka, F. Sun, and A. Zammit-Mangion. A Case Study Competition Among Methods for Analyzing Large Spatial Data. *arXiv e-prints*, art. arXiv:1710.05013, Oct 2017.
- IPCC. *Summary for Policymakers*, book section SPM, pages 1–30. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013. ISBN ISBN 978-1-107-66182-0. doi: 10.1017/CBO9781107415324.004. URL www.climatechange2013.org.
- R. Kalman. A new approach to linear filtering and prediction problems. *Transactions of ASME – Journal of Basic Engineering*, 82:35–45, 1960.
- I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer Graduate Texts in Mathematics. Springer-Verlag, "2nd" edition, 1998. doi: 10.1007/978-1-4684-0302-2.
- M. Katzfuss, J. Guinness, and W. Gong. Vecchia approximations of Gaussian-process predictions. *arXiv e-prints*, art. arXiv:1805.03309, May 2018.
- M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000. ISSN 00063444. URL <http://www.jstor.org/stable/2673557>.

- M. Laine, N. Latva-Pukkila, and E. Kyrölä. Analysing time-varying trends in stratospheric ozone time series using the state space approach. *Atmospheric Chemistry and Physics*, 14(18):9707–9725, 2014. doi: 10.5194/acp-14-9707-2014. URL <https://www.atmos-chem-phys.net/14/9707/2014/>.
- M. Lassas and S. Siltanen. Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Problems*, 20:1537, 08 2004. doi: 10.1088/0266-5611/20/5/013.
- S. Lauritzen. *Graphical Models*. Oxford Statistical Science Series. Clarendon Press, 1996. ISBN 9780191591228.
- K. Law, A. Stuart, and K. Zygalakis. *Data Assimilation*. Springer International Publishing, 2015. ISBN 9783319203249.
- F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011. doi: 10.1111/j.1467-9868.2011.00777.x. URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x>.
- D. Lu, D. Ricciuto, A. Walker, C. Safta, and W. Munger. Bayesian calibration of terrestrial ecosystem models: a study of advanced markov chain monte carlo methods. *Biogeosciences*, 14(18):4295–4314, 2017. doi: 10.5194/bg-14-4295-2017. URL <https://www.biogeosciences.net/14/4295/2017/>.
- J. Mäkelä, J. Susiluoto, T. Markkanen, M. Aurela, H. Järvinen, I. Mammarella, S. Hagemann, and T. Aalto. Constraining ecosystem model with adaptive Metropolis algorithm using boreal forest site eddy covariance measurements. *Nonlinear Processes in Geophysics*, 23(6):447–465, 2016. doi: 10.5194/npg-23-447-2016. URL <http://www.nonlin-processes-geophys.net/23/447/2016/>.
- J. Mäkelä, J. Knauer, M. Aurela, A. Black, M. Heimann, H. Kobayashi, A. Lohila, I. Mammarella, H. Margolis, T. Markkanen, J. Susiluoto, T. Thum, T. Viskari, S. Zaehle, and T. Aalto. Parameter calibration and stomatal conductance formulation comparison for boreal forests with adaptive population importance sampler in the land surface model jsbach. *Geoscientific Model Development*, 12(9):4075–4098, 2019. doi: 10.5194/gmd-12-4075-2019. URL <https://www.geosci-model-dev.net/12/4075/2019/>.
- J. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012. doi: 10.1137/1.9781611972344. URL <https://epubs.siam.org/doi/abs/10.1137/1.9781611972344>.

- J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 01 1965. ISSN 0010-4620. doi: 10.1093/comjnl/7.4.308. URL <https://doi.org/10.1093/comjnl/7.4.308>.
- H. Nguyen, M. Katzfuss, N. Cressie, and A. Braverman. Spatio-temporal data fusion for very large remote sensing datasets. *Technometrics*, 56(2):174–185, 2014. doi: 10.1080/00401706.2013.831774. URL <https://doi.org/10.1080/00401706.2013.831774>.
- J. Nocedal. Updating Quasi-Newton matrices with limited storage. *Math. Comput.*, 35(151):773–782, 07 1980.
- C. W. O'Dell, B. Connor, H. Bösch, D. O'Brien, C. Frankenberg, R. Castano, M. Christi, D. Crisp, A. 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. Corrigendum to "The ACOS CO₂ retrieval algorithm – Part 1: Description and validation against synthetic observations" published in *atmos. meas. tech.*, 5, 99–121, 2012. *Atmospheric Measurement Techniques*, 5(1):193–193, 2012. doi: 10.5194/amt-5-193-2012. URL <https://www.atmos-meas-tech.net/5/193/2012/>.
- B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer Berlin Heidelberg, 2010. ISBN 9783642143946.
- B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 60(3): 550–591, 2018. doi: 10.1137/16M1082469. URL <https://doi.org/10.1137/16M1082469>.
- G. Peyré and M. Cuturi. Computational Optimal Transport. *arXiv e-prints*, art. arXiv:1803.00567, Mar 2018.
- M. Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. Report, Department of Applied Mathematics and Theoretical Physics, Cambridge, UK, 08 2009.
- J. Pulliainen, M. Aurela, T. Laurila, T. Aalto, M. Takala, M. Salminen, M. Kulmala, A. Barr, M. Heimann, A. Lindroth, A. Laaksonen, C. Derksen, A. Mäkelä, T. Markkanen, J. Lemmetyinen, J. Susiluoto, S. Dengel, I. Mammarella, J.-P. Tuovinen, and T. Vesala. Early snowmelt significantly enhances boreal spring-time carbon uptake. *Proceedings of the National Academy of Sciences*, 114(42): 11081–11086, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1707889114. URL <http://www.pnas.org/content/114/42/11081>.
- M. Raivonen, S. Smolander, L. Backman, J. Susiluoto, T. Aalto, T. Markkanen, J. Mäkelä, J. Rinne, O. Peltola, M. Aurela, A. Lohila, M. Tomasic, X. Li, T. Larmola, S. Juutinen, E.-S. Tuittila, M. Heimann, S. Sevanto, T. Kleinen, V. Brovkin,

- and T. Vesala. HIMMELI v1.0: Helsinki Model of MEthane buiLd-up and emis-
sion for peatlands. *Geoscientific Model Development*, 10(12):4665–4691, 2017.
doi: 10.5194/gmd-10-4665-2017. URL <https://www.geosci-model-dev.net/10/4665/2017/>.
- C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. Adaptive
computation and machine learning series. University Press Group Limited, 2006.
ISBN 9780262182539.
- A. D. Richardson, D. Y. Hollinger, G. G. Burba, K. J. Davis, L. B. Flanagan, G. G.
Katul, J. W. Munger, D. M. Ricciuto, P. C. Stoy, A. E. Suyker, S. B. Verma,
and S. C. Wofsy. A multi-site analysis of random error in tower-based measure-
ments of carbon and energy fluxes. *Agricultural and Forest Meteorology*, 136(1–2):
1 – 18, 2006. ISSN 0168-1923. doi: <http://dx.doi.org/10.1016/j.agrformet.2006.01.007>. URL <http://www.sciencedirect.com/science/article/pii/S0168192306000281>.
- G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scal-
ing of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120,
02 1997. doi: 10.1214/aoap/1034625254. URL <http://dx.doi.org/10.1214/aoap/1034625254>.
- E. Roeckner, G. Bäuml, L. Bonaventura, R. Brokopf, M. Esch, M. Giorgetta,
S. Hagemann, I. Kirchner, L. Kornblueh, E. Manzini, A. Rhodin, U. Schlese,
U. Schulzweida, and A. Tompkins. The atmospheric general circulation model
ECHAM5. Part i: Model description. Report 349, Max-Planck-Institut für Meteo-
rologie, Hamburg, 2003a.
- E. Roeckner, G. Bäuml, L. Bonaventura, R. Brokopf, M. Esch, M. Giorgetta,
S. Hagemann, I. Kirchner, L. Kornblueh, E. Manzini, A. Rhodin, U. Schlese,
U. Schulzweida, and A. Tompkins. The atmospheric general circulation model
ECHAM5. Part ii: Simulated climatology and comparison with observations. Re-
port 354, Max-Planck-Institut für Meteorologie, Hamburg, 2003b.
- L. Roininen, S. Lasanen, M. Orispää, and S. Särkkä. Sparse approximations of frac-
tional Matérn fields. *Scandinavian Journal of Statistics*, 45(1):194–216, 2018. doi:
10.1111/sjos.12297. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12297>.
- Y. Rozanov. *Random Fields and Stochastic Partial Differential Equations*. Kluwer
Academic Publishers, 1998. ISBN 0792349849.
- W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1987. ISBN 0070542341.
- T. P. Runarsson and Xin Yao. Search biases in constrained evolutionary optimization.
*IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and
Reviews)*, 35(2):233–243, May 2005. doi: 10.1109/TSMCC.2004.841906.

- T. Santner, B. Williams, and W. Notz. *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer Verlag New York, 2003. ISBN 0387954201.
- S. Särkkä. *Bayesian Filtering and Smoothing*. Bayesian Filtering and Smoothing. Cambridge University Press, 2013. ISBN 9781107030657.
- B. Stevens, M. Giorgetta, M. Esch, T. Mauritsen, T. Crueger, S. Rast, M. Salzmann, H. Schmidt, J. Bader, K. Block, R. Brokopf, I. Fast, S. Kinne, L. Kornblueh, U. Lohmann, R. Pincus, T. Reichler, and E. Roeckner. Atmospheric component of the MPI-M Earth system model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, 5(2):146–172, 2013. doi: 10.1002/jame.20015. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jame.20015>.
- D. Stroock. *Elements of Stochastic Calculus and Analysis*. CRM Short Courses. Springer International Publishing, 2018. ISBN 9783319770383.
- A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010. doi: 10.1017/S0962492910000061.
- J. Susiluoto, A. Spantini, H. Haario, and Y. Marzouk. Efficient multi-scale gaussian process regression for massive remote sensing data with satgp v0.1. *Geoscientific Model Development Discussions*, 2019:1–30, 2019. doi: 10.5194/gmd-2019-156. URL <https://www.geosci-model-dev-discuss.net/gmd-2019-156/>.
- J. M. Tadić, X. Qiu, S. Miller, and A. M. Michalak. Spatio-temporal approach to moving window block kriging of satellite data v1.0. *Geoscientific Model Development*, 10(2):709–720, 2017. doi: 10.5194/gmd-10-709-2017. URL <https://www.geosci-model-dev.net/10/709/2017/>.
- A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 2005. ISBN 9780898715729.
- R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 00359246. URL <http://www.jstor.org/stable/2346178>.
- L. Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4): 1701–1728, 12 1994. doi: 10.1214/aos/1176325750. URL <https://doi.org/10.1214/aos/1176325750>.
- L. Tierney and A. Mira. Some adaptive monte carlo methods for bayesian inference. *Statistics in Medicine*, 18(17-18):2507–2515, 1999. doi: 10.1002/(SICI)1097-0258(19990915/30)18:17/18<2507::AID-SIM272>3.0.CO;2-J. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/>

- M. K. Titsias and O. Papaspiliopoulos. Auxiliary gradient-based sampling algorithms. *arXiv e-prints*, art. arXiv:1610.09641, Oct 2016.
- L. Trefethen and D. Bau. *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1997. ISBN 9780898719574.
- A. V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(2):297–312, 1988. ISSN 00359246. URL <http://www.jstor.org/stable/2345768>.
- R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi: 10.1017/9781108231596.
- M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008. doi: 10.1561/2200000001. URL http://www.eecs.berkeley.edu/~wainwrig/Papers/WaiJor08_FTML.pdf.
- D. Williams. *Probability with Martingales*. Cambridge University Press, 1991. ISBN 9780521406055.
- A. Zammit-Mangion, N. Cressie, A. L. Ganesan, S. O’ Doherty, and A. J. Manning. Spatio-temporal bivariate statistical models for atmospheric trace-gas inversion. *ArXiv e-prints*, Sept. 2015.
- A. Zammit-Mangion, N. Cressie, and C. Shumack. On statistical approaches to generate level 3 products from satellite remote sensing retrievals. *Remote Sensing*, 10(1), 2018. ISSN 2072-4292. doi: 10.3390/rs10010155. URL <http://www.mdpi.com/2072-4292/10/1/155>.
- Z. Zeng, L. Lei, L. Guo, L. Zhang, and B. Zhang. Incorporating temporal variability to improve geostatistical analysis of satellite-observed CO₂ in China. *Chinese Science Bulletin*, 58(16):1948–1954, Jun 2013. ISSN 1861-9541. doi: 10.1007/s11434-012-5652-7. URL <https://doi.org/10.1007/s11434-012-5652-7>.
- Z.-C. Zeng, L. Lei, K. Strong, D. B. A. Jones, L. Guo, M. Liu, F. Deng, N. M. Deutscher, M. K. Dubey, D. W. T. Griffith, F. Hase, B. Henderson, R. Kivi, R. Lindenmaier, I. Morino, J. Notholt, H. Ohyama, C. Petri, R. Sussmann, V. A. Velasco, P. O. Wennberg, and H. Lin. Global land mapping of satellite-observed CO₂ total columns using spatio-temporal geostatistics. *International Journal of Digital Earth*, 10(4):426–456, 2017. doi: 10.1080/17538947.2016.1156777. URL <https://doi.org/10.1080/17538947.2016.1156777>.

PAPER I

Jouni Susiluoto, Alessio Spantini, Heikki Haario, Youssef Marzouk

Efficient multi-scale Gaussian process regression for massive remote sensing data with satGP v0.1

In *Geoscientific Model Development Discussions*, 2019.

<https://doi.org/10.5194/gmd-2019-156>

Copyright © Authors 2018. CC Attribution 3.0 License. Reprinted with permission.



Efficient multi-scale Gaussian process regression for massive remote sensing data with satGP v0.1

Jouni Susiluoto^{1,2,3}, Alessio Spantini¹, Heikki Haario^{2,3}, and Youssef Marzouk¹

¹Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 77 Massachusetts Avenue, 33-207, Cambridge MA 02139 USA

²Lappeenranta University of Technology, School of Engineering Science, P.O. Box 20, FI-53851 Lappeenranta, Finland

³Finnish Meteorological Institute, Erik Palménin aukio 1, FI-00560 Helsinki, Finland

Correspondence: Jouni Susiluoto (jsusiluo@mit.edu)

Abstract. Satellite remote sensing provides a global view to processes on Earth that has unique benefits compared to measurements made on the ground. The global coverage and the enormous amounts of data produced come, however, with the price of spatial and temporal gaps and less than perfect data quality. Meaningful statistical inference from such data requires overcoming these problems and that calls for developing efficient computational tools.

5 We design and implement a computationally efficient multi-scale Gaussian process (GP) software package, satGP, geared towards remote sensing applications. The software is designed to be able to handle problems of enormous sizes and is able to compute marginals and sample from a random process with at least over hundred million observations.

The mean function of the Gaussian process is described by approximating marginals of a Markov random field (MRF). For covariance functions, Matern, exponential, and periodic kernels are utilized in a multi-scale kernel setting to describe the spatial
10 heterogeneity present in data. We further demonstrate how winds can be used to inform the covariance kernel formulation. The covariance kernel parameters are learned by calculating an approximate marginal maximum likelihood estimate and this is utilized to verify the validity of the multi-scale approach in synthetic experiments.

For demonstrating the techniques above, data from the Orbiting Carbon Observatory 2 (OCO-2) satellite is used. The satGP program is released as open source software.

15

1 Introduction

Climate change is one of the most important current global environmental challenges, to the point where it is drawing constant widespread attention even in mainstream media. The underlying reason is the anthropogenic carbon emissions: among the well-mixed greenhouse gases, carbon dioxide (CO₂) has currently the strongest effect on warming the planet, with the radiative
20 forcing of ca. 1.68 W m⁻² according to the latest IPCC report (IPCC, 2013).



The resulting global interest in atmospheric carbon along with technological advances has resulted in several CO₂-measuring satellites continuously monitoring the Earth and producing enormous quantities of data, which are processed to local estimates of CO₂ by solving a complicated inverse problem (Crisp et al., 2012). These include the Greenhouse gases Observing SATellite (GOSAT) from Japan (Yokota et al., 2009), which has been operational since January 2009, the OCO-2 from NASA, launched 5 in July 2014, and the Chinese TanSat (Yi et al., 2018), which was launched in December 2016. October 2018 saw the launch of GOSAT-2, and in May 2019 the OCO-3 instrument was taken to the International Space Station. In addition to the CO₂-measuring instruments, also other types of data are produced by remote sensing. For instance the European TROPOspheric Monitoring Instrument (TROPOMI) produces measurements of nitrogen dioxide, formaldehyde, carbon monoxide, aerosols, methane, and ozone.

10 Common denominators among most non-gridded remote sensing data sets include: a large number of observations, global coverage but small area observed at any given time, sensitivity to prevailing weather conditions and cloud cover, unknown and/or unreported error covariances, and predetermined positioning that rules out freely observing at a given time and location. These shortcomings can be partly remedied with computational statistics. The many steps of producing carbon flux estimates from readings produced by satellites are summarized by e.g. Cressie (2018). In this work a tool to solve one of those steps, the 15 production of gridded level 3 data sets with uncertainties from pointwise level 2 column integrated dry air CO₂ mole fraction (XCO₂) data, is introduced. Even though we demonstrate the capabilities of the software with OCO-2 data, the methods are not constrained by the quantity of interest observed.

The purpose of this manuscript is four-fold. First, to introduce satGP, a fast computer program that estimates Gaussian process covariance and mean function parameters from data, computes posterior marginal distributions, and samples from GP 20 priors and posteriors conditioning on over hundred million observations in situations where several hundred million marginals need to be computed. While lots of advances have recently been made in the field, we are not aware of any literature or software solving problems of quite this scale so far. Second, computational methods that allow the solution of problems of such scales are introduced. Third, covariance function and mean function formulations, some of which we have not seen used in the remote sensing community, are presented. In particular, the multi-scale formulation avoids excessive smoothing, allowing one to see 25 local effects where observations become available. Fourth, these methods are demonstrated with the XCO₂ data from the OCO-2 satellite.

Several interesting kriging studies have been published before in the context of satellite measurements of CO₂. Zeng et al. (2013) analyzed the variability of CO₂ in both space and time over China producing monthly maps from GOSAT data with slightly over 10000 observations. Nguyen et al. (2014) used a four times larger set of observations with Kalman Smoothing in 30 a reduced dimension with GOSAT and the Atmospheric InfraRed Sounder (AIRS) data from NASA. A map of atmospheric carbon dioxide derived from GOSAT data was presented at the higher resolution of 1×1.25 degrees in space and 6 days in time by Hammerling et al. (2012). In another publication by the same authors, synthetic OCO-2 observations were considered with the same spatial resolution.

A global dataset derived from GOSAT was presented by Zeng et al. (2017), with the spatiotemporal resolution of three days 35 and one degree. The results were validated against both Total Carbon Column Observing Network (TCCON) and modeling



5 results from CarbonTracker and the Goddard Earth Observing System with atmospheric chemistry (GEOS-Chem). This study evaluated also the temporal trend of the XCO₂. Similarly Tadić et al. (2017) describe a moving window block kriging algorithm to introduce time dependence into GOSAT-based XCO₂ map construction process using a quasi-probabilistic screening method for subsampling observations, thinning the data for computational reasons. Other recent studies have also contained analyses of OCO-2 data. For example, Zammit-Mangion et al. (2018) present fixed rank kriging (FRK) results based on OCO-2 data using a 16-day moving window. The results again appear very smooth.

10 An interesting approach is presented by Ma and Kang (2017), who describe a fused Gaussian process, combining a graphical model with a Gaussian process and applying that to sea surface temperature data. Another interesting approach for atmospheric trace gas inversion is presented by Zammit-Mangion et al. (2015), who simultaneously model both flux fields and concentrations using a bivariate spatiotemporal model, utilizing Hamiltonian Monte Carlo (Neal, 2011) for sampling the posterior. However, due to computational challenges the footprint area is very small.

15 For overcoming the difficulties posed by large numbers of data, various methods have been proposed. Lindgren et al. (2011) provide an explicit link between some random fields arising as solutions to certain stochastic partial differential equations and Markov random fields. A recent review of Vecchia-type approximations (Vecchia, 1988) is given by (Katzfuss et al., 2018) and a comparison of the performance of several recently developed methods is given by Heaton et al. (2018), with applications to MODIS data. The difficulty of ordering the observations for effective inference with Gaussian processes, especially as the dimension of the inputs grows, is underlined by Ambikasaran et al. (2016).

20 In this work we describe an approach to solve spatial statistics problems with hundreds of millions of data points. We do this by combining various ideas and techniques that come close to those applied in Vecchia-type and nearest neighbor Gaussian processes while utilizing random sampling and aggressive pre-filtering of uninformative data when possible. The presentation of the general Gaussian process problem is based on the one given by Santner et al. (2003) and Rasmussen and Williams (2006).

25 A generic space and time dependent mean function of the Gaussian process is found by solving marginals of a Markov random field (MRF). For covariance modeling, a multi-scale covariance kernel formulation is given. The validity of the multi-scale approach is established via a synthetic study. Approximate methods to learn the parameters of both the covariance kernel and the mean function as implemented in satGP are outlined. Additionally, a non-stationary covariance kernel formulation for utilizing wind data for computation, partly inspired by (Nassar et al., 2017), is proposed.

30 The capabilities of this early version satGP are demonstrated in practice by computing global XCO₂ concentrations for a duration of 1526 days at 0.5° spatial and daily temporal resolution with XCO₂ data from OCO-2 utilizing over 116 million observations. The number of computed marginals is over 350 million. An example of how these results look like is given by Fig. 7.

35 The key advances of this work are the capability to compute Gaussian process predictions with enormous remote sensing data sets, a practical way of learning the multi-scale kernel parameters and mean function parameters from data, and introduction of the flexible open source software, of which this is a first released version. Describing these developments is approached from the perspective of how the various parts of computation are implemented in the current version of satGP.



The rest of the manuscript is organized in the following manner: Section 2 describes the methods both generally and as implemented in satGP. An overview of computation in satGP is given in Sect. 3, and Sect. 4 presents and discusses simulation results, including a multi-scale synthetic parameter identifiability study and two applications to the OCO-2 v9 data set. In the concluding Sect. 5 some possible future directions are briefly mentioned.

5 2 Methods

In geosciences, kriging (Cressie and Wikle, 2001; Chiles and Delfiner, 2012) is often used for performing spatial statistics tasks such as gap-filling or representing data in a grid. The semivariogram models used in kriging are closely related to the covariance models used in the Gaussian process formalism (Santner et al., 2003; Rasmussen and Williams, 2006; Gelman et al., 2013), where instead of learning the variogram model from the data, a form of a covariance function is prescribed and its parameters learned.

Intuitively, one would like to learn properties of a spatio-temporal surface from some observational data of some quantity of interest. To each point in space and time corresponds a Gaussian distribution of that quantity, whose mean and variance can be calculated by solving a local regression problem at each desired point. This can also be crudely thought about as optimally solving a spatio-temporal interpolation problem when the observations have Gaussian errors.

The underlying theory related to Bayesian statistics, Gaussian processes, and Markov random fields is well known and therefore the novel aspects in this section have to do with the computational methods and modifications that are presented, such as observation selection schemes in Sect. 2.6 or approximate marginal maximum likelihood computation in Sect. 2.7. These modifications trade precision for tractability, but in a way that the results still remain valid. Due to the size of the problem, some sacrifices need to be made in order to be able to obtain any solution.

This section goes through the Gaussian process formalism, and both generic and the satGP-specific forms of mean and covariance functions are described. This is followed by discussion of how observation selection is carried out and how model parameters are learned.

2.1 Gaussian process regression

A Gaussian process is a stochastic process, which can be thought of as an infinite-dimensional Gaussian distribution in that the joint distributions at any finite set A of space-time points are multivariate normal. We denote the vector of these points by $x \in \mathbb{R}^q$ and underline that they contain both space and time components. In this work $q = 3$, even though this restriction can be overcome if needed, and satGP does have limited support for space-only problems.

The Gaussian process is denoted by

$$\Psi(x) \sim \text{GP}(m(x; \beta), k(x, x'; \theta)), \quad (1)$$



where $m : \mathbb{R}^q \rightarrow \mathbb{R}$ and $k : \mathbb{R}^{q^2} \rightarrow \mathbb{R}$ are respectively the mean and covariance functions of the process parameterized by hyperparameter vectors $\beta \in \mathbb{R}^{n_\beta}$ and $\theta \in \mathbb{R}^{n_\theta}$. Note, that with these functions x and x' refer to coordinates of a single location in the spatio-temporal domain, while below it may also refer to multiple locations, depending on context.

The function m above is called the drift in kriging literature, and the expected value of the process in areas with no data will tend to the value of the mean function in that area. It is chosen to reflect the deterministic patterns in the data, and these choices also affect how the function k and parameters θ in Eq. (1) need to be chosen. With inadequate modeling of the mean function, the obtained uncertainty estimates may end up being unnecessarily large. For instance linear trends, constant factors, seasonal and other periodic fluctuations should be included if they are known. An example of what is used with the OCO-2 data is shown later in Eq. (11).

The covariance function $k(x, x'; \theta)$ controls the smoothness of the draws ψ from Ψ . The parameter vector θ typically contains at least one scale parameter ℓ and a parameter controlling the maximum covariance τ^2 . The ℓ parameters correspond to the length scales of the random fluctuations of the realizations around the mean function, and the τ parameters describe the amplitude of that fluctuation. The functions m and k are fully described in Sect. 2.3 and 2.5, respectively. Additional practical guidelines are given in Appendix A.

In what follows the domain $\mathbb{R}^q \ni x$ is divided into two disjoint parts, one of which, $\mathcal{X}^{\text{train}} \subset \mathbb{R}^q$, contains the part where observation data (training data) was measured, and another one, $\mathcal{X}^{\text{test}} = \mathbb{R}^q \setminus \mathcal{X}^{\text{train}}$, where observations were not made. Any $x \in \mathcal{X}^{\text{test}}$ is below called *test input* as is often done in the GP literature, and these points are generally denoted by x^* .

In practice marginals of the random function Ψ in Eq. (1) or samples ψ from it are evaluated (computed) only at a finite set of points. Let $\psi^{\text{obs}} \in \mathbb{R}^n$ denote a vector of observations — synthetic or real — generated by the Gaussian process at locations $x^{\text{obs}} \in \mathbb{R}^{n \times q}$. Given a set of functions f_i for constructing the mean function, the matrix with elements $f_i(x_j; \delta(x_j^s))$ corresponding to locations x_j with regression coefficients $\beta(x_j^s)$ is denoted by $F(x)$. For a single input, instead of $F(x)$ the notation $f : \mathbb{R}^q \rightarrow \mathbb{R}^{n_\beta}$ is used, and with that, $f(x^*) = [f_1(x^*), \dots, f_{n_\beta}(x^*)]^T$. The joint distribution of the field at observed locations is then given by

$$\psi^{\text{obs}} \sim \mathcal{N}(F(x^{\text{obs}})\beta, K), \quad (2)$$

where the covariance matrix K is defined by its elements $K_{i,j} = k(x_i^{\text{obs}}, x_j^{\text{obs}}; \theta)$. For the mean function, in this work a specific form

$$m(x; \beta, \delta) = f(x; \delta)^T \beta(x) \equiv \tilde{f}(x^t; \delta(x^s))^T \beta(x^s) \quad (3)$$

is used, where the superindexes s and t refer to the spatial and temporal parts of the generic coordinate x , respectively, and $\delta(x^s)$ are auxiliary parameters which are potentially space-dependent. The purpose of the function \tilde{f} is purely illustrative, showing that given the parameters δ , the function f does not depend on the spatial part of x , and similarly that the β parameters do not depend on x^t . This definition of m is very general and can describe in practice a large number of realistic scenarios. However, the form of Eq. (3) imposes the strong assumption of separation of space and time in that the β and δ parameters do not depend on time. The explicit form of functions f_i used to model the OCO-2 data are given below in Sect. 2.3.



Bayesian statistics is a standard paradigm for analyzing data and uncertainties, and it is also widely used in geosciences (Rodgers, 2000; Gelman et al., 2013). From the vantage point it provides, given the observed data $\Psi^{\text{obs}} = \psi^{\text{obs}}$ at some finite set of points x^{obs} , the object of interest of the inference problem in this work is the joint posterior distribution of the Gaussian process and the parameters,

$$5 \quad p(\psi, \beta, \delta, \theta | \psi^{\text{obs}}) = \frac{p(\psi^{\text{obs}} | \psi, \beta, \delta, \theta) p(\psi | \beta, \delta, \theta) p(\beta, \delta, \theta)}{p(\psi^{\text{obs}})}, \quad (4)$$

where $p(\psi | \beta, \delta, \theta)$ is the Gaussian process prior and $p(\beta, \delta, \theta)$ is a prior on the Gaussian process hyperparameters. This calculation is not generally tractable for a huge number of inputs x , but posterior estimates of the GP, $p(\psi | \psi^{\text{obs}}, \hat{\beta}, \hat{\delta}, \hat{\theta})$, can be calculated by conditioning on parameter point estimates $\hat{\theta}$, $\hat{\beta}$, and $\hat{\delta}$. The first of these may be found by minimizing some loss function \mathcal{L} , described below in Sect. 2.7,

$$10 \quad \hat{\theta} = \arg \min_{\theta} \mathcal{L}(\theta), \quad (5)$$

and for the second a closed-form expression, given a point estimate of the parameters θ and δ , is given by

$$\mathbb{E}[\beta | \Psi^{\text{obs}} = \psi^{\text{obs}}, \theta, \delta] = (F^T K^{-1} F)^{-1} F^T K^{-1} \psi^{\text{obs}} \quad (6)$$

$$\nabla[\beta | \Psi^{\text{obs}} = \psi^{\text{obs}}, \theta, \delta] = (F^T K^{-1} F)^{-1}. \quad (7)$$

The δ parameters can be found approximately by finding a point estimate of parameters β and δ before computing Eq. (6), and by re-calibrating δ alone after. In practice this produces stable results with the OCO-2 data, and for pathological data sets, repeated alternating optimization of the parameters may be performed.

Even though a full posterior distribution of the parameters is not obtained this way, the solution of the Gaussian process itself is Bayesian in that the posterior marginals at each x are found by conditioning on the observations. In the satGP software, the space-dependent β and δ parameters are fitted first, and any learning of the covariance parameters is done only after that.

20 For prediction in the context of Gaussian random functions, the properties of multivariate normal distributions are exploited for calculating marginals of the random field Ψ at any set of points x .

The posterior distribution $p(\psi^* | \psi^{\text{obs}}, \hat{\theta}, \hat{\beta})$ of the Gaussian process at a finite set of test inputs x^* can, given point estimates $\hat{\beta}$ and $\hat{\theta}$, be modeled according to Eq. (2) with

$$\begin{pmatrix} \Psi^* \\ \Psi^{\text{obs}} \end{pmatrix} \sim \mathcal{N} \left(\begin{bmatrix} f(x^*)^T \\ F(x^{\text{obs}}) \end{bmatrix} \hat{\beta}, \begin{bmatrix} K(x^*, x^*) & K(x^*, x^{\text{obs}}) \\ K(x^{\text{obs}}, x^*) & K(x^{\text{obs}}, x^{\text{obs}}) \end{bmatrix} \right) \quad (8)$$

25 where Ψ and x have been divided into two parts - one for the test inputs x^* , and the other one for the observations x^{obs} . The predictive distribution at x^* can then be written as $\Psi^* | \hat{\beta}, \hat{\theta} \sim \mathcal{N}(\mu^*, \Sigma^*)$, where its moments are given by

$$\mu^* = f(x^*)^T \hat{\beta} + K(x^*, x^{\text{obs}}) K(x^{\text{obs}}, x^{\text{obs}})^{-1} (\psi^{\text{obs}} - F \hat{\beta}) \quad (9)$$

and

$$\Sigma^* = K(x^*, x^*) - K(x^*, x^{\text{obs}}) K(x^{\text{obs}}, x^{\text{obs}})^{-1} K(x^{\text{obs}}, x^*), \quad (10)$$

30 and where the covariance Σ^* is the Schur complement of $K(x^*, x^*)$.



2.2 Overview and objectives of satGP

The satGP program is meant to be a general purpose Gaussian process toolbox with emphasis on applicability to large remote sensing datasets. It features a selection of covariance kernels and routines for learning space-dependent mean function parameters and covariance parameters from data. With a given set of parameters, it computes posterior marginals and uncertainties at the spatial resolution desired by the user, or generates samples from the process. Drawing samples from the prior is also supported, and this can be utilized for devising synthetic data experiments to study the identifiability of the GP covariance kernel parameters. This section goes through these capabilities and relevant computational details. Since the software is applied in Sect. 4 to OCO-2 data, details pertaining to that particular case are included for illustration.

2.3 Mean functions in satGP

The most general mean function form available in satGP is given by Eq. (3). The functions f_i above are user-defined and, for ease of use, functionality for using a zero mean function, a spatially independent mean function, and an arbitrary gridded array of values are available. The specific forms of f_i used for the OCO-2 experiments in Sect. 4 are given by

$$\left. \begin{aligned} f_1(x) &= \sin(2\pi x^t \Delta_{\text{year}}^{-1} + \delta_{x^s}) \\ f_2(x) &= \cos(4\pi x^t \Delta_{\text{year}}^{-1} + \delta_{x^s}) \\ f_3(x) &= 1 \\ f_4(x) &= x^t \end{aligned} \right\} \quad (11)$$

where Δ_{year} is the duration of one year, and δ_{x^s} is a space-dependent phase shift. The function f_1 fits the summer-winter cycle, and f_2 fits the semiannual cycle. It is assumed that these can be modeled with the same δ_{x^s} parameters. The constant term is given by f_3 , and f_4 gives the slow global trend. The fit to the global mean values of XCO₂ from OCO-2 can be seen in Fig. 1.

2.4 Learning $\beta(x^s)$ as a Markov random field

When not learning GP covariance parameters or generating synthetic training sets, the finite set of test inputs x^* for GP calculation is taken in satGP to be a grid with predefined geographical and temporal extents and resolution. Solving the GP marginalization and sampling problems then amounts to solving Eq. (9) and (10) at each corresponding space-time point. Since e.g. sources, sinks and timing of seasons are local, the mean function should be different from one spatial grid point to another. This is achieved by modeling the $\beta(x^s)$ parameters as a Markov random field, which are often used in geophysics as a computational tool to solve large spatial statistics or inference problems. In practice what follows explains how the spatial dependence can be resolved using computational statistics. The MRF imposes the condition that neighboring grid cells should not be too different from each other. How different they are allowed to be is a modeling choice, see Appendix A.

This MRF is an undirected graphical model $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ (Lauritzen, 1996) with the set of vertices $\mathcal{V} = \{\nu^{ij} | i = 1 \dots m_{\text{lat}}, j = 1 \dots m_{\text{lon}}\}$ and edges $\mathcal{E} = \{(\nu^{i,j}, \nu^{i+1,j}) | i = 1 \dots m_{\text{lat}} - 1, j = 1 \dots m_{\text{lon}}\} \cup \{(\nu^{k,l}, \nu^{k,l+1}) | k = 1 \dots m_{\text{lat}}, l = 1 \dots m_{\text{lon}} - 1\}$. The vertices ν^{ij} correspond to the mean function parameters β^{ij} at grid point (i, j) . This Markov property implies that the prob-

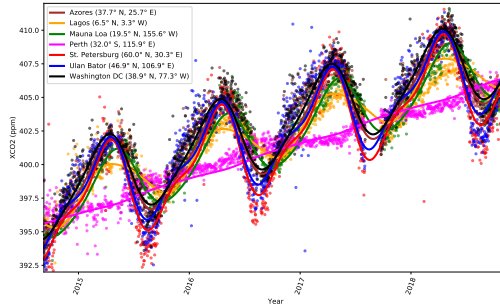


Figure 1. Mean function m with components of f given by Eq. (11). The solid lines give the mean function value, fitted to local data, and the corresponding daily means are shown as dots of the same color. The fit is not perfect at all times due to e.g. smoothness constraints of the field, but it works well as the Gaussian process mean function.

ability of the β -parameters of latitude i and longitude j is given by $p(\nu^{ij}) = \int_{\partial\nu^{ij}} p(\nu^{ij}|\partial\nu^{ij})p(\partial\nu^{ij})$, where $\partial\nu^{ij} = \{\nu \in \mathcal{V} | (\nu, \nu^{ij}) \in \mathcal{E}\}$.

Since the maximal cliques of this graph are the connected pairs of vertices, according to Hammersley and Clifford (1971) the full joint distribution of the graph $p(\mathcal{V})$ factors as $\prod_{(\nu, \nu') \in \mathcal{E}} \frac{1}{Z} \phi(\nu, \nu')$, where Z is called a partition function and ϕ are compatibility functions. One reasonably efficient way to solve marginals for each vertex in such a graph is to use the variable elimination algorithm, which is an exact standard algorithm suitable for undirected graphs of moderate size. To make the computation faster, satGP currently uses a modified version to compute each diagonal in the graph in parallel from $\nu^{0,0}$ to $\nu^{n_{\text{lat}}, n_{\text{lon}}}$ and back, conditioning each ν^{ij} on the previously evaluated vertices in $\partial\nu^{ij}$ without introducing the diagonal edges of the reconstituted graph, as would be normally done. The program also inversely weights the edges exponentially according to the distances between the (geographical) coordinates corresponding to the connected nodes. This rate of exponential decay is user-configurable. The structure of the MRF and the approximate elimination order are shown in Fig. 2.

In the particular form used for OCO-2 data in Eq. (11), the phase-shift parameter δ cannot be estimated with regression like β in Eq. (9) and (10). For this reason, the nonlinear space-dependent δ -parameters are found with an optimization algorithm from the NLOpt package, by default the BFGS algorithm, before finding $\hat{\beta}$ with Eq. (9) and (10), and after obtaining $\hat{\beta}$ the δ parameter is re-optimized given the $\hat{\beta}$. For calibrating the δ parameters for vertex ν , the quantity $\sum_{j=1}^n (m(x_\nu; \beta_\nu, \delta_\nu) - \psi_j)^2 + \sum_{j' \in \partial\nu} (\delta_\nu - \delta_{j'})^2$ is minimized. Here the first sum runs over the training data selected by the observation selection method described in Sect. 2.6. This optimization problem is very simple since there are few β or δ parameters for the individual vertices. The complexity introduced by the interactions described by the edges is taken care of by the approximate elimination algorithm described above.

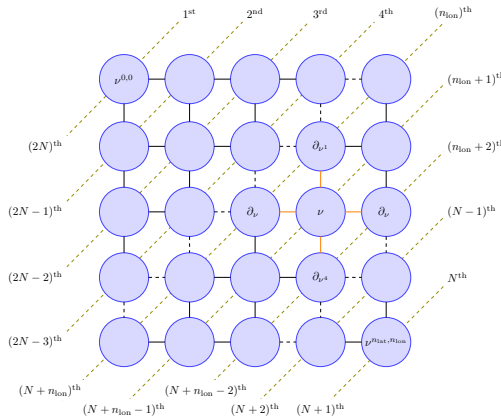


Figure 2. The marginal distribution of vertex ν , $p(\nu)$, is conditional only on the neighbors $\partial_{\nu,1} \dots \partial_{\nu,4}$ (red edges) due to the Markov structure in the pictured lattice graph. Each connected pair is a maximal clique in this particular case. For effective solving, the vertices on the diagonal dashed lines are computed simultaneously making the algorithm non-exact. The order numbers labeling the diagonal lines represent an ordering in which the diagonals can be computed in parallel to get all the marginals in $\mathcal{O}(N)$ wall time, where $N = n_{lat} + n_{lon} - 1$. The $(N + 1)^{\text{th}}$ computation in the corner is not conditioned on already-computed neighbors to avoid double counting data.

2.5 Covariance functions in satGP

The smoothness, amplitude, and length scale of the Gaussian process are determined by the covariance kernel used, and this choice much determines how the result of the computation looks like. The satGP program supports several different types of covariance function components for forming the full covariance function k in Eq. (1). The options available reflect the properties that can be expected in remote sensing data – varying smoothness and meridional and zonal length scales, potential periodicity, and changing the orientation of the data-informed and uninformed axes according to wind speed and direction. This section lists the available covariance function formulations. For further intuition regarding the parameters, also see Appendix A.

For convenience, let

$$10 \quad \xi_{\ell_I}^\gamma(x, x') = \sum_{c \in I} \left| \frac{x^c - x'^c}{\ell_c} \right|^\gamma = \|P^I(x) - P^I(x')\|_\Gamma^\gamma, \quad (12)$$

where $\gamma > 0$ is the exponent, $I \subseteq \{x^s, x^t\}$ is a set of dimensions of the input, with x^s referring to latitude and longitude and x^t to time. The P^I matrix projects x onto indices I , and Γ is a diagonal covariance matrix with elements ℓ_c^γ , and the notation $\|r\|_\Gamma$ stands for $\sqrt{r^T \Gamma^{-1} r}$. The space-only variables are denoted I_S and spatial and temporal variables together are denoted I_{ST} .



The exponential family of covariance functions with parameters $\theta = (\gamma, l, \tau)$ is defined by the covariance function

$$k_{\text{exp}}(x, x'; \theta, I) = \tau^2 \exp(-\xi_{\ell_I}^\gamma(x, x')). \quad (13)$$

The exponent γ controls the smoothness of the samples from the Gaussian process, with $\gamma = 2$ yielding infinitely differentiable realizations.

5 The Matérn family of covariance functions, with $\theta = (\nu, \ell_I, \tau)$ is given by the covariance

$$k_{\text{M}}(x, x'; \theta) = \frac{\tau^2 s^\nu}{\Gamma(\nu) 2^{\nu-1}} K_\nu(s), \quad (14)$$

where $s = 2\sqrt{\nu} \xi_{\ell_I}^\nu(x, x')$ and ν controls the smoothness parameter usually denoted by α via $\alpha = \nu + \frac{q}{2}$. The function K_ν is the modified Bessel function of the second kind of order ν . With $q = 1$, the value $\nu = \infty$ corresponds to the squared exponential kernel and $\nu = 0.5$ to the exponential kernel with $\gamma = 1$. Despite this similarity between the Matérn and exponential kernels,
 10 the realizations of the random function from the processes with values $\frac{1}{2} < \nu < \infty$ do not correspond to those with the kernel k_{exp} with any value of γ .

A periodic kernel with $\theta = (\tau, \ell_{\text{per}}, \theta_{\text{exp}})$ is defined in satGP by

$$k_{\text{per}}(x, x'; \theta, I) = \tau^2 \exp\left(-\frac{2 \sin^2\left(\pi \left[\frac{x^t - x'^t}{\Delta_{\text{period}}}\right]\right)}{\ell_{\text{per}}^2} - \xi_{\ell_S}^\gamma(x, x')\right), \quad (15)$$

15 and the term θ_{exp} defines the parameters for the exponential functions ξ , while ℓ_{per} controls the periodic (inter-period) covariance length. While the periodic kernel is not utilized with the OCO-2 case studies below, it can be a useful tool in many other situations, such as with OCO-3, which due to not being on a Sun-synchronous orbit will make observations at varying local times.

An additional covariance function formulation available in satGP is one based on local wind information. The underlying
 20 rationale is that winds affect how quantities of interest such as gases in the atmosphere or algae blooms in the surface water spread. Therefore, if wind data is available, it is natural to use it in the Gaussian process.

The wind-informed covariance has parameters $\theta = (\tau, \ell_I, \rho, w^*)$ and is defined by

$$k_W(x, x'; \theta, I) = k_{\text{exp}}(x_W, x'_W; \theta^W, ST), \quad (16)$$

where the difference between x_W and x'_W is represented using transformed axes parallel and perpendicular to the wind direction at the test input x^* . The spatial scaling parameters in Eq. (13) for k_W , corresponding to the parallel to wind and
 25 perpendicular to wind directions, are given by

$$\ell^{\parallel} = \ell \sqrt{1 + w^* \rho} \quad \ell^{\perp} = \ell, \quad (17)$$

where w^* is the wind velocity at the test input x^* and ρ scales the effect of the wind. The parameter vector for the exponential kernel $\theta^W = (\tau, \gamma, \ell^{\parallel}, \ell^{\perp}, \ell_t, 2)$, where the last element denotes the exponent γ used by the exponential kernel. The resulting
 30 covariance ellipses are shown in Fig. 3 for several wind vectors and values of ρ .

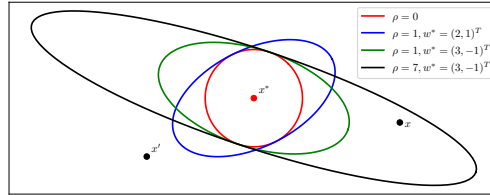


Figure 3. Equicovariance ellipses from the wind-informed kernel with various wind vectors w^* and values of ρ . The wind values are taken at the test input x^* , but the covariance function k is evaluated also for each pair of observations x and x' .

The covariance functions used in this work to model Ψ are sums of several kernels - sums of valid Gaussian process kernels remain valid kernels. The general form of this *multi-scale kernel* is given by

$$k(x, x'; \theta) = \delta_{x, x'} \sigma_x^2 + k_{\text{per}}(x, x'; \theta, I_S) + k_M(x, x'; \theta) + k_{\text{exp}}(x, x'; \theta, I_{ST}) + k_W(x, x'; \theta, I), \quad (18)$$

- 5 where the first term, which in kriging is called the nugget, contains the observation error variances, and the parameter θ is understood to be different for each component. Not all kernels are included in all experiments - rather, the simulations in Sect. 4 utilize kernels with one to three components. The kernel components of a multi-scale kernel are below called *subkernels*.

2.6 Covariance localization and observation allocation for the multi-scale kernel

Using a large number of observations makes solving the Gaussian process Eq. (9) and (10) untractable as the cost of inverting the covariance matrix scales as $\mathcal{O}(n_{\text{obs}}^3)$. This creates a need for finding approximate solutions while introducing as little error as possible. In satGP covariance localization is used to utilize only a subset of observations for computing Eq. (9) and (10). To do this, a maximum subkernel covariance matrix size κ and minimum covariance parameter σ_{min}^2 are defined by the user.

Assume that the multi-scale kernel defined by the user contains n_{ker} subkernels. For each test input x^* and for each subkernel k_l let the set of observations feasible for inclusion in K in Eq. (8) be

$$15 \quad A_{*,l}^{\text{obs}} = \{\psi_i \in \psi^{\text{obs}} | k_l(x_i, x^*) < \sigma_{\text{min}}^2, \psi_i \notin A_{*,j}^{\text{obs}} \forall j < l\}, \quad (19)$$

where the last condition prevents observations from being added by several subkernels. From these candidate observations, $\min(|A_{*,l}^{\text{obs}}|, \kappa)$ are selected, either greedily selecting the κ observations with highest $k(x_i, x^*)$, or choosing the observations uniformly randomly sampling from those training data for which the minimum covariance threshold is exceeded, see Appendix A for additional details. When $|A_{*,l}^{\text{obs}}| < \kappa$ and $l < n_{\text{ker}}$, the parameter κ will be grown for the next kernel to compensate for the deficit by setting $\kappa \leftarrow \kappa + (\kappa - |A_{*,l}^{\text{obs}}|)$. This is done to allow the full kernel size to grow to $n_{\text{ker}}\kappa$ when possible.

Since the kernels are handled sequentially, the order of the different kernels may slightly affect which observations are selected due to the exclusion in Eq. (19), and to grow the full kernel to size $n_{\text{ker}}\kappa$ as often as possible, it is recommended to



specify the subkernel with the largest ℓ parameters as the last one. After selecting all observations for all kernels, the covariance matrix K is constructed by evaluating the full covariance function k according to Eq. (18) for all pairs of selected observations.

For learning the locally varying parameters in the mean function with Eq. (6) – (7), the observation selection is performed by disregarding the time component, i.e. setting $x_i^t \leftarrow x^{*t}$ for all x_i .

- 5 Observation allocation could be done also by selecting observations based on values of k instead of each k_l individually, or by other approaches, such as the one presented by Schäfer et al. (2017). However, while the method of observation selection does have an effect on the inferred posterior marginals, the screening property of Gaussian processes ensures that this effect is not major as long as observational noise is small and the nearest observations are included in all directions.

- 10 Out of the two methods available in satGP, random selection avoids observation sorting and is therefore faster, especially if a huge number of data are near the test input x^* . This comes at the cost of producing slightly noisier fields of marginal posterior means. For covariance parameter estimation random selection works well. The current nearest-neighbor-in-covariance approach is only one possibility, but is justified by the parameter identifiability results in Sect. 4.1.

2.7 Learning the covariance parameters θ

From Eq. (4), the log marginal likelihood of observations ψ^{obs} given a set of parameters θ , β and δ is given by

$$15 \quad 2 \log p(\psi^{\text{obs}} | \beta, \delta, \theta) = -\|(\psi^{\text{obs}} - F\beta)\|_K - \log |K| - n_{\text{obs}} \log(2\pi), \quad (20)$$

where the covariance function parameters θ are implicitly in K and the non-linear mean function parameters in F , for which the shorthand notation $F = F(x^{\text{obs}})$ is used in this section. The maximum (marginal) likelihood estimate (MLE) $\hat{\theta}$ of θ can be found via minimizing

$$\mathcal{L}(\theta) = \left\{ \|(\psi^{\text{obs}} - F\hat{\beta})\|_K + \log |K| + n \log(2\pi) \right\} \quad (21)$$

- 20 as stated in context of Eq. (5).

In the presence of a huge number of observations, calculating the determinant of the full covariance $|K|$ is not feasible, and the log likelihood is approximated with the block diagonal form, resulting in

$$\hat{\theta}_{\text{MLE}} = \arg \min_{\theta} \sum_{x_i \in E_{\text{ref}}} \left\{ \|(\psi_i^{\text{obs}} - F_i\hat{\beta})\|_{\tilde{K}_i} + \log |\tilde{K}_i| \right\}, \quad (22)$$

- 25 where E_{ref} is a set of randomly sampled points from the specified spatio-temporal domain. While the selection of inputs included in E_{ref} has an effect on the obtained parameter estimate, that effect has proven in simulations to be small. The vector $\psi_i^{\text{obs}} \in \mathbb{R}^{d_i}$ contains observations closest in covariance to x_i , chosen according to the observation allocation rules outlined in Sect. 2.6. The last term in Eq. (21) is dropped, since while varying θ in Eq. (22) changes d_i , the size of ψ^{obs} stays the same.

- 30 While this method is suitable for finding point estimates for the parameters θ , the log-likelihood has an unknown scaling factor resulting in an unknown multiplicative factor for the variance term in the exponent of the Gaussian distribution, and hence information about the true size of the posterior of the covariance parameters $p(\theta | \psi^{\text{obs}}, \beta, \delta)$ is lost.



The covariance parameter optimization can be performed by using optimization algorithms such as COBYLA or SBPLEX available in NLOpt (Johnson, 2014). An alternative is to explore the scaled posterior by using the Adaptive Metropolis (AM) Markov chain Monte Carlo (MCMC) algorithm (Haario et al., 2001), an implementation of which is included in the satGP source code. Using MCMC is feasible since the forward model is just sampling from a multivariate normal distribution which is very fast, and also due to that the parameter dimension is moderate, even with multiple subkernels.

3 Overview of Computation

The satGP code is written in C with visualization scripts written in Python and parallelization implemented with OpenMP directives. The program reads data from netCDF files and the configuration from a C header file. For linear algebra, the C interfaces of LAPACK and BLAS, LAPACKE and CBLAS, are utilized and for optimization tasks, algorithms in the NLOpt library are used. The computations are carried out in single precision both in order to save memory resources with the largest data sets and also in anticipation of implementing the covariance function routines in a way that allows computation on graphics processing units.

The most important configuration variables are listed in Table 1. The user needs to define whether parameters are learned or prescribed and whether marginals or samples from the GP are to be computed. The mean and covariance kernel need to be defined by initializing corresponding structs with parameters and their limits if calibration is to be performed. For computing GP marginals or drawing samples from the random process, the geographic and temporal extents need to be specified and the mean function and the covariance kernel used must be given. For more details than is described below, see Appendix A.

For computational efficiency, several parameters can be tweaked, including all of those in the second and last sections of Table 1. The first main bottleneck for computing a marginal at x^* is sorting the observations for selecting the most informative ones to be used in the covariance matrices, see Sect. 2.6. This requires roughly $\mathcal{O}(r_l \log r_l + \kappa \log \kappa)$ operations, where $r_l \propto \prod_{i=1}^q \ell_i^l$ is the number of grid locations (test inputs) x_i^* in the spatial grid such that for the l^{th} subkernel, $k_l(x_i^*, x^*) < \sigma_{\min}^2$. Here the parameters ℓ_i^l are the corresponding length scale parameters over all the dimensions of the inputs x – this controls the size of the hypersphere inside which observations are considered for each x^* . The second bottleneck is calculating the Cholesky decompositions of the covariance matrices K with cost $\mathcal{O}((n_{\text{ker}}\kappa)^3)$. The cost of calculating the means and variances of the GP in a grid for a set of n_{times} points on the time axis is therefore given by

$$\text{cost} = \mathcal{O} \left(\frac{A n_{\text{times}}}{\omega^2} \left[(n_{\text{ker}}\kappa)^3 + \sum_{l=1}^{n_{\text{ker}}} (r_l \log r_l + \kappa \log \kappa) \right] \right), \quad (23)$$

where A is the grid area in degrees squared and ω is the grid resolution. When the random observation selection method mentioned in Sect. 2.6 is used, the $r_l \log r_l$ in Eq. (23) becomes just r_l .

The execution of the program is presented in Fig. 4. The names of the subprograms here deviate from those in the code to improve readability.

The function `AddToState()` reads observations (asynchronously) into a `state` object that tracks the proximity of each observation to each grid point. Only part of data is added, and what part, is controlled on l. 6 by the parameter η_{train}^i , which



Table 1. Most important satGP control variables and high level C structs: first section contains parameters for program logic, second for domain specification, third for covariance and mean function definition, and last for observation handling. This list is by no means exhaustive – the configuration file contains lots of variables that can control the program. Some additional tweaking is possible by changing hard-coded values directly in the source code, such as those listed in Appendix A.

Variable	Type	Low	High	Notes
learn_k	int	0	2	(0) Don't train θ , (1) generate observations and learn θ , (2) learn θ from non-synthetic data.
learn_m	int	0	1	(0) Don't train local β and δ , (1) find local β and δ as in Sect. 2.4.
sampling	int	0	2	(0) Skip sampling, (1) calculate GP marginals at each grid point, (2) sample from GP.
area	char*	-	-	Area definition setting longitude and latitude minimum and maximum values
n_days	int	1	∞	Number of days to be simulated
ω	float	> 0	180	1-d grid resolution in degrees – small values degrade esp. posterior sampling performance.
n_ker	int	1	10	Number of subkernels k_l in k
cfc	struct*	-	-	Recursive struct pointer defining $k_1 \dots k_{n_{ker}}$ and corresponding θ , see Sect. 2.5.
mf	struct*	-	-	Struct pointer for defining type of $m(\cdot, \cdot)$ and associated (initial) β and δ , see Sect. 2.3.
ζ_{train}	float	0	∞	Determines what fraction of observations are randomly included in ψ^{obs} when learning θ .
ζ_{sample}	float	0	∞	Determines what fraction of observations are randomly included in ψ^{obs} when sampling $\neq 0$.
σ_{min}^2	float	0	∞	Discard observation at x_i for x^* if $k(x_i, x^*) < \sigma_{min}^2$, see Sect. 2.6.
n_ref	int	0	∞	Number of reference points in E_{ref} in Eq. (22) for training θ
n_synthetic	int	0	∞	Number of random locations where synthetic data is generated for training θ
$\sigma_{synthetic}^2$	float	0	∞	Variance of Gaussian noise added to synthetic observations
κ	int	1	∞	Maximum subkernel size, values $\kappa > n_{ker}^{-1} 1000$ will be slow due to $\mathcal{O}(\kappa^3)$ scaling.

corresponds to the inclusion probability of each observation. This probability depends on ζ_{train} in Table 1 via

$$\eta_{train}^i = \frac{d(x_i, x_{i_{prev}})}{\omega \zeta_{train}} \wedge 1, \quad (24)$$

where $d(x_i, x_{i_{prev}})$ is the Euclidean distance to the previous added point and \wedge is the standard notation for minimum. Hence with $\zeta = 0$, all observations will be added.

5 For computing the marginals, the spatial domain can be decomposed with `Decompose()`, line 23, into several spatial subdomains (sd) so that arbitrary-size grids can be computed. This makes solving large problems with limited amount of memory possible, but only works with `sampling = 2`.

The state object is emptied by `ReInitializeState()` which also potentially sets new subdomain extents. Function `SampleFromPrior()` actually performs the computations on lines 30-37, but with the set of points x_i^* in a random pattern
 10 instead of in a grid as is the case in l. 27-38.



The `AddSubdomainData()` method on l. 29 adds data as on lines 3-9, but only to the current subdomain. After that, the `SelectObservations()` method (l. 31) carries out selecting the best observations as described in Sect. 2.6. For constructing the set of potential observations, the grid is searched for locations that may have informative observations for the current test input stored in the `state` object. These locations are first ordered into categories with decreasing potential covariance and
5 for the best locations, that together hold at least 2κ observations, the covariance function with the test input is evaluated. Out of these, the κ best are chosen. The factor 2 can be increased for the wind-informed kernel and the value 8 is used in the demonstration of the wind-informed kernel in Sect. 4.7.

The function `ComputeMarginal()` constructs the covariance matrix K , inverts via the Cholesky decomposition, and solves Eq. (9) and (10) to find the marginal distribution at any test input x^* . That function returns the negative log likelihood
10 and is therefore directly used in learning the covariance parameters θ in `FindCovfunCoeffs()` on line 18.

The Gaussian process algorithm is an interpolation algorithm when observation noise is zero, and interpolation algorithms may misbehave when used for extrapolation. In a spatio-temporal large grid, when `sampling = 2`, i.e. when draws of the Gaussian process are generated in a regular spatio-temporal grid, computing conditionals based on the previous predictions would amount to extrapolation if done in order. For this reason, a deterministic sparse ordering is used, which ensures that
15 test inputs corresponding to simultaneous predictions are far from each other so that their mutual covariance is negligible. For this reason conditioning on already computed values is for the vast majority of GP evaluations interpolation instead of extrapolation.

4 Results and discussion

In this section, several simulation studies are presented. In the first experiment, parameter identifiability with the multi-scale
20 kernel is examined with satGP-generated data. After that, the MRF of mean function β coefficients is trained with OCO-2 data and those fields are then briefly analyzed.

Based on a locally varying mean function of the form in Eq. (3), the covariance parameters of the OCO-2 XCO₂ spatio-temporal field are learned. Knowing both the mean and the covariance functions, the Gaussian process is then solved globally in a grid and snapshots of the mean and uncertainty fields are presented. The section is concluded by a demonstrating how the
25 wind-informed kernel works. The covariance function parameters are learned from data.

4.1 Parameter identifiability with the multi-scale kernel

A synthetic study was performed to confirm the identifiability of the multi-scale covariance function parameters. For this, sampling with a random spatial pattern from the prior was carried out, adding 1% noise, and then estimating the parameters by computing the posterior mean estimates using Adaptive Metropolis.

The identifiability experiment was performed with various kernels, and the more complex the kernel, the more difficult
30 recovering the true parameters was. With a single Matern, exponential, or periodic kernel, the parameters could be recovered very easily. This was also true for a combination of exponential and Matern kernels with a relatively small κ parameter.



Data: filelist containing files with observation data $y_i = (\mu_{\psi_i}, \sigma_{\psi_i}^2)$ indexed by location x_i , input variables from Table 1.

Result: Optimized β parameters for mean function and θ parameters for covariance kernel, gridded Gaussian process marginal means and variances or a sample from the Gaussian process evaluated in a grid.

```

1 Initialization: Create grid according to area and  $\omega$ ,
  define  $k(x, x')$  and  $m(x, t)$ , initialize state;
2 if learn_m = 1 or learn_k = 2 then
3   for file in filelist do
4     D  $\leftarrow$  ReadData (file);
5     for  $(x_i, y_i) \in D$  do
6       if Bernoulli( $\eta_{\text{train}}^i$ ) then
7         AddToState (state,  $x_i, y_i$ );
8       end
9     end
10 end
11 if learn_m then FindLocalMeanfunCoeffs (state);
12 if learn_k = 1 then
13   ReInitializeState (state, fulldomain);
14   for  $i \leftarrow 1$  to  $n_{\text{synthetic}}$  do
15      $(x_i, y_i) \leftarrow$  SampleFromPrior ();
16     AddToState (state,  $x_i, y_i$ );
17   end
18 end
19 if learn_k  $\neq$  0 then
20   FindCovfunCoeffs ( $n_{\text{ref}}$ )
21 end
22 if not sampling then
23    $(n_{\text{sd}}, (sd_i)_{i=1}^{n_{\text{sd}}}) \leftarrow$  Decompose ( $n_{\text{dom}}^{\text{max}}, \text{area}, \omega$ );
24 else
25   assert ( $n_{\text{gp}} < n_{\text{dom}}^{\text{max}}$ );
26 end
27 if sampling then for  $i \leftarrow 1$  to  $n_{\text{sd}}$  do
28   ReInitializeState (state,  $sd_i$ );
29   AddSubdomainData (state, filelist,  $sd_i, \eta_{\text{sample}}$ );
30   for  $x^* \in sd_i$  do
31      $A_*^{\text{obs}} \leftarrow$  SelectObservations (state,  $x^*$ );
32      $\mu^*, \sigma_*^2 \leftarrow$  ComputeMarginal ( $x^*, A_*^{\text{obs}}$ );
33     if sampling = 2 then
34        $\widehat{\psi}^* \leftarrow$  Normal ( $\mu^*, \sigma_*^2$ );
35       AddToState (state,  $x^*, (\widehat{\psi}^*, \sigma_{\text{synthetic}}^2)$ )
36     end
37   end
38 end
39 ..
    
```

Figure 4. Overview of satGP. After initialization data is read for training m and k , after which possible MRF computation is carried out. This is followed by sampling the prior if a synthetic study is performed, and learning the θ parameters controlling k . Gaussian process marginals are then computed in a grid, potentially by decomposing the domain for large grids. Finally, samples from the GP may be drawn.



The covariance kernel parameters were still recoverable with a combination of three kernels – Matern with $\nu = \frac{5}{2}$, exponential, and periodic, but for this, a larger κ was needed – the simulation shown used $\kappa = 256$. With small κ , some of the parameters had a tendency to end up at the lower boundary, possibly due to effects of the covariance cutoff on the determinant of the covariance matrix in Eq. (20). Optimization using minimization algorithms such as Nelder-Mead, COBYLA, or BOBYQA tended to often end up in local minima, and for this reason MCMC was used instead. The number of random reference points in E_{ref} in Eq. (22) was set to 12, which was enough to reliably recover parameters close to the true value.

The parameter limits, true values, and posterior means of the synthetic experiment with three kernels are given in Table 2. In total 200,000 observations were created in the region between -10 and 10 latitude and -10 and 10 longitude over a period of four years according to the values in Table 2. A total of 10 million Metropolis-Hastings iterations were carried out to make sure that the posterior covariance stabilized. The posterior, with first 50% of the chain discarded as burn-in, is shown in Fig. 5

Table 2. Lower and upper limits, with true and estimated parameter values. The three-kernel synthetic covariance function parameter estimation problem is already very difficult, here resulting in slight overestimation of the parameters of the smallest kernel.

	low	high	true	est	$\frac{\text{est}-\text{true}}{\text{true}}$
τ^{mat}	0.05	1	0.5	0.652	0.304
$\ell_{\text{lat}}^{\text{mat}}$	0.003	0.02	0.007	0.00989	0.413
$\ell_{\text{lon}}^{\text{mat}}$	0.003	0.02	0.01	0.0135	0.350
ℓ_t^{mat}	1d	14d	7d	8.06d	0.15
τ^{per}	0.01	2	1	1.073	0.073
$\ell_{\text{lat}}^{\text{per}}$	0.001	0.04	0.02	0.0207	0.035
$\ell_{\text{lon}}^{\text{per}}$	0.001	0.04	0.02	0.0220	0.1
ℓ_{per}	0.01	0.3	0.1	0.1075	0.075
τ^{exp}	0.5	3	1	0.927	-0.077
$\ell_{\text{lat}}^{\text{exp}}$	0.005	0.1	0.025	0.0352	0.408
$\ell_{\text{lon}}^{\text{exp}}$	0.005	0.1	0.04	0.0405	0.0125
ℓ_t^{exp}	7d	30d	21d	24.83d	0.182

4.2 The OCO-2 v9 data

The simulations with real remote sensing data utilize the v9 data from the OCO-2 satellite. The OCO-2 satellite was launched in 2014, and it orbits the Earth on a Sun-synchronous orbit (Crisp et al., 2012; O’Dell et al., 2012). The footprint area of each measurement is roughly 1.29 by 2.25 kilometers, but the data is very sparse in time and in space. The satellite completes 14.57 revolutions around Earth overpasses in one day. In the presence of clouds, the satellite is not able to produce measurements, and this poses a challenge for areas with persistent cloud covers, such as Northern Europe in the winter.

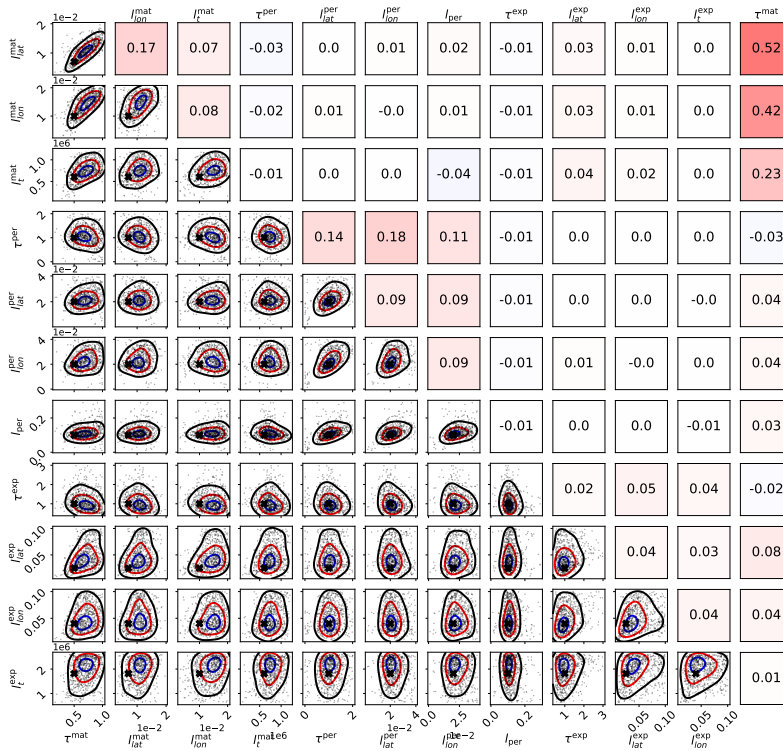


Figure 5. Scaled MCMC posteriors from a synthetic study showing identifiability of multi-scale Gaussian process kernel parameters. On lower left, the pairwise marginal distributions are shown, with the black crosses denoting true values. The axis labels are on the left and below the figure. On upper right, sample correlations are shown, with axis labels on the left and on the top. Small within-kernel component positive covariances are present. The contours shown include 85% (black), 50% (red) and 15% (blue) of the posterior mass.

The present work utilizes the XCO₂ data, its reported uncertainties, associated coordinate information, and zonal and meridional wind speeds that are contained in the data files. Only observations flagged good are used, and there are in total 116489342 such observations for the time period considered.

4.3 Solving the mean function for OCO-2 v9

- Solving the mean function from OCO-2 v9 XCO₂ data, as described in Sect. 2.4, produces best estimates for the coefficients of Eq. (11) shown in Fig. 6. The upper left quadrant shows the semiannual seasonality of the XCO₂ concentration, which explains the color shift along the equator. The lower left quadrant shows the amplitude of the twice faster oscillations, and like β_1 , also β_2 shows the highest amplitude oscillations in the boreal region.



The constant term β_3 in the upper right quadrant shows the background concentration. Some of the reddest areas such as East China, both coasts of the United States, Central Europe, and the Persian Gulf stand out and are areas where major emission sources are known to exist. The observation of a local elevated concentration compared to the surrounding areas approaches the work of Hakkarainen et al. (2016), where empirically defined time-integrated local XCO₂ anomalies are interpreted as possible emission sources.

The phase shift is modeled separately, and the field in the lower right quadrant is obtained by optimization, conditioning on the β factors. This partly explains the slightly different spatial pattern. The figure shows how the phases of the XCO₂ annual cycles differ in some regions, such as the Amazon or the Central African rain forests and the Sahel. The trend component β_4 was here set to be constant, as CO₂ over time mixes in the atmosphere.

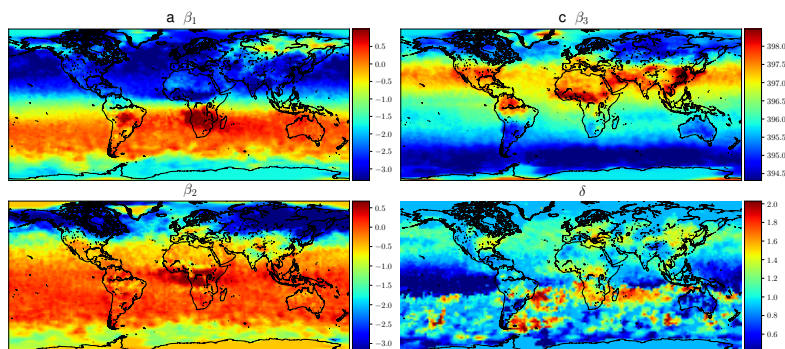


Figure 6. Mean values of mean function coefficients that were described as a Markov random field, calculated in a $2^\circ \times 2^\circ$ grid. The β_i coefficients multiply the f_i functions in Eq. (11). Panel (d) shows how the phase parameter δ can vary more in the southern hemisphere where β_1 and β_2 are small. The mean function and fitted daily means for several locations with the corresponding mean function parameters are shown in Fig. 1.

10 4.4 Covariance parameters of the OCO-2 v9 data

The OCO-2 data has several natural length scales, both spatially and temporally. The distance between adjacent observations is only one to two kilometers in space and some hundredths of a second in time, but the distance between consecutive orbits is thousands of kilometers in space and several hours in time. On consecutive days the satellite passes close to the trajectory of the previous day at a distance of tens to three hundred kilometers depending on the latitude. The Earth has natural temporal diurnal and annual cycles, but since OCO-2 is Sun-synchronous, only the latter matters. Since the annual cycle is already fitted in the particular form of the mean function used, Eq. (11), a periodic kernel component is not included, and the data is modeled with a kernel consisting of a larger-scale exponential and smaller scale Matern component.

The covariance parameters for the two-component kernel, which are the median values from sampling the posterior with MCMC, are given in Table 3. With learning the parameters from a data set with natural length scales, the posterior may appear



multi-modal, with some of the modes only having relatively little mass. In such a case, the median provides a more robust estimate for the parameters than the mean. The $\ell_{lon}^{(\cdot)}$ and $\ell_t^{(\cdot)}$ parameters of the posterior mean were slightly larger, which would result in slower computation. Selecting the median is further justified by the slight overestimation of some parameters in the synthetic study in Sect. 4.1.

- 5 Learning the covariance parameters from OCO-2 v9 data used the following configuration parameters for satGP: $\zeta_{train} = 0$, $\kappa = 256$, and $n_{ref} = 12$. A total of 1.1184 million MCMC iterations were completed, with the first 50% discarded as burn-in to produce statistics. The reference points were randomly picked from a rectangle with corners at (0°S, 65°E) and (60°N, 145°E). While using the whole globe would have been a principled choice, MCMC requires lots of iterations, and for any claim of global coverage n_{ref} would have needed to be much larger.

Table 3. Covariance function parameter values learned from OCO-2 data. First column shows the Matern kernel parameters, and the second column the exponential kernel parameters. The length scale along the parallels, $\ell_{lon}^{(\cdot)}$ is much larger than that along the meridians, $\ell_{lat}^{(\cdot)}$.

	(\cdot) = mat	(\cdot) = exp
$\tau^{(\cdot)}$	0.899	2.72
$\ell_{lat}^{(\cdot)}$	0.00513	0.0418
$\ell_{lon}^{(\cdot)}$	0.0363	0.397
$\ell_t^{(\cdot)}$	20h 22min	16d 20h 12min

10 4.5 Posterior predictive distributions of XCO2 from the OCO-2 v9 data

The marginal posterior predictive distribution at test points x^* , given by Eq. (9) and (10), were calculated globally in a half-degree grid between 80°S and 80°N at a daily resolution. The first day of simulation was September 6 2014, and the last day was November 10 2018, spanning in total 1526 days. For each day, 230400 marginals were computed, resulting in a collective 351 million inverted covariance matrices. The satGP parameters used were $\zeta_{sample} = 0$ and $\kappa = 256$, and the covariance kernel used was the one learned in Sect. 4.4, with parameters given in Table 3. The simulation time was 25 days on a moderately fast Intel i7-8700K CPU utilizing the available 12 CPU threads and 32 GiB memory.

Global fields of the mean values and marginal uncertainties are presented in Fig. 7 and 8, with a subset (to avoid excessive over-drawing) of observations shown as a scatter plot. For this simulation, a maximum distance of 1100 km (10° on the equator) was specified for speeding up searching for closest observations in the direction along parallels. This constraint can be seen as discontinuities in uncertainty when no observations are nearby, especially close to the poles. The (b) parts of the figures show how uncertainty is reduced with the overpass of OCO-2. This uncertainty reduction diminishes fast due to the Matern component of the multi-scale kernel having a very short length scale parameter in the time dimension. In the upper figures, the background color (posterior mean) usually matches the observations. Due to observational noise, the GP mean is not strictly interpolation, however.

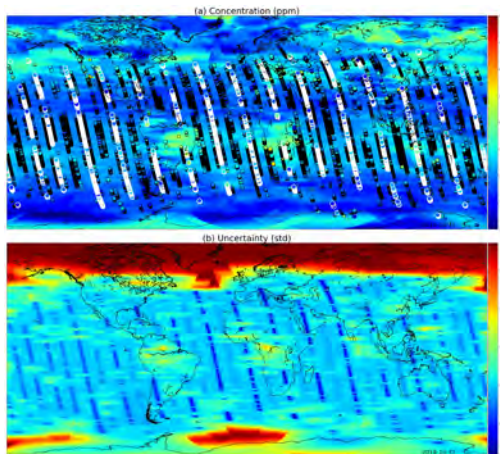


Figure 7. Global XCO₂-distribution posterior mean values (a) and their uncertainties (b) on last day of October 2014. The most informative observations are shown with the concentrations, with the large white circles being from October 31st 2014, medium circles from one day before or after, and small circles from two days before or after. The OCO-2 utilizes sunlight for retrieval, and that is why there are very few observations above 60°N.

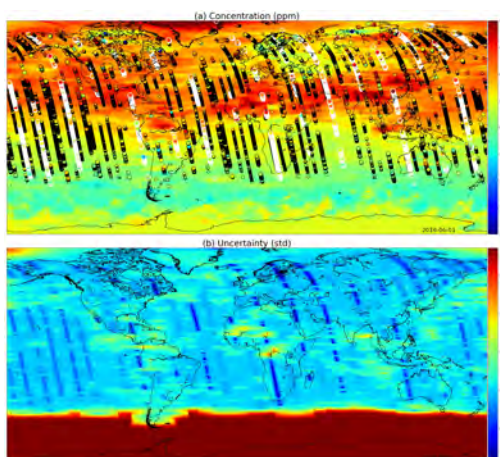


Figure 8. Global XCO₂-distribution posterior mean values (a) and their uncertainties (b) on June 1st 2016. While photosynthesis in the Northern Hemisphere is already reducing the carbon dioxide concentrations globally, the observations condition the Gaussian process to higher mean values than in Fig. 7. In the summer months the uncertainty stays high close to the South Pole.



4.6 Comparison of single- and multi-scale kernels with OCO-2 data

How the multi-scale kernel formulation affects the predictive posterior distributions can be demonstrated with OCO-2 data. In Fig. 9 posterior marginals from September 15 2014 are shown. The first row (a-b) contains results from the multi-scale kernel described in Sect. 4.4, and the second row (c-d) shows fields from only the exponential part of the multi-scale kernel. The parameters of the multi-scale kernel are shown in Table 3. The bottom row (e-f) contains the difference fields between the first and the second rows. The single-kernel uncertainty is very low in Fig. 9 (d) since lots of observations fall into regions of high covariance with almost any test input, with the exception of the Northern side of Ireland, which does not have any observations nearby. Since the covariance kernel parameters were trained for the multi-scale kernel, the parameters used for the single kernel are not the ones describing the XCO₂ field best.

Figure (a) shows that as intended, the multi-scale approach leads to local enhancements of the XCO₂ mean field. Far from the measurements, the smaller Matern kernel no longer reduces the predicted marginal uncertainties, and this leads to an increase in uncertainty in these areas. Figure (e) shows additional enhancements of the XCO₂ mean fields, which are in this case due to the different maximum covariances between the multi-scale and single-scale kernels.

The total kernel size was kept at 1024 ($\kappa = 512$ for (a-b) and $\kappa = 1024$ for (c-d)) in both experiments. Additionally $\zeta_{\text{sample}} = 5$, and $\omega = 0.5^\circ$ in this case. The very same observations were used in both cases.

4.7 Wind-informed kernel with OCO-2 data

The wind-informed kernel, Eq. (16), lets local wind data at test input x^* rotate and scale the coordinate axes. Modeled winds are included with OCO-2 data, and they can be used to produce gridded winds that can then be used locally with the computation of each marginal posterior predictive distribution.

The covariance parameters for a single wind kernel were learned by taking the median of an MCMC posterior, similarly as was done in Sect. 4.4. The resulting parameters were $\tau = 2.07$, $\ell = 0.038$, and $\rho = 56.7$. The variance of ρ was high, possibly due to the square root in the current formulation in Eq. (17). For this simulation, $\zeta = 1$, $\kappa = 1024$, and $\omega = 0.7$, and the simulation time for the area from $(27^\circ N, 115^\circ E)$ to $(40^\circ N, 145^\circ E)$ for the single day was 2.652s (walltime) on the i7-8750H laptop CPU.

The simulation results are shown in Fig. 10. Low uncertainties shown in blue color on the right spread with the winds, as do the concentration estimates on the left both due to the high reading in South Korea and the low reading close to Shanghai.

Optimally the wind-informed kernel should utilize winds that are not recomputed from the observations as was done for convenience, but directly from a weather or climate model. The satGP program contains configuration options for doing this. The optimal covariance function parameter values are conditional on the wind data, so the values should be learned separately for each new application and wind data set.

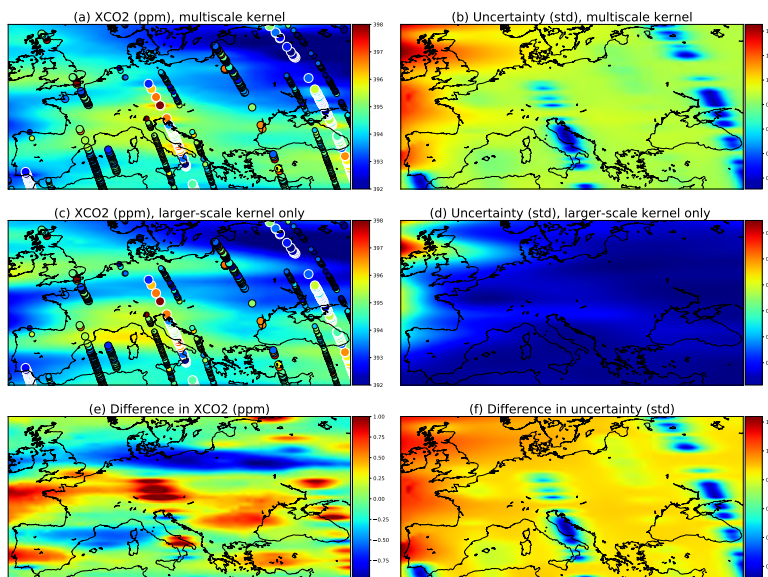


Figure 9. Comparison of a multi-scale kernel with the two components described in Sect. 4.4 and a single component kernel defined by the parameters of the exponential kernel. These parameters were given in Table 3. The observations used are the same and are shown in panels (a) and (c) as circles. The large ones with white borders are observations from the present day, September 15th 2014, medium circles are observations from 14th and 16th, and small circles from 13th and 17th.

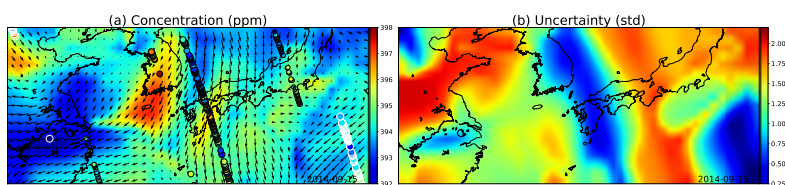


Figure 10. (a) GP posterior mean of XCO2 and (b) its uncertainties with the wind-informed kernel. The area shown contains the Korean peninsula in the center, China on the left, and Japan on the center-right. The large circles with the white edges are present-day observations, medium circles are observations from adjacent days, and the smallest ones are observations from two days away. Wind direction and magnitude are given by the black arrows, and uncertainty is clearly reduced where wind is blowing directly towards or away from the observations.



5 Conclusions and future work

In this work we have introduced the first version of a new fast Gaussian process software, satGP v.0.1. It aims at being a general purpose Gaussian process toolbox, especially meant to be used with remote sensing data. The software solves the problems of spatially varying mean function, learning its parameters via computation of marginals of an MRF, and also allows learning the parameters of the multi-scale covariance function using either optimization algorithms or adaptive Markov chain Monte Carlo. On top of these, satGP allows to conduct synthetic parameter identification studies via sampling Gaussian process prior and posterior distributions, and this can be done with any kernel prescribed, including a non-stationary wind-informed kernel. We are not aware of open source remote sensing-oriented software that would provide this combination of features. The satGP program was demonstrated with the enormous data set produced by the NASA Orbiting Carbon Observatory 2.

There are various aspects of satGP that could be improved in future versions. These include addition of routines for doing model selection to select the components of the multi-scale kernel, improving the observation selection/thinning scheme for statistical optimality, and finding joint posterior predictive distributions. For the last one, a multi-grid version can be developed, and this could be potentially useful for flux inversion studies.

The satGP software utilizes various approximations for computational tractability, and the connection between parameters such as length scales ℓ , thinning parameter ζ , maximum kernel size κ , and prediction accuracy could be studied further, as well as changing the grid resolution according to density of observations.

The methodology and code presented can be also used with other data sources. For instance, combining data from the various satellites that measure CO₂, such as GOSAT, GOSAT-2, OCO-2, TANSAT, and the OCO-3, would be particularly interesting. That more and more instruments are about to provide data from the orbit in the near future will lead to a need to understand the properties of even larger data sets.

Code and data availability. The satGP code is available from the contact author upon a reasonable request. The OCO-2 v9 data used is freely available directly from JPL/NASA.

Appendix A: Input parameters and variables in satGP

The satGP software by design allows for lots of flexibility for defining how to model the quantity of interest as a Gaussian random field. In this section the possibilities are discussed along with some recommendations. The parameters in Table 1 are described in more detail than earlier, along with some other configuration variables in the configuration file `config.h`. Practical aspects of defining mean functions and covariance kernels are also included. Some of the details in this section may change for future versions of satGP.

Of the four sections in Table 1, the first is obvious, as those parameters control the main logic of satGP. It is recommended to first learn the mean function, then with that mean function learn the covariance function, and only after that calculate the means



and variances of the Gaussian process with `sampling = 1`. The setting `sampling = 2` can be used for illustration purposes, for understanding how the different realizations of the random function would look like.

The `area` parameter defines the longitude-latitude extents of the domain where `satGP` is wished to be used. The strings and the corresponding areas are defined in the beginning of the file `gaussian_proc.h`, and can be changed there as needed.

5 Current available areas contain e.g. `NorthAmerica`, `Europe`, `EastAsia`, `World`, and `TESTAREA`.

The parameter `n_days` defines how many days are to be simulated after the starting day. Currently the starting day is hardcoded in the code base to the first day of OCO-2 data. However, if `use_daylist ≠ 0` in the configuration file, a list of days can be used. This list can quite easily generated by modifying a python script `create_daylist.py`, which is included with `satGP`.

The ω parameter determines how much spatial detail is resolved when sampling or computing marginals of the random field.
10 A small value like 0.1 will make computing very expensive, and using such values might be unnecessary when the smallest covariance subkernel length scale parameters are large. These ℓ parameters are in the scale of distances on the unit ball, and therefore on the equator an ℓ parameter of 0.05 corresponds to a length scale of around 2.9° , so the ω parameter should rarely be much less than half of that. On the other hand, if the observations are spatially very close to each other and describing local variation is aimed for, then the ℓ parameters need also to be small. Given computational constraints, larger values or different
15 `area` parameters may need to be used.

In the third section of Table 1, the first parameter `n_ker` denotes the number of subkernels. Even though the hard limit is set at 10, in practice this should be between one and three since the parameters of more than three subkernels are not necessarily identifiable. More kernels means also more computational cost, due to the κ parameter, which is the last one in the table and discussed later.

20 The parameters `cfc` and `mf` are not strictly input variables, but C struct pointers that are created based on input variables. These variables are described in the configuration file, and they amount to choosing the covariance kernels from prescribed types (e.g. Matern, exponential, and periodic), and then defining the parameters for those kernels. The best parameters are those that are learned with `learn_k = 2` when non-synthetic data is used.

The learning is best performed with MCMC, and the posterior mean and median have proven to be a useful values. For
25 unimodal posterior distributions these values are very close. The number of MCMC iterations is controlled by the variable `mcmc_iters`, for which 10^6 is a large enough value, and for computing the log-likelihood in Eq. (22), the number of reference points `n_ref` in the set E_{ref} can be set to a low value of e.g. number of CPU threads, if at least 12 are available. If with MCMC the chain gets stuck in local minima, the value of the `mcmc->scalefactor` in the `mcmc()` function in `mcmc.h` may be shrunk, and equally well, if the posterior ends up being flat with respect to many parameters, it may be increased.

30 For learning the covariance parameters, parameter limits need to be given. These should correspond to the expected length scales in the data – e.g. long-range fluctuations with low amplitude, and short-scale variations due to local effects. It is in practice best if the parameter ranges do not overlap.

If the exponent of the exponential kernel needs to be changed, that needs to be done by changing the `exponent` variable in the `covfun_dyn()` function in the file `covariance_functions.h`. Similarly, if the order of the Matern kernel needs to



be changed, that can be done by changing the variable `n` in functions `covfun_matern52()` and `initialize_covfunconfig()` in that same file.

For constructing the mean function, the configuration file contains the parameter `mftype`. The possible values are: 0) a zero mean function is used, 1) a mean function that changes only in time is used, 2) a (time-dependent) field is read in and used - this can be e.g. the mean value from a previous Gaussian process simulation, and 3) a space and time dependent mean function is used. The function itself is given as a function pointer to variable `mean_function` in the configuration file, and this function needs to be defined somewhere - e.g. in the file `mean_functions.h`. For the mean function, another variable, `mfccoeff`, needs to be set. This is the total number of parameters (β and δ in Eq. (3)) if `mftype` \in $\{1, 3\}$. If the mean function parameters are learned, the parameter `nnonbetas`, the number of mean function non-linear δ parameters, needs to be set to the appropriate value in the function `fit_beta_parameters_with_unc()` in `mean_functions.h`. For global mean function coefficients, the values of those coefficients are given in the configuration file. Additionally, parameter limits for learning the space-dependent mean function parameters are set in the configuration file. Finally, when learning the space-dependent mean function parameters, the smoothness of the field may be controlled by changing the `dscale` parameter in the configuration file, and to a lesser extent by modifying the `dfmin` and `dfmax` parameters in function `fit_beta_parameters_with_unc()` in file `mean_functions.h`.

In the last section, the ζ_{train} parameter controls data thinning when learning covariance kernel parameters and the ζ_{sample} parameter has the same effect for when `sampling` \neq 0. How the thinning takes place was explained in the context of Eq. (24). While with few observations no thinning needs to be done at all, i.e. ζ may be set to zero, with large data sets the representability of data may be improved when a coarse grid is used for computation, and also memory bottlenecks may be avoided. These parameters may be also increased if faster execution is required, e.g. for debugging purposes.

The σ_{min}^2 parameter controls which observations are not considered at all when computing at a location x^* , as described by Eq. (19). The higher this is, the more data is discarded. Setting σ_{min}^2 to a very low value makes searching for candidate observations slow, while picking too high a value may make posterior fields look edgy. In practice values between 10^{-7} and 10^{-3} seem to work well. This parameter is not actually meant to be changed, and it is for that reason set in `create_config()` in the file `gaussian_proc.h`.

The variable `nsynthetic` defines how many synthetic observations are generated when `learn_k` = 1. Very large values are once again expensive, and instead a smaller `area` should rather be used with more moderate values of `nsynthetic`. Those values can be in practice up to 10^5 or more. With very low values, it may be that spatial patterns specified by the prescribed covariance kernel are not represented appropriately, and therefore values less than 10^4 should be avoided, except for maybe in settings with only a single subkernel. If $\sigma_{\text{synthetic}}^2$ is high, parameter identifiability suffers. Varying this parameter could be used for understanding how complex a multi-scale kernel can be useful with particular data sets. The values also depend on the maximum covariance parameters of the Gaussian process, given by the τ^2 parameters in the formulas of Sect. 2.5.

The last parameter in Table 1, κ , defines the maximum subkernel size. The larger this parameter is, the more data is included for constructing the covariance matrix K , whose Cholesky decomposition needs to be computed to solve the local regression problem inherent to Gaussian processes. In practice the full kernel size should be kept under 1000, and in order to compute GP



calculations fast, a full kernel size of less than 500 is recommended. However, with a very small number of marginals, values up to 10^4 may be experimented with. When $n_{\text{ker}} < 64$, the speed-up due to solving the GP formulas faster decreases, since at that point computing Cholesky decompositions no longer takes up majority of computing time. This lower bound depends on the CPU architecture and the sizes of the various CPU caches.

5 Whether the observations for computing the local values are chosen at random or greedily is determined by the variable `select_closest` in function `pick_observations()` in file `covariance_functions.h`. The value used should normally be non-zero, since with random selection adjacent grid points often do not utilize the best available observations closest by, leading to noisiness or graininess in the computed mean field.

10 In addition to the parameters and variables listed here, there are also other parameters in the configuration file and in the code, even though those should not need to be changed. Any variables that the user might want to tweak are generally accompanied by at least some comments describing their effects.

15 In the current version, the satGP program is run with the script `gproc.sh`, whose comments describe the various options. Compiling and running require a modern GCC version (such as version 8) and the meson build system, and additionally all the needed libraries listed in Sect. 3. The current low version number reflects the fact that as of now, installing and using the software will require a degree of technical knowledge, including some Python, C, and BASH programming skills.

Author contributions. JS, AS, HH, and YM designed the study. JS prepared this manuscript, wrote the satGP code, chose, tested and implemented the computational methods, and performed the simulations.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements.



References

- Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., and O'Neil, M.: Fast Direct Methods for Gaussian Processes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38, 252–265, <https://doi.org/10.1109/TPAMI.2015.2448083>, 2016.
- Chiles, J.-P. and Delfiner, P.: *Geostatistics*, Wiley, 2012.
- 5 Cressie, N.: *Mission CO2ntrol: A Statistical Scientist's Role in Remote Sensing of Atmospheric Carbon Dioxide*, *Journal of the American Statistical Association*, 113, 152–168, <https://doi.org/10.1080/01621459.2017.1419136>, 2018.
- Cressie, N. and Wikle, C.: *Statistics for Spatio-Temporal Data*, Wiley, 2001.
- Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J.,
10 O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO₂ retrieval algorithm; Part II: Global XCO₂ data characterization, *Atmospheric Measurement Techniques*, 5, 687–707, <https://doi.org/10.5194/amt-5-687-2012>, <https://www.atmos-meas-tech.net/5/687/2012/>, 2012.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D.: *Bayesian Data Analysis*, Chapman and Hall/CRC, 3rd edn., 2013.
- 15 Haario, H., Saksman, E., and Tamminen, J.: An Adaptive Metropolis Algorithm, *Bernoulli*, 7, 223–242, <http://www.jstor.org/stable/3318737>, 2001.
- Hakkaraian, J., Ialongo, I., and Tamminen, J.: Direct space-based observations of anthropogenic CO₂ emission areas from OCO-2, *Geophysical Research Letters*, 43, 11,400–11,406, <https://doi.org/10.1002/2016GL070885>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL070885>, 2016.
- 20 Hammerling, D. M., Michalak, A. M., O'Dell, C., and Kawa, S. R.: Global CO₂ distributions over land from the Greenhouse Gases Observing Satellite (GOSAT), *Geophysical Research Letters*, 39, <https://doi.org/10.1029/2012GL051203>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL051203>, 2012.
- Hammersley, J. and Clifford, P.: *Markov random fields on finite graphs and lattices*, unpublished manuscript, 1971.
- Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M.,
25 Lindgren, F., Nychka, D. W., Sun, F., and Zammit-Mangion, A.: A Case Study Competition Among Methods for Analyzing Large Spatial Data, *Journal of Agricultural, Biological and Environmental Statistics*, <https://doi.org/10.1007/s13253-018-00348-w>, 2018.
- IPCC: *Summary for Policymakers*, book section SPM, pp. 1–30, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/CBO9781107415324.004>, www.climatechange2013.org, 2013.
- Johnson, S. G.: *The NLOpt nonlinear-optimization package*, <http://github.com/stevengj/nlopt>, 2014.
- 30 Katzfuss, M., Guinness, J., and Gong, W.: Vecchia approximations of Gaussian-process predictions, *arXiv e-prints*, [arXiv:1805.03309](https://arxiv.org/abs/1805.03309), 2018.
- Lauritzen, S.: *Graphical Models*, Oxford Statistical Science Series, Clarendon Press, 1996.
- Lindgren, F., Rue, H., and Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 423–498, <https://doi.org/10.1111/j.1467-9868.2011.00777.x>, <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x>, 2011.
- 35 Ma, P. and Kang, E. L.: Fused Gaussian Process for Very Large Spatial Data, *ArXiv e-prints*, 2017.



- Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B. A., and Crisp, D.: Quantifying CO₂ Emissions From Individual Power Plants From Space, *Geophysical Research Letters*, 44, 10,045–10,053, <https://doi.org/10.1002/2017GL074702>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074702>, 2017.
- Neal, R. M.: MCMC using Hamiltonian dynamics, in: *Handbook of Markov Chain Monte Carlo*, edited by Brooks, S., Gelman, A., Jones, G., and Meng, X., Chapman & Hall/CRC Handbooks of Modern Statistical Methods, CRC Press, 2011.
- Nguyen, H., Katzfuss, M., Cressie, N., and Braverman, A.: Spatio-Temporal Data Fusion for Very Large Remote Sensing Datasets, *Technometrics*, 56, 174–185, <https://doi.org/10.1080/00401706.2013.831774>, 2014.
- O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Crisp, D., Eldering, A., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: Corrigendum to "The ACOS CO₂ retrieval algorithm - Part 1: Description and validation against synthetic observations" published in *Atmos. Meas. Tech.*, 5, 99–121, 2012, *Atmospheric Measurement Techniques*, 5, 193–193, <https://doi.org/10.5194/amt-5-193-2012>, <https://www.atmos-meas-tech.net/5/193/2012/>, 2012.
- Rasmussen, C. and Williams, C.: *Gaussian Processes for Machine Learning*, MIT Press, <http://www.gaussianprocess.org/gpml/chapters/>, 2006.
- Rodgers, C.: *Inverse Methods for Atmospheric Sounding: Theory and Practice*, Series on atmospheric, oceanic and planetary physics, World Scientific, 2000.
- Santner, T., Williams, B., and Notz, W.: *The Design and Analysis of Computer Experiments*, Springer Verlag New York, first edn., 2003.
- Schäfer, F., Sullivan, T. J., and Owhadi, H.: Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity, arXiv e-prints, arXiv:1706.02205, 2017.
- Tadić, J. M., Qiu, X., Miller, S., and Michalak, A. M.: Spatio-temporal approach to moving window block kriging of satellite data v1.0, *Geoscientific Model Development*, 10, 709–720, <https://doi.org/10.5194/gmd-10-709-2017>, <https://www.geosci-model-dev.net/10/709/2017/>, 2017.
- Vecchia, A. V.: Estimation and Model Identification for Continuous Spatial Processes, *Journal of the Royal Statistical Society. Series B (Methodological)*, 50, 297–312, <http://www.jstor.org/stable/2345768>, 1988.
- Yi, L., Jing, W., Lu, Y., Xi, C., Zhaonan, C., Dongxu, Y., Zengshan, Y., Songyan, G., Longfei, T., Naimeng, L., and Daren, L.: TanSat Mission Achievements: from Scientific Driving to Preliminary Observations, *Chinese Journal of Space Science*, 38, 627, <https://doi.org/10.11728/cjss2018.05.627>, http://www.cjss.ac.cn/EN/abstract/article_2600.shtml, 2018.
- Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., and Maksyutov, S.: Global Concentrations of CO₂ and CH₄ Retrieved from GOSAT: First Preliminary Results, *SOLA*, 5, 160–163, <https://doi.org/10.2151/sola.2009-041>, 2009.
- Zammit-Mangion, A., Cressie, N., Ganesan, A. L., O'Doherty, S., and Manning, A. J.: Spatio-temporal bivariate statistical models for atmospheric trace-gas inversion, *Chemometrics and Intelligent Laboratory Systems*, 149, 227 – 241, <https://doi.org/https://doi.org/10.1016/j.chemolab.2015.09.006>, 2015.
- Zammit-Mangion, A., Cressie, N., and Shumack, C.: On Statistical Approaches to Generate Level 3 Products from Satellite Remote Sensing Retrievals, *Remote Sensing*, 10, <https://doi.org/10.3390/rs10010155>, <http://www.mdpi.com/2072-4292/10/1/155>, 2018.
- Zeng, Z., Lei, L., Guo, L., Zhang, L., and Zhang, B.: Incorporating temporal variability to improve geostatistical analysis of satellite-observed CO₂ in China, *Chinese Science Bulletin*, 58, 1948–1954, <https://doi.org/10.1007/s11434-012-5652-7>, 2013.
- Zeng, Z.-C., Lei, L., Strong, K., Jones, D. B. A., Guo, L., Liu, M., Deng, F., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Henderson, B., Kivi, R., Lindenmaier, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Sussmann, R., Velasco, V. A., Wennberg, P. O.,

<https://doi.org/10.5194/gmd-2019-156>
Preprint. Discussion started: 30 August 2019
© Author(s) 2019. CC BY 4.0 License.



and Lin, H.: Global land mapping of satellite-observed CO₂ total columns using spatio-temporal geostatistics, *International Journal of Digital Earth*, 10, 426–456, <https://doi.org/10.1080/17538947.2016.1156777>, 2017.

PAPER II

Jouni Susiluoto, Maarit Raivonen, Leif Backman, Marko Laine, Jarmo Makela, Olli Peltola, Timo Vesala, Tuula Aalto

Calibrating the sqHIMMELI v1.0 wetland methane emission model with hierarchical modeling and adaptive MCMC

In *In Geoscientific Model Development*, 11, 1199-1228, 2018.
www.geosci-model-dev.net/11/1199/2018/

Copyright © Authors 2018. CC Attribution 3.0 License. Reprinted with permission.



Calibrating the sqHIMMELI v1.0 wetland methane emission model with hierarchical modeling and adaptive MCMC

Jouni Susiluoto^{1,3,4}, Maarit Raivonen², Leif Backman^{1,2}, Marko Laine¹, Jarmo Makela^{1,4}, Olli Peltola², Timo Vesala^{2,5}, and Tuula Aalto¹

¹Climate research, Finnish Meteorological Institute, P.O. Box 503, Helsinki, Finland

²Institute for Atmospheric and Earth System Research/Physics, Faculty of Science, University of Helsinki, Helsinki, Finland

³Lappeenranta University of Technology, School of Science, Lappeenranta, Finland

⁴University of Helsinki, Department of Mathematics and Statistics, Helsinki, Finland

⁵Institute for Atmospheric and Earth System Research/Forest Sciences, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, Finland

Correspondence: Jouni Susiluoto (jouni.susiluoto@fmi.fi)

Received: 12 March 2017 – Discussion started: 21 April 2017

Revised: 25 February 2018 – Accepted: 1 March 2018 – Published: 29 March 2018

Abstract. Estimating methane (CH₄) emissions from natural wetlands is complex, and the estimates contain large uncertainties. The models used for the task are typically heavily parameterized and the parameter values are not well known. In this study, we perform a Bayesian model calibration for a new wetland CH₄ emission model to improve the quality of the predictions and to understand the limitations of such models.

The detailed process model that we analyze contains descriptions for CH₄ production from anaerobic respiration, CH₄ oxidation, and gas transportation by diffusion, ebullition, and the aerenchyma cells of vascular plants. The processes are controlled by several tunable parameters. We use a hierarchical statistical model to describe the parameters and obtain the posterior distributions of the parameters and uncertainties in the processes with adaptive Markov chain Monte Carlo (MCMC), importance resampling, and time series analysis techniques. For the estimation, the analysis utilizes measurement data from the Siikaneva flux measurement site in southern Finland.

The uncertainties related to the parameters and the modeled processes are described quantitatively. At the process level, the flux measurement data are able to constrain the CH₄ production processes, methane oxidation, and the different gas transport processes. The posterior covariance structures explain how the parameters and the processes are related. Additionally, the flux and flux component uncertain-

ties are analyzed both at the annual and daily levels. The parameter posterior densities obtained provide information regarding importance of the different processes, which is also useful for development of wetland methane emission models other than the square root Helsinki Model of Methane build-up and emission for peatlands (sqHIMMELI).

The hierarchical modeling allows us to assess the effects of some of the parameters on an annual basis. The results of the calibration and the cross validation suggest that the early spring net primary production could be used to predict parameters affecting the annual methane production.

Even though the calibration is specific to the Siikaneva site, the hierarchical modeling approach is well suited for larger-scale studies and the results of the estimation pave way for a regional or global-scale Bayesian calibration of wetland emission models.

1 Introduction

Methane is the third most important gas in the atmosphere in terms of its capacity to warm the climate, after water vapor and carbon dioxide, currently with the radiative forcing of 0.97 W m⁻² (IPCC, 2013). This is a sizable part of the total effect of well-mixed greenhouse gases, which is approximately 3.0 W m⁻². According to IPCC (2013), the amount of CH₄ in the atmosphere has risen to its highest level in at

least the last 800 000 years due to human activity, and based on ice core measurements, also its growth rate is presently very likely at its highest level in the last 22 000 years.

The sources of CH₄ are both anthropogenic and natural. In the years 2003–2012, 60 % of the global emissions were anthropogenic (range 50–65 %), and about one-third came from natural wetlands. The most important source of uncertainty in the global methane budget is attributable to emissions from wetlands and other inland waters. Combining top-down and bottom-up estimates, natural wetland emissions range from 127 to 227 Tg CH₄ yr⁻¹ (Saunois et al., 2016). Anthropogenic sources include rice paddies, landfills, enteric fermentation and manure, incomplete combustion of hydrocarbons, and natural gas leaks (Ciais et al., 2013).

The methane from wetlands is produced by prokaryotic archaea under anaerobic conditions. The main sink for atmospheric CH₄ is its oxidation in troposphere by OH, and the average lifetime of a CH₄ molecule in the atmosphere is 9.1 ± 0.9 years (Prather et al., 2012; IPCC, 2013).

The wetlands in the boreal zone are a significant contributor to the total CH₄ emissions from wetlands (Kirschke et al., 2013), and for this reason the CH₄ emissions from them have been intensively studied, also with models, during the past years (Wania et al., 2010; Kaiser et al., 2016; Petrescu et al., 2015). However, major discrepancies between predictions from those models remain (Melton et al., 2013; Bohn et al., 2015).

The need for improved wetland methane emission modeling is amplified by the fact that although annual mean precipitation is projected to increase in the boreal zone (Ruosteenoja et al., 2016), changes in the frequency and duration of severe drought may follow an alternate path (Lehtonen et al., 2014), manifesting the need to study wetland responses to extreme events.

Changes to hydrological conditions such as draining or recurring low water table depth can alter the balance of greenhouse gas emissions (Frolking et al., 2011; Petrescu et al., 2015). Modeling and calibrating for such exceptional events can be difficult, as was found, for instance, by Mäkelä et al. (2016).

The Helsinki Model of Methane build-up and emission for peatlands (HIMMELI) is a relatively full-featured wetland/peatland CH₄ emission model and it is described in detail in Raivonen et al. (2017). The model contains process descriptions for CH₄ production from anaerobic respiration, O₂ consumption and CO₂ production from oxic respiration, and gas transport processes via diffusion, ebullition, and plant transport. Modeling the concentrations of CH₄, O₂, and CO₂ in the peat column is explicitly included. The peat column depth can be set at any desired value, and the water table movement determines the part of the peat column that is favorable for CH₄ production. The version of HIMMELI in this work has additional processes, described in Sect. 3.1, and the modified model is referred to as sqHIMMELI (square root HIMMELI), as it contains a description of CH₄ production

from root exudates. The sqHIMMELI model is geared towards site-level studies, whereas HIMMELI is more suited for integration directly as a component in, e.g., land surface schemes.

Even well-constructed computer models describing environmental processes accumulate error at many levels (Sanso et al., 2008). The sources include time and space discretization, compromises in model physics and biochemistry descriptions due to computational constraints, insufficient information about the initial states of the model, and numerical errors, along with parameterization-induced inaccuracies of the subgrid-size processes. This leads to a need to calibrate and optimize models, as the physical variables do not necessarily exactly correspond to the model variables, and hence the model parameters cannot often be directly measured. Of course, any physically insightful interpretation of calibration results makes sense only for a well-constructed physical model.

Several current CH₄ models include the important physical processes controlling both CH₄ production and transport in the peat column (Kaiser et al., 2016; Lai, 2009b; Müller et al., 2015; Grant and Roulet, 2002). The modeled peat column depth affects the total modeled CH₄ emission from the peatlands and it is directly included in some models (Lai, 2009b; Walter and Heimann, 2000). These models are in general highly sensitive to changes in the values of the parameters (van Huissteden et al., 2009). However, even though algorithmic parameter optimization has been done in some studies, the stress is often on parameter efficiencies (van Huissteden et al., 2009) or optimal values (Müller et al., 2015), and hence the full uncertainty of the values of parameters in these models is not well understood.

Methane models typically use measured values from field campaigns and parameters estimated from those studies where applicable (Lai, 2009b; Walter and Heimann, 2000; Tang et al., 2010; Riley et al., 2011), and, when needed, include extra tuning parameters for processes (Walter and Heimann, 2000). This is a practical and much-used route, as information regarding all of the needed parameters is not available at all sites (van Huissteden et al., 2009; Walter and Heimann, 2000). Wide variability can be expected from some parameters, such as those controlling CH₄ oxidation (Segers, 1998). Emissions from different areas of the same wetland can also vary, due to microtopography and differences between how fast the peat decomposes in different areas (Lai, 2009a; Cresto Aleina et al., 2016), making straightforward parameter value assignment difficult.

Due to these uncertainties, values of parameters vary widely from research to research. For instance, for the Q_{10} value controlling the temperature dependence of CH₄ production, Walter and Heimann (2000) use the value 6, hand-picking it from the interval of 1.7–16, whereas van Huissteden et al. (2009) use a range of 3–8, and Müller et al. (2015) constrain the value between 1 and 10, with the default value of 1.33 and eventually optimize it to the value of 1 for two

of the three optimizations presented. For other parameters, such as those controlling diffusion rates in peat, the situation is similar.

Calibration done for the models is usually quite basic. Wania et al. (2010) tune their model by running it with parameters from a parameter grid, containing only three values for each of the seven parameters tested, and Riley et al. (2011) follow a similar procedure for the wetland CH₄ model component, CLM4Me, of the Community Land Model. Such sensitivity studies obviously are not able to find out how a model is able to perform at its best. Müller et al. (2015) have further optimized the CLM4Me model using an emulator combined with a simple minimization algorithm, with respect to several different sites, which are bound to have quite different physical characteristics, and are yielding optimal values often at the borders of the prescribed allowed area of variation. In a sensitivity analysis of the PEATLAND-VU model, a derivative of the Walter Heimann model, van Huissteden et al. (2009) look at the efficiencies of the different parameters but do not elaborate on other qualities of the posterior.

Using hierarchical modeling to estimate annually varying parameters is sensible, since the flux measurement site has both properties that change from year to year (e.g., small changes in vegetation, plant roots, and microbe populations) and properties that are more permanent (e.g., peat quality and plant species). With fixed parameter values for all years, the model sometimes does not accurately and appropriately describe the observations. On the other hand, with different parameters for all the years, the parameters are easily overfitted, meaning that while the resulting model fits the data well, it does not accurately predict future fluxes (Gelman et al., 2013). Hierarchical modeling provides a solution for these problems.

In the present study, the sqHIMMELI model is calibrated using adaptive Markov chain Monte Carlo (MCMC) and importance resampling techniques to evaluate a hierarchical statistical model for the model parameters. The calibration is done for the boreal Siikaneva site. This study complements the work in Raivonen et al. (2017) in describing the effects of various parameters on the processes and fluxes, and analyzing what kinds of configurations best describe the studied boreal wetland.

Merely optimizing model parameters may lead to misleading results due to the presence of several local minima in the objective function; for example, Müller et al. (2015) reported in a study where they used a surrogate model to calibrate the parameters of the CH₄ model component of the Community Land Model. This multimodality can be accommodated for by using MCMC techniques. Utilizing MCMC methods for optimizing environmental models and studying their uncertainties is not new (Laine, 2008; Ricciuto et al., 2008; Hararuk et al., 2014), but to our knowledge they have not been used for wetland CH₄ model parameter estimation before. Moreover, the research that the authors are aware of

does not investigate the interannual variability of parameters, as is done in this study.

The main objective of this work is to analyze the capabilities and limitations of a modern feature-filled wetland CH₄ model by looking into the shape of the posterior parameter distributions, parameter correlations, and the roles, identifiabilities, interdependencies, and interconnections of the parameters and the processes they control. As a part of this work, knowledge about how the methane and carbon dioxide flux data are able to constrain the parameters and processes is obtained.

2 Siikaneva wetland flux measurement site and model input data

Methane and carbon dioxide flux measurements were needed for estimating the model parameters, and for that purpose observational data from the Siikaneva peatland flux measurement site in southern Finland (61°50' N, 24°12' E) were used. The site is a boreal oligotrophic fen with a peat depth of up to 4 m.

Measurement of ecosystem-scale gas fluxes started in 2005, and in this work eddy covariance (EC) CH₄ and CO₂ flux measurements from the years 2005 to 2014 were used. In the current application of the EC method, the gas fluxes were calculated from the wind speed and direction, and CH₄ and CO₂ concentration information. All these variables were sampled with 10 Hz and fluxes were calculated over 30 min averaging time in order to capture to the whole spectrum of turbulent exchange. During the measurement period, several different instruments were used for methane concentration measurements: Campbell TGA-100 (2005–2007 and April–August 2010), Los Gatos RMT-200 (January 2008–February 2014), Picarro G1301-f (April 2010–October 2011), and Los Gatos FGGGA (2014). Carbon dioxide concentrations were measured throughout the period with a LI-7000 manufactured by LI-COR Inc. The wind velocity vector was analyzed by a USA-1 acoustic anemometer by METEK (Rinne et al., 2007). All the EC data were post-processed in a consistent manner using an in-house software EddyUH (Mammarella et al., 2016). Flux data were screened for instrumental problems and for insufficient turbulent mixing. Due to instrument problems, data from 2009 were not available.

For this study, daily means of CH₄ fluxes were calculated from the screened data that contained gaps. This is a viable approach, since CH₄ fluxes do not show a diel pattern at this site (Rinne et al., 2007). However, before calculating the daily values of net ecosystem exchange of CO₂, standard gap-filling methods for peatland CO₂ fluxes were applied (Aurela et al., 2001, 2007). In short, the gap-filling algorithm estimated the CO₂ flux dependency on photosynthetic photon flux density, air temperature, and water table position, and the algorithm was used to fill periods when CO₂

Table 1. Description of the data used.

Data	Description	Usage	Units	Source	Comments
LAI	leaf area index	input	–	modeled	Gaussian curve to approximate the seasonal cycle
WTD	water table depth	input	m	measured	gap-filled at various times
NPP	net primary prod.	input	$\text{mol m}^{-2} \text{s}^{-1}$	modeled	generated by a separate NPP model
T_{soil}	soil temperature	input	$^{\circ}\text{C}$	measured	interpolated from fewer observation depths
CH_4	CH_4 flux	objective function	$\text{mol m}^{-2} \text{s}^{-1}$	measured	used in the objective function formulation
CO_2	CO_2 flux	objective function	$\text{mol m}^{-2} \text{s}^{-1}$	measured	used in the objective function formulation

fluxes were missing; see more details in Aurela et al. (2001) and Aurela et al. (2007) about the gap-filling procedure. After gap-filling the daily means of CO_2 fluxes were calculated and used in this study.

For using these carbon dioxide data with the cost function, the CO_2 flux produced by sqHIMMELI was matched with the sum of net ecosystem exchange and the net primary production of all plants. We assumed that the share of aerenchymatous plants is 70 % of the total net primary production (NPP). The fact that the net primary production is not a measured but modeled quantity (see below) introduces some uncertainty into the CO_2 flux against which the model is calibrated.

The required inputs for sqHIMMELI are daily soil temperatures, water table depths (WTDs), NPP, and leaf area indices (LAIs). The soil temperature profile for the grid used was generated by interpolating from measurement data between the measurement depths (–5, –10, –20, –35, and –50 cm) and assuming that at –3 m and below the temperature is a constant +7 $^{\circ}\text{C}$. This was the mean temperature of all the years at –50 cm depth. The WTD data used were available as measurement data, and where data were missing they were gap-filled by repeating the previous measured value. Net primary production cannot be measured in a direct way, and hence values obtained from a regression model were used. The methodology is explained in Appendix E and still further in Raivonen et al. (2017). Similarly for LAI, a simple model was used for obtaining the input. The details are, again, given in Appendix E. A summary of the data used is given in Table 1.

3 The sqHIMMELI model

The HIMMELI model (Raivonen et al., 2017) is a detailed model for estimating CH_4 emissions from wetlands. It was developed at the University of Helsinki in collaboration with the Finnish Meteorological Institute and the Max Planck Institute for Meteorology in Hamburg. The model is designed to be used as a submodel in different modeling environments, such as regional and global biosphere models. It contains processes describing the production of CH_4 and CO_2 including anaerobic production of CO_2 , the loss of CH_4 and O_2 , and transport of CH_4 , O_2 , and CO_2 between the soil and the

atmosphere. The CH_4 transport can take place by diffusion in peat (in water and in the air), by ebullition (transport by bubble formation), and by diffusion in the porous aerenchyma tissues in vascular plants. The model is driven with peat temperature, WTD, and LAI of the aerenchymatous plants. The process descriptions are mainly adopted from previous wetland CH_4 models such as Arah and Stephen (1998), Wania et al. (2010), and Tang et al. (2010). The version of the model used here differs slightly from that presented in Raivonen et al. (2017) and is therefore called with the different name of sqHIMMELI to avoid confusion.

The model simulates the processes in a discretized peat column. The number and thickness of the peat layers can be varied, but in this work six 10 cm layers are used, similarly to, e.g., Kaiser et al. (2016), with one thicker bottom layer under these, so that the total modeled peat column depth is 85 % of the maximum observed 4 m depth of the wetland, i.e., 3.4 m. The water table divides the column into water-filled and air-filled parts, and CH_4 is produced only in the inundated anoxic layers. In the present configuration, the NPP-related CH_4 production is allocated into the layers according to the vertical distribution of the root mass, described in Sect. 3.2. The internal time resolution of the model is dynamically adjusted depending on the model state, and the output interval is set to 1 day.

At present, the model does not contain descriptions for processes related to snow pack or ice such as diffusion through snow, or release of accumulated gas bubbles under ice in springtime as described by, e.g., Mastepanov et al. (2013) and Sriskantharajah et al. (2012).

HIMMELI itself, as presented in Raivonen et al. (2017), does not simulate carbon uptake (photosynthesis) or peat carbon pools but instead it takes as input the rate of anoxic respiration. The differences between HIMMELI and sqHIMMELI are described below in Sect. 3.1 and 3.2 and in Sect. 3.5.3.

For each modeled process in sqHIMMELI, there are parameters regulating the process, affecting the concentrations of CH_4 , O_2 , and CO_2 in the peat column, and the wetland methane emissions. The equations describing the physics relevant to the optimized parameters are listed in Sect. 3.4. Other relevant model equations are listed in Sect. 3.5.

3.1 Root exudates and peat decomposition

Methanogens prefer recently assimilated fresh carbon as their energy source, for instance, the root exudates of vascular plants (Joabsson and Christensen, 2001). A connection between ecosystem productivity and CH₄ emission has been observed in several wetland studies (Bellisario et al., 1999; Whiting and Chanton, 1993). However, anoxic decomposition of litter and older peat also produces CH₄ (Hornibrook et al., 1997). Many models form CH₄ substrates by extracting directly a fraction of the net primary production (van Huissteden et al., 2009; Wania et al., 2010), and some rely on heterotrophic peat respiration only (Riley et al., 2011). In sqHIMMELI, both primary production and anaerobic peat decomposition were included.

The modified sqHIMMELI model contains an exudate pool description, from which it produces methane (Eqs. 3 and 15). The exudate pool itself is described by Eq. (4), detailing how the modeled NPP turns into root exudates. Effectively, a fraction of NPP determined by the parameter ζ_{exu} (–) produces root exudates, which are then distributed as anaerobic respiration according to the root distribution into the peat column at the rate determined by the model parameter τ_{exu} (s). The part ending up under the water table produces CH₄ and CO₂, depending on the oxygen content of the water, and above the water table the exudates are respired into CO₂.

The second source of anaerobic respiration, the anaerobic peat decomposition, is modeled in sqHIMMELI with a simple Q_{10} model adopted from Schuldt et al. (2013). The peat under the water table is prescribed a turnover time, based on which anaerobic respiration and CH₄ are produced according to Eqs. (5) and (16).

3.2 Root distributions

The sqHIMMELI model differs from HIMMELI in the details regarding the root distribution model. Compared to measurement data of root distributions of aerenchymatous sedges from Saarinen (1996), the original root distribution $\pi(z)$, adopted from Wania et al. (2010) and described by

$$\pi(z) \propto \exp(-z/\lambda_{\text{root}}), \tag{1}$$

does not describe the distribution of roots well. Here, z is depth, and λ_{root} is a parameter describing the steepness of the decaying exponential curve. This formula is replaced with

$$\pi(z) \propto C_0 \exp\left[-\frac{(z-z_0)^2}{\lambda_{\text{root}}^2}\right] + C_1. \tag{2}$$

With the Gaussian shape, the new root density decreases faster with depth. Without this change, the optimization process calibrates the model to have very high root masses below 50 cm underground. The other difference between the models is that in the original model there are vanishingly few roots below the depth of 1 m, but according to Saarinen (1996),

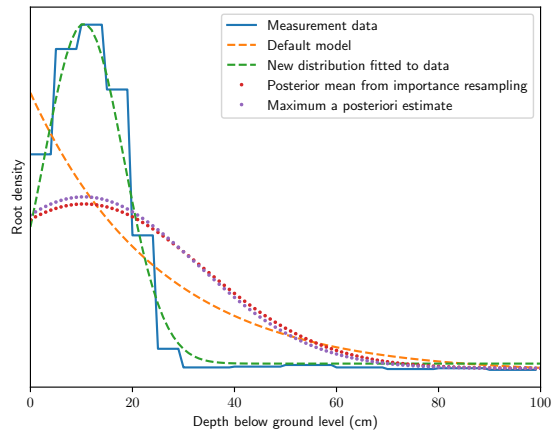


Figure 1. The different root distribution descriptions. The original description is shown as the decaying exponential, and the graph with discrete steps shows measurement data from Saarinen (1996). The new root distribution curve with optimized parameters is shown along with the curves resulting from the MCMC optimization. The original distribution gives more root mass to depths of 50–80 cm than the MCMC-optimized curves of the new root distribution. All curves are normalized to the same total root mass.

sedge roots can reach to as low as 2.3 m under the surface. The term C_1 in Eq. (2) was added to remedy this.

Before starting the optimization, the parameters C_0 , C_1 , and z_0 were fitted to data from Saarinen (1996), resulting in values of $C_0 = 215$, $C_1 = 6$, and $z_0 = 0.105$. The different root distributions are shown in Fig. 1.

3.3 Peat depth

Methane is produced from anaerobic peat decomposition at all peat depths in the sqHIMMELI model, and its transport and oxidation affect the modeled CH₄ emission. The homogeneous model description of the peat column is highly idealized, as in reality the peat column varies from place to place with respect to CH₄ production rate, production depth, and gas transport. We model the peat column to be 3.4 m deep, which is 85 % of the maximum observed depth of the Sikaneva wetland. Small uncertainty in the value of the parameter is acceptable since the parameter τ_{cato} , which regulates the rate of peat decomposition into CH₄, can partly compensate for this uncertainty.

3.4 Parameter descriptions for sqHIMMELI

The parameters for the optimization were chosen to constrain the processes most important for the CH₄ emission. Of the optimized parameters, all but ζ_{exu} (–) and Q_{10} (–) are the same for all years. However, ζ_{exu} and Q_{10} change year to year to reflect the changes in the relative CH₄ input to the

Table 2. Parameters that were not calibrated. Based on an initial sensitivity analysis, the Michaelis–Menten parameters K were not constrained by the data enough strongly and consistently to include them in the optimization. The same applies for the ebullition half-life, which is understandable given the temporal resolution of the observed data. The peat porosity was dropped from optimization in favor of the diffusivity parameters $f_{D,w}$ and $f_{D,a}$, and the specific leaf area (SLA) was not chosen for optimization since the optimized parameters τ (m m^{-1}) and ρ ($\text{m}^2 \text{kg}^{-1}$) are already part of Eq. (22) where SLA appears. The parameter $g_{\text{CH}_4}^{Q_{10}}$ was left out in favor of parameter τ_{cato} , despite their functions regarding CO_2 being different but trusting the prior value.

Parameter	Equation	Value	Units	Description	Source
$g_{\text{CH}_4}^{Q_{10}}$	16	0.4	–	peat decay to CH_4 fraction	Schuldts et al. (2013)
K_R	19	0.022	mol m^{-3}	Michaelis–Menten coeff.	Nedwell and Watson (1995)
K_{CH_4}	20	0.044	mol m^{-3}	Michaelis–Menten coeff.	Nedwell and Watson (1995)
K_{O_2}	20	0.033	mol m^{-3}	Michaelis–Menten coeff.	Nedwell and Watson (1995)
SLA	22	23	$\text{m}^2 \text{kg}^{-1}$	specific leaf area	Vile et al. (2005)
k	23	$\log(2)/1800$	s^{-1}	ebullition rate constant	–
σ	23	0.5	–	peat porosity	Rezanezhad et al. (2016)

system from peat decomposition and NPP-based production. This will allow to analyze the year-to-year changes in relative importance of the production pathways. The setup is natural; for example, Bergman et al. (2000) report the Q_{10} values changing from measurement date to another, even within a single year. As the values reported for minerotrophic lawn in Bergman et al. (2000) indicate that they may vary quite irregularly within a growing season, the modeling performed here does not take intra-annual variations into account and concentrates on the year-to-year variation. Possible mechanisms for the parameter variations include variations in substrate supply and desiccation stress, and are discussed in, e.g., Davidson et al. (2006). Table 2 shows the parameters that are used in the equations below but not optimized in this work, along with their values and explanations of why they were left out. The list of calibrated parameters along with their physical meanings is presented below.

CH₄ production-related parameters

1. τ_{exu} (s) controls the decay rate of exudates, ν , from the root exudate pool P_{exu} ,

$$\nu = \frac{P_{\text{exu}}}{\tau_{\text{exu}}}, \quad (3)$$

2. ζ_{exu} (–) is the fraction of NPP carbon that goes to the root exudate pool.

$$\frac{dP_{\text{exu}}}{dt} = -\nu + \psi_t \zeta_{\text{exu}}, \quad (4)$$

where ψ_t is the rate of NPP at time t , P_{exu} is size of the root exudate pool, and ν was given by Eq. (3).

3. τ_{cato} (y) controls the base rate of peat decomposition into CH_4 in Eq. (5).

4. Q_{10} (–) controls the temperature dependence of the rate of peat decomposition into CH_4 in anaerobic conditions via factor k_{cato} , given by

$$k_{\text{cato}} = Q_{10}^{\frac{(T-273.15)}{10}} / \tau_{\text{cato}}. \quad (5)$$

5. $f_{\text{CH}_4}^{\text{exu}}$ (–) is the fraction controlling the methane production from anaerobic respiration of root exudates in Eq. (15).

$$R_{\text{CH}_4}^{\text{exu}}(z) = \frac{f_{\text{CH}_4}^{\text{exu}}}{dz} \nu \frac{\pi(z)}{1 + \eta C_{\text{O}_2}(z)}$$

Here, $\pi(z)$ is the root distribution from Eq. (2), and ν is described in Eq. (3). The equation is discussed in Sect. 3.5.2.

Oxidation and respiration parameters

6. V_{R0} ($\text{mol m}^{-3} \text{s}^{-1}$) is the respiration parameter controlling the rate of heterotrophic respiration, which consumes O_2 and produces CO_2 . This affects the rate of temperature dependent heterotrophic respiration, $V_R(z)$, given by

$$V_R(z) = V_{R0} \exp\left(\frac{\Delta E_R}{R} \left(\frac{1}{283} - \frac{1}{T(z)}\right)\right). \quad (6)$$

Here, ΔE_R (J mol^{-1}) is a parameter affecting the temperature dependence of the heterotrophic respiration, R is the universal gas constant, and $T(z)$ is temperature at depth z .

7. ΔE_R (J mol^{-1}) is described above in the context of Eq. (6).

8. V_{O0} ($\text{mol m}^{-3} \text{s}^{-1}$) is the CH_4 oxidation parameter controlling the potential rate of CH_4 oxidation V_O :

$$V_O(z) = V_{O0} \exp\left(\frac{\Delta E_{\text{oxid}}}{R} \left(\frac{1}{283} - \frac{1}{T(z)}\right)\right). \quad (7)$$

9. ΔE_{oxid} is described in Eq. (7), affecting temperature response of CH_4 oxidation.

Gas transport-related parameters

10. λ_{root} (m) controls how the root mass is distributed; see Eq. (2).
11. ρ ($\text{m}^2 \text{kg}^{-1}$) is the root-ending area per root biomass, affecting root conductance; see Eq. (8).
12. τ (m m^{-1}) is the root tortuosity parameter affecting the root conductance K_R . A tortuosity of 1 means that the roots are not decreasing the conductance via their curvedness. The equation for the conductance is

$$K_R(z) = \frac{D_{\text{air}} m \rho \pi(z)}{\tau z}, \quad (8)$$

where $\pi(z)$ is the root mass density as a function of depth, over which the sum of the density is 1, and m is the total root mass per square meter, set to be proportional to LAI.

13. $f_{D,a}$ (–) is the fraction of the diffusion rate in air-filled peat divided by the diffusion rate in free air. The parameter affects the diffusion and the plant transport fluxes in the model: the higher this parameter is, the more diffusion there is, as it takes a shorter time for the CH_4 to exit the peat, reducing the possibility of oxidation and increasing the concentration gradient driving diffusion. The equation is

$$D_{\text{air}} = f_{D,a} D_{\text{air}}^{273} \left(\frac{T}{298}\right)^{1.82}, \quad (9)$$

where D_{air} is the diffusion rate in air-filled peat, D_{air}^{273} is the diffusion base rate at 273 K, and T is the temperature. The effect on plant transport comes via Eq. (8).

14. $f_{D,w}$ (–) is the same as above but in water. The equation describing the peat–water diffusion rate is

$$D_{\text{water}} = f_{D,w} D_{\text{water}}^{298} \frac{T}{298}, \quad (10)$$

where the terms are analogous to the ones in Eq. (9).

3.5 The sqHIMMELI model equations

The version of HIMMELI presented here describes processes for CH_4 production and transport. It differs from the version presented in Raivonen et al. (2017) in that the model presented there does not contain the processes for anaerobic respiration but rather takes them as input, the idea being that such input would be available when using HIMMELI as a part of a larger model. Hence, the equations presented in Sect. 3.5.2 are specific to the version used in this study. The other difference between the models is the difference between the root distributions described in Sect. 3.2.

3.5.1 Governing equations

The gas concentrations of CH_4 , carbon dioxide, and oxygen in the peat column are governed by the equations

$$T_X(t, z) = Q_X^{\text{diff}} + Q_X^{\text{plant}} + Q_X^{\text{ebu}} \quad (11)$$

$$\frac{\partial[\text{CH}_4]}{\partial t}(t, z) = -T_{\text{CH}_4} + R_{\text{CH}_4}^{\text{exu}} + R_{\text{CH}_4}^{\text{peat}} - R_{\text{CH}_4}^{\text{oxid}} \quad (12)$$

$$\frac{\partial[\text{O}_2]}{\partial t}(t, z) = -T_{\text{O}_2} - R_{\text{aerob}}^{\text{peat}} - R_{\text{CO}_2}^{\text{exu}} - 2R_{\text{CH}_4}^{\text{oxid}} \quad (13)$$

$$\begin{aligned} \frac{\partial[\text{CO}_2]}{\partial t}(t, z) = \\ -T_{\text{CO}_2} + R_{\text{CO}_2}^{\text{exu}} + R_{\text{CO}_2}^{\text{peat}} + R_{\text{CH}_4}^{\text{oxid}} + R_{\text{aerob}}^{\text{peat}}, \end{aligned} \quad (14)$$

where $T_X(t, z)$ describes transport of gas X containing the diffusion, ebullition, and plant transport components, and R stands for production or consumption. The different terms in the equations are described below.

3.5.2 Anaerobic respiration producing CH_4

The equations presented in this section are specific to the version of HIMMELI used in this study. The version in Raivonen et al. (2017) takes the rate of anaerobic decomposition of carbon as input and does not treat the different sources of that carbon separately.

The carbon for methane production in this model version comes from two sources: root exudates and anaerobic peat decomposition. The methane production from anaerobic respiration of that carbon is given by the terms $R_{\text{CH}_4}^{\text{exu}}$ and $R_{\text{CH}_4}^{\text{peat}}$, described by

$$R_{\text{CH}_4}^{\text{exu}}(z) = \frac{f_{\text{CH}_4}^{\text{exu}}}{dz} \nu \frac{\pi(z)}{1 + \eta C_{\text{O}_2}(z)} \quad (15)$$

$$R_{\text{CH}_4}^{\text{peat}}(z) = k_{\text{cato}}(z) g_{\text{CH}_4}^{Q_{10}} \frac{\rho_{\text{cato}} f_{\text{Ccato}}}{M_C}, \quad (16)$$

where, in Eq. (15), ν is the decay rate of root exudates from Eq. (3), η is an oxygen inhibition parameter, $C_{\text{O}_2}(z)$ is the oxygen concentration at depth z , and $\pi(z)$ is the normalized proportion of the total anaerobic root mass, also at depth z , given in an unnormalized form in Eq. (2). The decay rate of

root exudates does not depend on the peat column thickness. The parameter $f_{\text{CH}_4}^{\text{exu}}$ (–) determines what fraction of root exudates in anaerobic conditions will turn into CH_4 . Equation (15) is only used below the water table. The anoxic peat decomposition described by Eq. (16) depends on the amount of peat and its temperature, among others. The factor $g_m^{Q_{10}}$ (–) is the proportion of the anaerobic peat decomposition process producing CH_4 , ρ_{cato} is the peat density in the catotelm, f_{Ccato} is the fraction of carbon in catotelm peat, and M_{C} is the molar mass of carbon. The parameter $k_{\text{cato}} = Q_{10}^{\frac{(T-273.15)}{10}} / \tau_{\text{cato}}$ is described in Eq. (5), and is zero above water table. The equations for CO_2 are similar:

$$R_{\text{CO}_2}^{\text{exu}}(z) = \nu\pi(z) - R_{\text{CH}_4}^{\text{exu}}(z) \quad (17)$$

$$R_{\text{CO}_2}^{\text{peat}}(z) = (1 - g_{\text{CH}_4}^{Q_{10}})k_{\text{cato}}(z) \frac{\rho_{\text{cato}} f_{\text{Ccato}}}{M_{\text{C}}}, \quad (18)$$

and the meanings of the symbols are analogous to the ones in equations for CH_4 .

3.5.3 Peat respiration and methane oxidation

Peat respiration (aerobic respiration) is described with an equation of the Michaelis–Menten form:

$$R_{\text{aerob}}^{\text{peat}}(z) = V_R(z) \frac{\alpha C_{\text{O}_2}^w(z)}{K_R + C_{\text{O}_2}^w(z)}, \quad (19)$$

where $C_{\text{O}_2}^w$ is the oxygen concentration in water. Above the water table, we assume a water phase that is in equilibrium with the gas phase, i.e., $C_{\text{O}_2}^w = \alpha C_{\text{O}_2}^a$. The parameter α is a dimensionless Henry solubility constant for oxygen. Parameter K_R is the Michaelis–Menten constant of the process, and $V_R(z)$ is given by Eq. (6). Methane oxidation is controlled by dual-substrate Michaelis–Menten kinetics,

$$R_{\text{CH}_4}^{\text{oxid}}(z) = V_O(z) \frac{C_{\text{O}_2}^w(z)}{K_{\text{O}_2} + C_{\text{O}_2}^w(z)} \frac{C_{\text{CH}_4}^w(z)}{K_{\text{CH}_4} + C_{\text{CH}_4}^w(z)}, \quad (20)$$

and here the terms are analogous to those in Eq. (19), except for that the term $V_O(z)$ is described by Eq. (7).

3.5.4 CH_4 transport

The transport term $T_X(t, z)$ in Eq. (11) consist of the following terms:

$$Q_X^{\text{diff}} = D_{\text{medium}}^X \frac{\partial}{\partial z} C_X^{\text{medium}} \quad (21)$$

$$Q_X^{\text{plant}}(z) = \frac{\rho\pi(z)D_{\text{air}}^X}{\tau^2} \frac{\text{LAI}}{\text{SLA}} \frac{C_X(t, z) - C_X^{\text{atm}}}{z} \quad (22)$$

$$Q_X^{\text{ebu}}(z) = -k\sigma \frac{\text{PP}_{i,X} \sum_i \text{PP}_i(z) - (P_{\text{atm}} + P_{\text{hyd}}(z))}{RT \sum_i \text{PP}_i(z)}. \quad (23)$$

The first of these is the diffusion, where the diffusion coefficients D are given by Eqs. (9) and (10), and “medium” refers

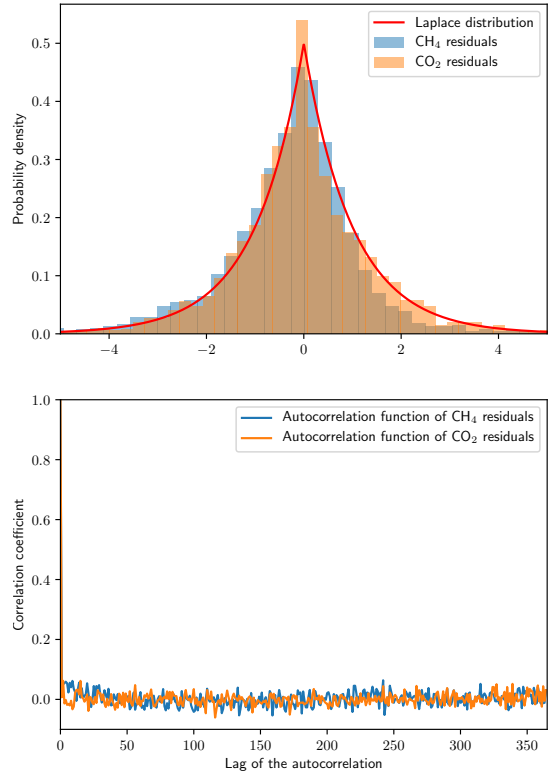


Figure 2. Residual histograms and autocorrelation functions of the error terms ϵ_t in the objective function, Eq. (24), show that neither the CO_2 nor the CH_4 residuals are autocorrelated and that they closely follow the Laplace distribution. The results shown are for the residuals from the posterior mean estimate.

to either air or water. The second equation is for plant transport, with ρ ($\text{m}^2 \text{kg}^{-1}$) and τ (m m^{-1}) described in context of Eq. (8), $\pi(z)$ the normalized root distribution mentioned above, and C_X^{atm} referring to the atmospheric partial pressure of gas X . LAI stands for the leaf area index, given as input, and SLA is the specific leaf area. The third equation is the ebullition component of the gas transport, where pp_i refers to the partial pressure of different gases indexed with i , R is the universal gas constant, k is an ebullition rate constant, and σ is the peat porosity. The parameters P_{atm} and $P_{\text{hyd}}(z)$ refer to the atmospheric pressure and hydrostatic pressure at depth z , respectively.

4 Model calibration

The model calibration consists of several steps but can be summarized as first estimating the posterior with MCMC and then based on those results, recalibrating the objective func-

Table 3. Parameter limits and prior distribution parameters. The priors are truncated Gaussian, with mean values μ and standard deviations σ , truncated at the values in the columns “low” and “high”.

	Low	High	Units	Prior μ	Prior σ	Source
$f_{D,a}$	0.01	1.0	–	0.8	0.2	Raivonen et al. (2017)
$f_{D,w}$	0.01	1.0	–	0.8	0.2	Raivonen et al. (2017)
V_{RO}	2×10^{-6}	1×10^{-4}	$\text{mol m}^{-3} \text{ s}^{-1}$	1×10^{-5}	2×10^{-5}	Nedwell and Watson (1995); Watson et al. (1997)
V_{O0}	2×10^{-6}	3×10^{-4}	$\text{mol m}^{-3} \text{ s}^{-1}$	1×10^{-5}	2×10^{-5}	Same as Raivonen et al. (2017); also Segers (1998)
λ_{root}	0.01	0.4	m	0.125	0.05	Fitted to data in Saarinen (1996)
τ	1.0	5.0	m m^{-1}	1.5	0.2	Stephen et al. (1998)
ρ	0.05	0.4	$\text{m}^2 \text{ kg}^{-1}$	0.085	0.0425	Stephen et al. (1998)
τ_{exu}	3	30	days	14	2.5	Wania (2007)
τ_{cato}	1000	30 000	years	–	–	Flat prior
ΔE_R	5000	200 000	J mol^{-1}	50 000	5000	Nedwell and Watson (1995)
ΔE_{oxid}	5000	200 000	J mol^{-1}	50 000	5000	Nedwell and Watson (1995)
$f_{\text{CH}_4}^{\text{exu}}$	0.5	0.77	–	0.635	0.06	Nilsson and öquist (2013)
Q_{10}	1.7	16.0	–	5.9	0.5*	Juottonen (2008); Gedney et al. (2004); Bergman et al. (2000)
ζ_{exu}	0.01	0.99	–	0.5	0.2*	Walker et al. (2003)

* For importance resampling, the hierarchical modeled parameters’ (Q_{10} (–) and ζ_{exu} (–)) priors were relaxed by a factor of 3 to allow for a more data-constrained resampling and to accommodate the low values of Q_{10} reported by Szafraniek-Nakonieczna and Stepińska (2014). Note that the values of the prior for these two parameters were sampled at each iteration with Gibbs sampling.

tion and using this new formulation for importance resampling. Importance resampling is typically used for obtaining posterior distributions from minor changes to the objective function descriptions (Gelman et al., 2013). This is also its purpose here.

In more detail, first, a posterior estimate was drawn running 500 000 iterations of sqHIMMELI simulations with the adaptive Metropolis Markov chain Monte Carlo algorithm with a Laplace-distributed error description and a first-order autoregressive model, AR(1), for the residuals. Second, for defining the more refined cost function for importance resampling, the optimal order for an autoregressive moving average (ARMA) time series model for the model residuals was identified from the maximum a posteriori estimate by minimizing the Akaike and Bayesian information criteria with respect to the model order. The third step was drawing a random sample of size 50 from the posterior estimate obtained with MCMC, with which the error model parameters α and γ , described in conjunction to the details of the error model in Eq. (A3), were calibrated by minimizing the Kullback–Leibler divergence (Kullback and Leibler, 1951) with respect to the standard Laplace distribution for the methane and carbon dioxide separately. The median of the obtained parameters was chosen for the second cost function used in the importance resampling. Fourth, a random sample of size 10 000 was drawn from the MCMC posterior and importance resampling was performed by drawing a subsample of size 1500 utilizing weights calculated with the new cost function values obtained from the abovementioned error model calibration as described by, e.g., Gelman et al. (2013).

The need for the importance resampling arises from the fact that the error-model-transformed methane and carbon dioxide residuals emerging from the maximum a posteriori

and posterior mean estimates from the calibration with the AR(1) model are not fully independent and identically distributed. The recalibration of the error model, and resampling from the simulated posterior using importance resampling, remedies this problem, as can be seen in the residual histogram and autocorrelation functions in Fig. 2.

4.1 Hierarchical description of parameters

In order to be able to assess the annual parameter and CH_4 transport pathway changes, a hierarchical description for two of the parameters was used. These parameters were Q_{10} (–) controlling the temperature dependence of the peat decomposition rate, and ζ_{exu} (–) regulating the production of root exudates from NPP.

The “hyperparameters” are the means and variances defining the Gaussian priors of the hierarchical parameters Q_{10} (–) and ζ_{exu} (–). They were updated using fixed Gaussian “hyperpriors” with Gibbs sampling. The sampling distribution depends on the current values of the hyperparameters. The role of the hyperprior is to constrain the distribution from which the hyperparameters are sampled.

Technically, a “Metropolis-within-Gibbs” method (Gelman et al., 2013) for sampling the hierarchical parameters, non-hierarchical parameters, and the hyperparameters was used, presented briefly in Appendix C. The model parameters (i.e., everything except the hyperparameters) were sampled with the adaptive Metropolis (AM) MCMC algorithm (Haario et al., 2001), which uses a Gaussian proposal distribution, whose covariance matrix is adapted as the chain evolves, and over time the acceptance rate gets closer to an optimal value, which is 0.23 for Gaussian targets in large dimensions (Roberts et al., 1997). If the algorithm proposes

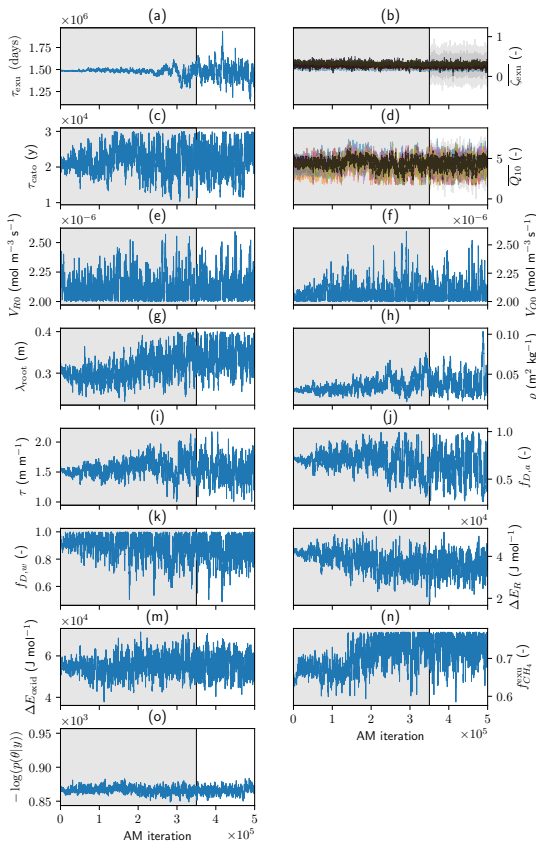


Figure 3. MCMC chains showing a thinned sample of the half million values in the chain. The first 70 % was discarded for the analyses as a warm up and is grayed out in the figures. The hierarchical parameters in panels (b) and (d) show the mean value in the middle as a black mass, and the colorful surroundings are the values of the parameters for the individual years. Panel (o) shows the value of the objective function.

values outside the hard parameter limits listed in Table 3, the model will not be evaluated and the value is rejected.

Our empirical data for the hierarchical model were the 9 years from 2006 to 2014, meaning that for each of these years there were corresponding ζ_{exu} (–) and Q_{10} (–) parameters in the optimization. The model was spun up for each annual flux estimation in order to have a realistic column of gas concentrations available. For this reason, the previous year was always also simulated, and for the likelihood only the residuals from the latter year were included in the calculations. Therefore, the year 2005 did not contribute directly to the values of the objective function. The different years were run in parallel to save execution time.

4.2 Objective functions for MCMC and importance resampling

As in many practical uncertainty quantification applications, a major part of the parameter estimation problem is the proper definition of the objective function. For MCMC, it is defined here based on a priori information about the measurement uncertainties, based on information from the model residuals, and based on additional prior information. For the importance resampling, we modify the error model for the CO_2 and CH_4 residual components of the objective function based on an analysis of the MCMC results.

4.2.1 Model residuals and error model

The form of the objective function is the same for both MCMC and importance resampling. The first two components of the objective function contain the contributions from the modeled differences to the daily CH_4 and CO_2 flux measurements. In the MCMC objective function, it is assumed that the daily flux estimate uncertainties are dependent on approximately a fraction α of the flux measurement (Richardson et al., 2006) and some constant error, γ (e.g., measurement device precision). The model error is expected to follow a similar form, and hence α and γ contain the contributions from both the model and measurement errors. For importance resampling, the description is the same except for that a 14-day running mean of the interannual variability is used for α . These parameters are set independently for both CH_4 and CO_2 .

When determining the parameters γ and α , the resulting residuals end up being autocorrelated. Therefore, they are treated as such with the AR(1) model for MCMC and with the ARMA(2,1) model for the importance resampling, described, e.g., in Chatfield (1989).

Since the primary interest is in the methane fluxes, the carbon dioxide residuals are scaled down to a fifth in the importance resampling cost function, which is enough to guide the parameter values since several years of CO_2 flux data are used. Furthermore, as the model does not contain descriptions for the effects of snow and ice on the fluxes, the fit cannot be expected to be very good in the winter months. Therefore, we further only consider 20 % of the contribution of the residuals in the winter season from December to February. The obtained residuals, denoted by the ϵ terms in the objective function, Eq. (24), are treated as Laplace distributed. The flux observation errors are reported to follow a distribution of this type, rather than a Gaussian distribution (Richardson et al., 2006). The error model is explained in more detail in Appendix A.

4.2.2 Prior information

The parameters affecting the CH_4 production of the wetland model are not known well, but despite this, not setting any

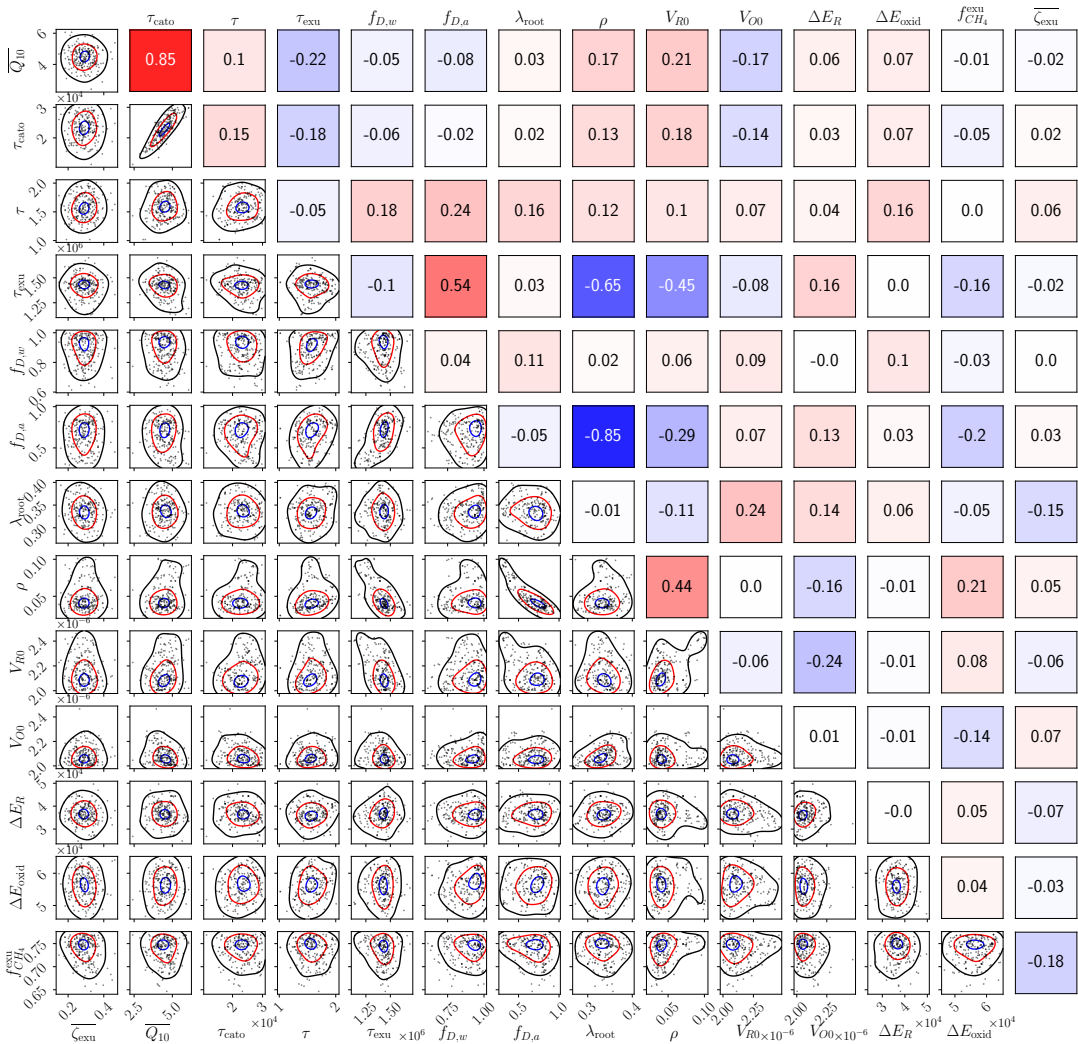


Figure 4. Posterior distributions of the parameters from the importance sampling. The two-dimensional marginal distributions of the posterior distribution are shown in the triangle on the lower left (labels on the left and at the bottom), and the correlations between parameters are shown in the upper triangle on the right (labels on the left and at the top). The images in the lower left triangle show the 90 % (black), 50 % (red), and 10 % (blue) contours, and points from a random sample of the posterior (black dots). On the upper right, each plot shows correlation coefficients between parameters, color coded to show negative correlations in blue and positive in red. The units are listed in Table 4.

prior distributions on parameters can lead to nonphysical parameter values in the posterior distribution.

The parameter priors are set to zero outside prescribed bounds. Within these bounds, the parameters are assigned Gaussian priors, with the exception of one parameter whose prior is set to be flat. The prior values are based on both lit-

erature and expert knowledge, and the information regarding the parameter values is summarized in Table 3.

4.2.3 The objective function

The “objective function” for the parameter optimization, $J(\theta)$, is the negative logarithm of the value of the unnormalized posterior probability density function at θ . It combines

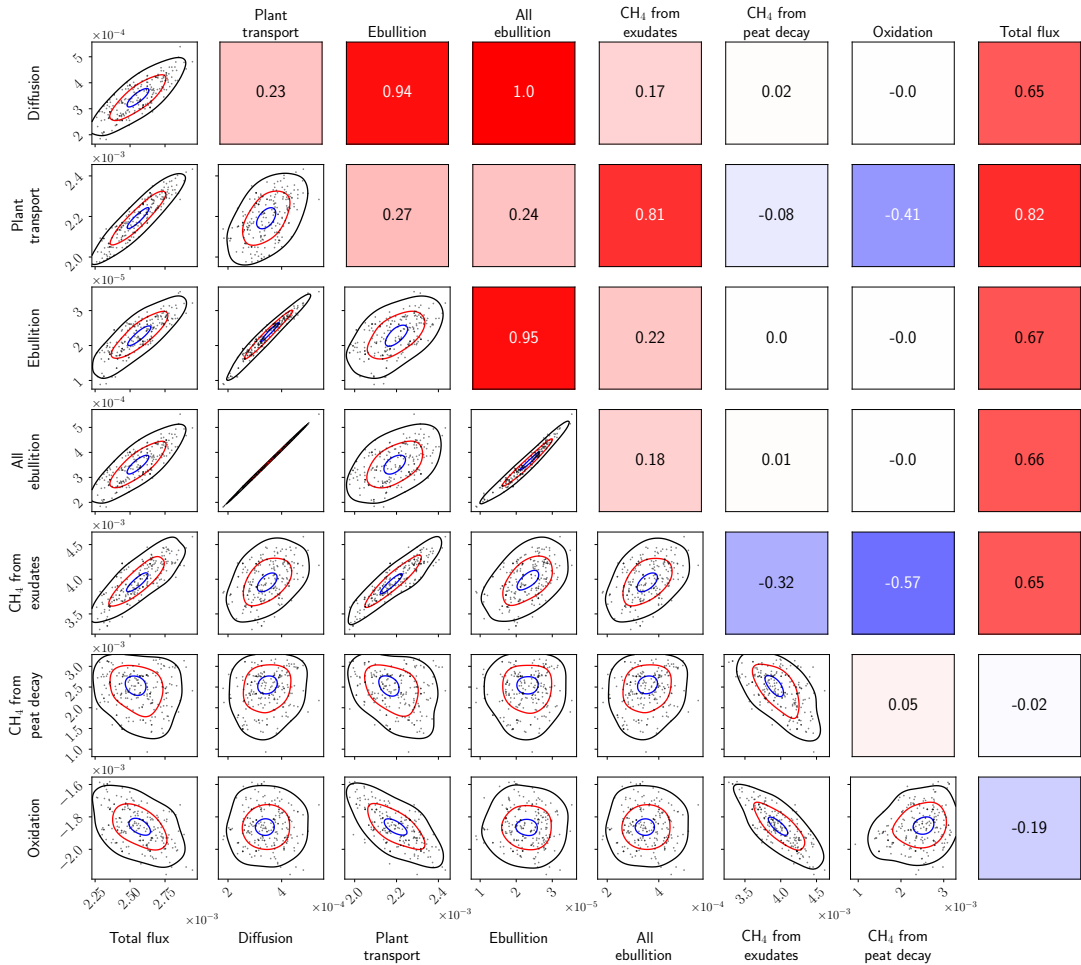


Figure 5. Posterior distributions and correlations of the annual means of the output from the modeled processes for the year 2012. The dynamics for the other years are mostly similar, but the strengths of the correlations vary somewhat. The results shown are based on 1000 random samples from the parameter posterior distribution. The two-dimensional marginal distributions in the triangle on the lower left have their labels on the left and at the bottom, and the correlations between the processes in the upper triangle on the right have their labels on the left and at the top. The images in the lower left triangle show the 90 % (black), 50 % (red), and 10 % (blue) contours. The all-ebullition and diffusion fluxes correlate almost fully, showing that the “diffusion” flux has a strong contribution from underground ebullition.

our statistical knowledge of the flux observations and parameter priors presented in Sect. 4.2.1–4.2.2 and is given by

$$\begin{aligned}
 J(\theta) &= -\log(p(\theta|\mathbf{y})) \\
 &= \sum_{i=1}^{N_{\text{obs}}^{\text{CH}_4}} |\epsilon_i^{\text{CH}_4}| + \sum_{j=1}^{N_{\text{obs}}^{\text{CO}_2}} |\epsilon_j^{\text{CO}_2}| + \frac{1}{2} \sum_{k=1}^{N_{\text{par}}} \frac{(\theta_k - \mu_k)^2}{\sigma_k^2}.
 \end{aligned}
 \tag{24}$$

Here, ϵ_i are the AR(1)- or ARMA(2,1)-transformed, Laplace-distributed residuals, and the last term is the prior contribution, where θ_k is the proposed parameter value, μ_k is the prior mean, and σ_k^2 is its variance. For further technical details, see Appendix A.

5 Results and discussion

The Markov chain Monte Carlo simulations yielded a chain of 500 000 samples. From these, 70 % from the start of the

Table 4. Parameter values obtained in the optimization of the sqHIMMELI model with importance resampling. The maximum a posteriori, posterior mean, non-hierarchical mean (mean values used for hierarchically varying parameters), and values from Raivonen et al. (2017) are shown. The horizontal line in the middle separates the hierarchically optimized parameters (including their priors) from the others.

Parameter	MAP	Posterior mean	Non-hier. mean	Default
$\tau_{\text{cato}} (\times 10^4 \text{ y})$	2.872	2.269	2.269	3.0
$\tau (\text{m m}^{-1})$	1.462	1.581	1.581	1.5
$\tau_{\text{exu}} (\times 10^6 \text{ s})$	1.187	1.411	1.411	1.21
$f_{D,w} (-)$	0.866	0.887	0.887	0.8
$f_{D,a} (-)$	0.427	0.65	0.65	0.8
$\lambda_{\text{root}} (\text{m})$	0.314	0.333	0.333	0.252
$\rho (\text{m}^2 \text{ kg}^{-1})$	0.081	0.049	0.049	0.085
V_{R0} ($\times 10^{-6} \text{ mol m}^{-3} \text{ s}^{-1}$)	2.366	2.153	2.153	10.0
V_{O0} ($\times 10^{-6} \text{ mol m}^{-3} \text{ s}^{-1}$)	2.013	2.09	2.09	10.0
ΔE_R ($\times 10^4 \text{ J mol}^{-1}$)	3.478	3.647	3.647	5.0
ΔE_{oxid} ($\times 10^4 \text{ J mol}^{-1}$)	5.358	5.575	5.575	5.0
$f_{\text{CH}_4}^{\text{exu}} (-)$	0.729	0.736	0.736	0.5
$\zeta_{\text{exu}} (-)$	0.343	0.292	–	–
$\zeta_{\text{exu}}^{\text{std}} (-)$	0.128	0.157	–	–
$\overline{Q}_{10} (-)$	5.721	4.425	–	–
$Q_{10}^{\text{std}} (-)$	0.587	0.616	–	–
$\zeta_{\text{exu}}^{2006} (-)$	0.212	0.182	0.292	0.4
$\zeta_{\text{exu}}^{2007} (-)$	0.251	0.244	0.292	0.4
$\zeta_{\text{exu}}^{2008} (-)$	0.28	0.276	0.292	0.4
$\zeta_{\text{exu}}^{2009} (-)$	0.202	0.243	0.292	0.4
$\zeta_{\text{exu}}^{2010} (-)$	0.34	0.314	0.292	0.4
$\zeta_{\text{exu}}^{2011} (-)$	0.251	0.258	0.292	0.4
$\zeta_{\text{exu}}^{2012} (-)$	0.327	0.324	0.292	0.4
$\zeta_{\text{exu}}^{2013} (-)$	0.368	0.313	0.292	0.4
$\zeta_{\text{exu}}^{2014} (-)$	0.334	0.323	0.292	0.4
$Q_{10}^{2006} (-)$	5.946	4.488	4.425	3.5
$Q_{10}^{2007} (-)$	4.882	3.857	4.425	3.5
$Q_{10}^{2008} (-)$	4.017	3.684	4.425	3.5
$Q_{10}^{2009} (-)$	5.469	4.14	4.425	3.5
$Q_{10}^{2010} (-)$	5.337	4.284	4.425	3.5
$Q_{10}^{2011} (-)$	6.306	4.305	4.425	3.5
$Q_{10}^{2012} (-)$	5.377	4.193	4.425	3.5
$Q_{10}^{2013} (-)$	5.219	4.211	4.425	3.5
$Q_{10}^{2014} (-)$	6.438	4.332	4.425	3.5
Cost function value	1205.22	1227.01	–	–

chain were discarded as a warm up (Fig. 3). A revised posterior distribution, obtained by first sampling 10 000 entries randomly from the chain, and after that obtaining 1500 entries from those with importance resampling, is shown in Fig. 4, and the correlation features are shown in the upper triangle of that figure. For the different processes, Fig. 5 shows an example of the posteriors and the process correlations.

Three different parameter estimates obtained from the posterior distribution were used to look at its features and fluxes: the maximum a posteriori (MAP) estimate, posterior mean estimate, and a non-hierarchical posterior mean estimate, where the mean values of the parameters $\zeta_{\text{exu}} (-)$ and $Q_{10} (-)$ over the different years were used. The “default” parameters in the text and figures refer to values adapted from Raivonen et al. (2017). If not stated otherwise, the maximum a posteriori and posterior mean estimates refer to the values obtained from the importance resampling, not from the MCMC.

5.1 Parameter values

The parameter values used in the analyses are shown in Table 4. The MAP and posterior mean estimates agree on the value of the water diffusion rate coefficient $f_{D,w} (-)$, and the posteriors shown in Fig. 6k show that the estimates are close to the middle of the marginal distribution and slightly above the prior value. In tests with a shallower peat column, smaller values of this variable were obtained (not shown).

Contrary to this, the air diffusion rate coefficient, $f_{D,a} (-)$, finds its best values lower, and the variability of the parameter is larger than that for the diffusion rate coefficient in water-filled peat.

The root distribution parameter, λ_{root} , is optimized larger than expected, and again the MAP estimate is close to the posterior mean. This implies that the model optimizes best when the CH_4 produced from the photosynthesis-induced exudate production goes relatively far below the surface: with a value of 0.3, 49% of the roots are deeper than 25 cm, 15 % of the roots are deeper than 50 cm, and just 2.5 % are deeper than 75 cm; see Fig. 1. In relation to these numbers, the water table depth is most of the time above the depth of -20 cm. Additionally, a larger λ_{root} will facilitate the emission of the CH_4 produced by peat decomposition in the catotelm.

The values of the exudate pool turnover time τ_{exu} are close to the default value of 2 weeks, with the MAP estimate at a little under 14 days and the posterior mean at 2.5 days more. The results from the importance resampling show that the spread is around 3 days around this posterior mean value. However, the value of ζ_{exu} controlling amount of exudates produced from photosynthesis is smaller than the default value at roughly 0.15–0.45, with the MAP and posterior mean estimates at 0.343 and 0.292, respectively. In contrast to this, and balancing the effect of a relatively low ζ_{exu} , the parameter $f_{\text{CH}_4}^{\text{exu}} (-)$, controlling how much methane is produced from anaerobic decomposition of exudates, has a

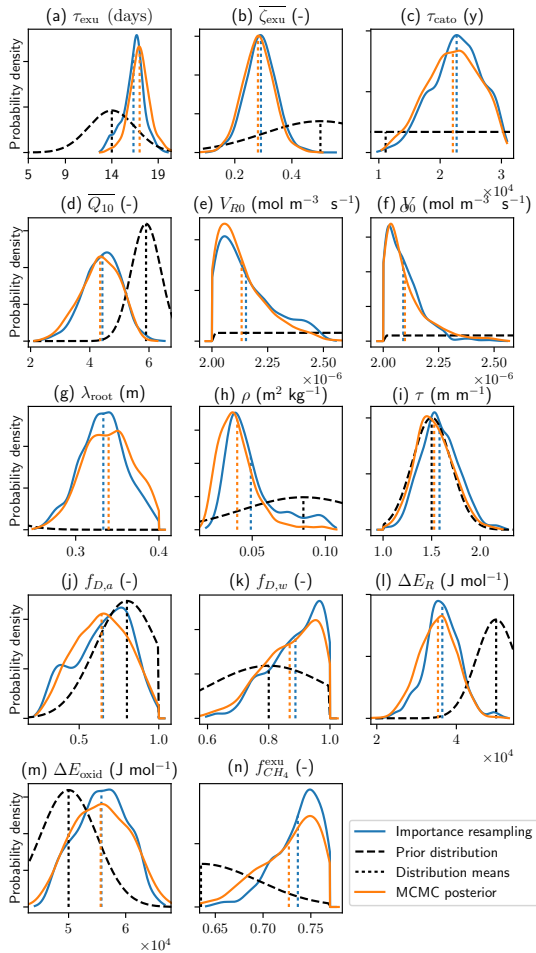


Figure 6. Posterior marginal and prior distributions from MCMC and importance resampling for all parameters: panels (a–d) and (n) are the production-related, (e–f) the respiration- and oxidation-related, and (g–k) the gas-transport-related parameters. The blue and orange curves shown are smoothed slightly using Gaussian kernel estimates for readability. To make these figures, 70 % from the start of the MCMC chain was discarded as a warm up (orange line). The dotted vertical lines show the prior mean values and the sample means from both MCMC and importance sampling. For the parameters ζ_{exu} (b) and Q_{10} (d), the prior distribution drawn is the hyperprior.

skewed posterior marginal distribution with most of the mass above the value of 0.7, as can be seen in Fig. 6.

The non-hierarchically optimized parameter, V_{R0} ($\text{mol m}^{-3} \text{s}^{-1}$), controlling the amount of CH_4 oxidation taking place is close to the minimum allowed value at one-fifth of the default value. This is also true for the parameter

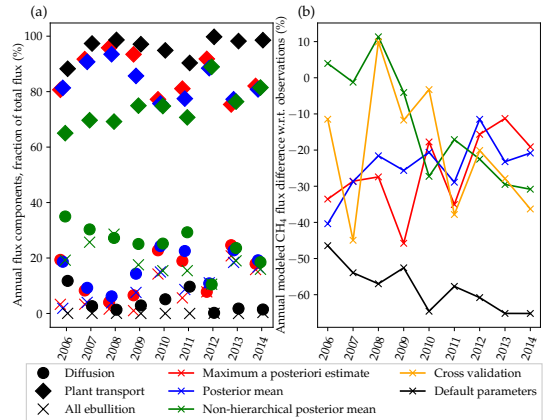


Figure 7. (a) Proportions of flux components as a function of the year. Diamonds are for plant transport, balls for the diffusion flux, and crosses describe the total ebullition taking place. The figure on the right shows the annual model–observation mismatch in percent for the methane flux, where only residuals from days with observation data available have been taken into account. The data in panel (a) have been spread slightly for readability in the x -axis direction. The orange line in panel (b) represents the results from the cross validation discussed in Sect. 5.6. Note that the optimization target was not to directly fit annual emissions.

controlling heterotrophic respiration, V_{R0} ($\text{mol m}^{-3} \text{s}^{-1}$), all of whose optimized estimates reside close to its minimum value, reducing the amount of heterotrophic respiration taking place. The posteriors are very narrow. In contrast to these narrow posteriors, the parameters ΔE_{oxid} (J mol^{-1}) and ΔE_R (J mol^{-1}), which are present in the same equations as the V_{O0} and V_{R0} parameters, have slightly wider posterior distributions, with the former slightly under and the latter slightly above the default values.

Table 4 shows that the hierarchically optimized parameter Q_{10} (–), controlling the temperature dependence of the CH_4 production from peat decomposition, has slightly different values for the MAP and posterior mean estimates, with the Gibbs-sampled mean value (mean of those values in the case of the posterior mean) at 5.72 and 4.43, respectively.

The parameter τ_{cato} (y), also controlling the peat decomposition rate in the catotelm, compensates for the differences of Q_{10} between the MAP and posterior mean estimates by having a faster turnover time for the posterior mean than the MAP estimate. That parameter has a wide posterior, ranging from around 10 000 to 30 000, which was the value used by Raivonen et al. (2017) and the upper limit of the parameters in our work. Our posterior density goes to zero towards the higher limit, and the posterior mean is found at the value of 22 690 years.

The interannual variability of Q_{10} (–) is mostly similar for both MAP and posterior mean estimates. For instance, the

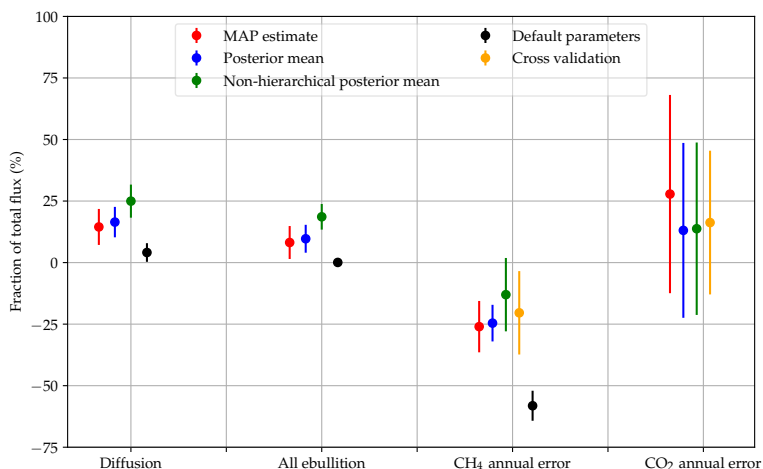


Figure 8. Fractions of the annual diffusive fluxes of the total fluxes. Means and 1σ error bars are shown. Almost all ebullition takes place when the water table is below the peat surface, and hence it is emitted to the atmosphere as part of the diffusion flux. Plant transport is not shown, as it is very close to the complement of the diffusive flux: together, these two streams add up to more than 98 % of the total flux. Plant transport variation is very close to that of diffusion. On the right side of the figure, the average annual errors are shown for the interannual variation of the fluxes. The results of the cross validation of the regression modeling of the hierarchically varying parameters, discussed in Sect. 5.6, are drawn in orange. The “default” parameters produce carbon dioxide fluxes that are above the upper limit of the chart.

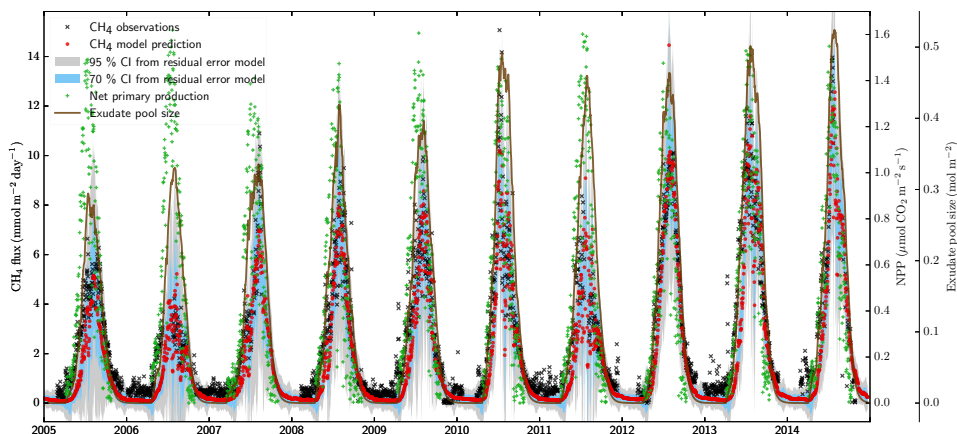


Figure 9. Output CH_4 flux (red dots) with parameters from the posterior mean. Methane observations (black crosses) and predicted fluxes with confidence intervals from ARMA(2,1) modeling of a set of 1000 residual time series are shown, as are the input net primary production (green dots) and the exudate pool sizes (brown line). Most of the observations are inside the confidence intervals, but note that the effects of the parameter variations in the posterior are not part of these confidence intervals. The constituents of the total flux are shown in Fig. 12.

years of the smallest values are 2007 and 2008 in both cases, and the values of the years 2006, 2011, and 2014 are the largest in both cases. For the other hierarchically calibrated parameter, ζ_{EXU} (–), these similarities do not exist.

5.2 Cost function values and model fit

Table 4 lists the cost function values for the MAP and posterior mean estimates, and the annual errors for the MAP, posterior mean, and non-hierarchical posterior mean estimates and default parameter values are shown for each parameter set in Fig. 7. The cost function value is unsurprisingly lower

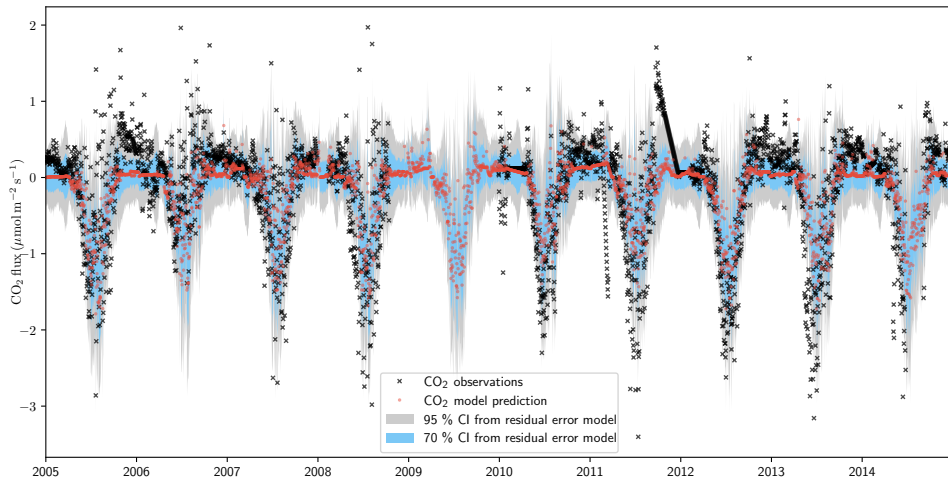


Figure 10. Output CO₂ flux (red dots) with parameters from the posterior mean. Carbon dioxide observations (black crosses) and model-predicted fluxes with confidence intervals from the ARMA(2,1) modeling of a set of 1000 residual time series are also shown. As with methane, most of the observations are inside the confidence intervals. The parameter variations in the posterior probability distribution are not reflected in these confidence intervals.

for the MAP estimate than for the posterior mean estimate, indicating a better fit in terms of the error model. In Fig. 7b, the non-hierarchical posterior estimate shows a large variance of the annual errors, with early years having a positive bias, and later years having a negative bias. Incidentally, the average discrepancy from observations over the whole period for the non-hierarchical posterior mean is small for both methane and carbon dioxide, as Fig. 8 indicates. However, the variation for methane is the largest, implying that the annual variation is not reflected well. The model estimates of the annual fluxes are good in that the variance of the errors is small for both MAP and posterior mean experiments, especially, even though the estimates show a negative bias of 25%. Compared to the default parameters, which strongly underestimate methane emissions (and even more overestimate the carbon dioxide emissions), the flux estimates are much improved. This is to be expected as the results shown are not for an independent validation dataset. Rather, the motivation with the MAP and posterior mean estimates is to see what the model fit looks like for optimized parameters and how the features differ from the unoptimized ones. It is, however, worth noting that the target objective function did not aim at minimizing annual discrepancies but daily residuals that were considered correlated.

A cross validation of the regression modeling in terms of the annual errors is shown in Figs. 7b and 8. While the annual estimates are not on average better than the ones from the simulation with the non-hierarchically obtained posterior mean, the spread of the errors are acceptable, particularly if the strong negative bias in 2007, which is mostly due to lack

of observations during the season, is disregarded. Additionally, the overall biases are surprisingly slightly better than with the optimized parameters, due to effects of the prior, different data resolution in the cost function, and the non-trivial error model used. The cross validation is described in Sect. 5.6.

The positive bias in the CO₂ may partly be due to the assumption that 70% of the NPP comes from the aerenchymatous plants, and this affected the data that the sqHIMMELI model results were matched with.

All years of hierarchically optimized experiments show at least a small negative annual bias in the methane flux when compared to the available observations. This can be due to the high day-to-day variability of the summertime fluxes, which dominate year-round total fluxes, and the fact that the model can not, without data about the fine structure and heterogeneity of the wetland, match the high variability fluxes. The proportional model–data residual error component αy_t (Appendix A) allows the model to underestimate the high peaks more than the low flux values. The error model favors the baseline of the lower values during periods when observed variance is very high, for instance, in the peak emission season of 2010. This is also true for periods of increased ebullition, and such fluxes are very difficult to fit into. These periods contribute to both the cost function values and the underestimation of the total methane flux. Any temporal shifts of peaks of seasons are penalized heavily, and the optimized parameter values rather produce less peaks than right size peaks at a slightly wrong time.

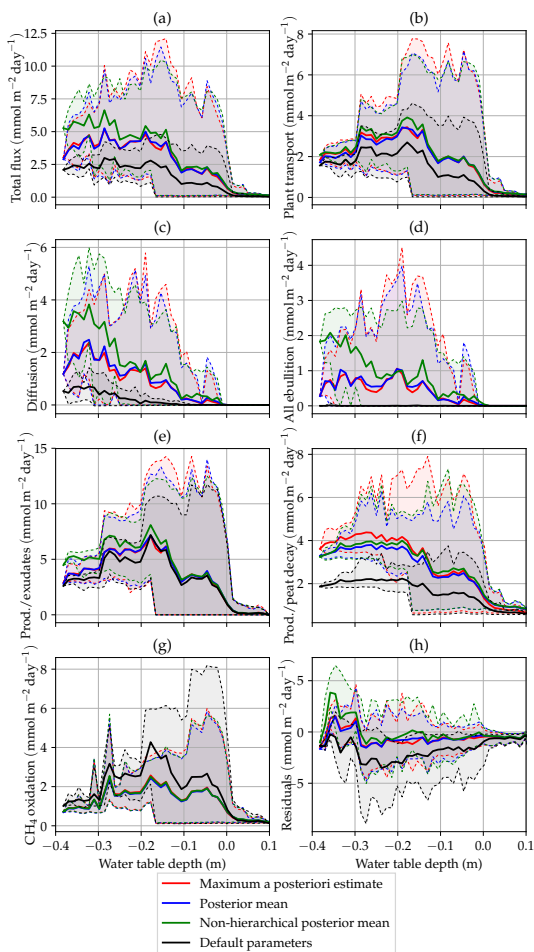


Figure 11. Means of total CH₄ emission (a), its components (b–c), total ebullition taking place (d), CH₄ production (e–f), CH₄ oxidation (g), and model residuals (h) as functions of water table depth. Shaded areas show the 5th and 95th percentiles. To look at the effect of the optimization, compare the black and the blue/red lines.

Another reason is that the carbon dioxide fluxes are overestimated by the model, leading to need to balance between the two, and as methane production in the wetland also produces carbon dioxide, the optimization algorithm will find a middle ground between the conflicting needs of minimizing carbon dioxide and maximizing methane production.

Additionally, the wintertime methane fluxes are underestimated systematically, and the emissions start slightly late in early summer, which produces a negative bias to the total flux even though visually the fit is good, as can be seen in Fig. 9. This figure also reveals that the observations for the vast majority fall within the confidence margins suggested by

the ARMA model for the residual. The variation from the full posterior is higher because the uncertainty shown in Fig. 9 does not take the parameter variations into account.

The carbon dioxide time series against flux observations are shown in Fig. 10. This figure reveals that sqHIMMELI and the error model most of the time are able to explain the carbon dioxide fluxes well, even though some of the largest deviations are not captured. Since in an observational time series outliers can come from an underlying process that is not well explained by these models, having a small number of such deviations is not surprising.

The input data have a role in affecting the model fit to the data, and since NPP is a modeled quantity, there is some additional uncertainty stemming from that modeling involved. For LAI, we note that even though in reality it is not identical every year, in the model, it follows the same pattern (see Appendix E). The parameter calibration must then favor parameters producing a good fit in terms of average model performance.

5.3 Parameter values and processes in sqHIMMELI

The sqHIMMELI model produces the CH₄ from anaerobic respiration that originates from peat decay and the decay of root exudates. These production components, along with the different output pathways, CH₄ oxidation, and model residuals, are plotted as functions of water table depth in Fig. 11 for the MAP, posterior mean, non-hierarchical posterior mean, and default parameter values. The process correlations and covariances are shown for the year 2012 in Fig. 5.

In the following, “all ebullition” refers to any ebullition in the peat column regardless of whether the bubbles reach the peat column surface. “Ebullition” refers to the part of all ebullition which reaches the surface. Most of the time, the water table is under the peat surface, and at those times ebullition is zero, although all ebullition can be substantial. In that case, the ebullition flux does not go directly into the atmosphere, but into the first air-filled peat layer above the water table level, and continues from there via other pathways. The reason for this separation comes from implementation details of HIMMELI. In all experiments, ebullition reaching the surface is a minor fraction of the total CH₄ emission.

For the posterior mean estimate, the flux components and oxidation are shown as time series in Fig. 12. Optimizing the model leads to increased production of methane from peat decay, as can be seen in Fig. 11f. A similar effect is seen also in the plant transport component in Fig. 11b.

Comparing results from simulations with optimized parameters to results using the default parameter values (shown in Table 4) shows that the optimization somewhat decreases the role of the plant transport pathway in favor of the diffusion pathway, especially for the years 2010, 2011, and 2013. Diffusion and all-ebullition fluxes are closely tied to each other, as can be seen in Fig. 7a, in that in many years (2007–2008, 2012–2014) their values are close to each other for all

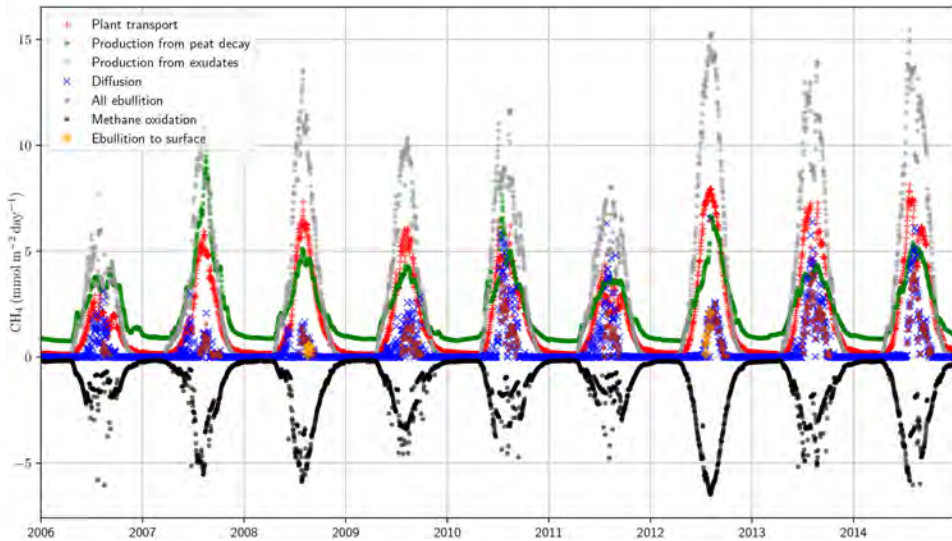


Figure 12. Diffusion, plant transport, ebullition, CH₄ production, and CH₄ oxidation time series for parameter values from the posterior mean estimate. The figure shows how only a minor part of ebullition in the end comes to the surface as ebullition. The total flux and the observations are shown in Fig. 9.

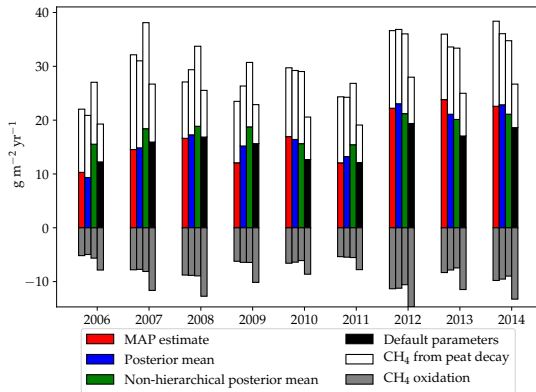


Figure 13. Annual CH₄ production in g m⁻² from root exudates (colored part) and peat decomposition (white part) for the different years. Oxidized CH₄ is shown as gray and negative.

estimates. This is also visible in the flux component time series in Fig. 12.

5.3.1 Methane production and oxidation

Figures 13 and 5 show that there is considerable interannual variation in the production of CH₄ from both of the production processes. The year 2007 has a high amount of production from peat decomposition, whereas the year 2006

shows a lot less, even though the ζ_{exu} -controlled proportion does not change equally much. Generally, though, in years of high emissions, the amount of CH₄ from both of the production sources is increased. The shape of the NPP input, shown in Fig. 9, does not change remarkably from year to year, but the emissions change considerably, as the model state and input affect the production non-linearly. For example, in times of low WTD in the peak emission season, the root exudates do not contribute to CH₄ production as much as during slightly wetter times, as much of the roots are located in the dry part of the peat column and the exudates are deposited there (Fig. 11e). Another explanation for changes in CH₄ production comes through the production-determining parameters, whose variation is in Sect. 5.6 found to be related to the springtime temperature and NPP.

The NPP-based CH₄ production controlled by the parameter ζ_{exu} (–) is not strongly constrained by its hyperprior as can be seen in Fig. 6b and the MAP and posterior mean estimates. The posterior means in Table 4 are between 0.182 and 0.323 for the different years. For the MAP values, the values are slightly higher, leading to a larger input to the root exudates pool. The effect of ζ_{exu} on the exudate pool sizes can be seen by comparing the posterior mean values to the exudate pool sizes in Fig. 9. The values obtained here are in line with values reported by Walker et al. (2003), who give a range of roughly 0.15–0.65 in terms of our $\zeta_{\text{exu}}^{\text{CH}_4}$ parameter, when also considering the mean value of the $f_{\text{exu}}^{\text{CH}_4}$. This parameter finds its maximum a posteriori value at 0.729, which is close to the prescribed upper limit of 0.77. The posterior mean is at

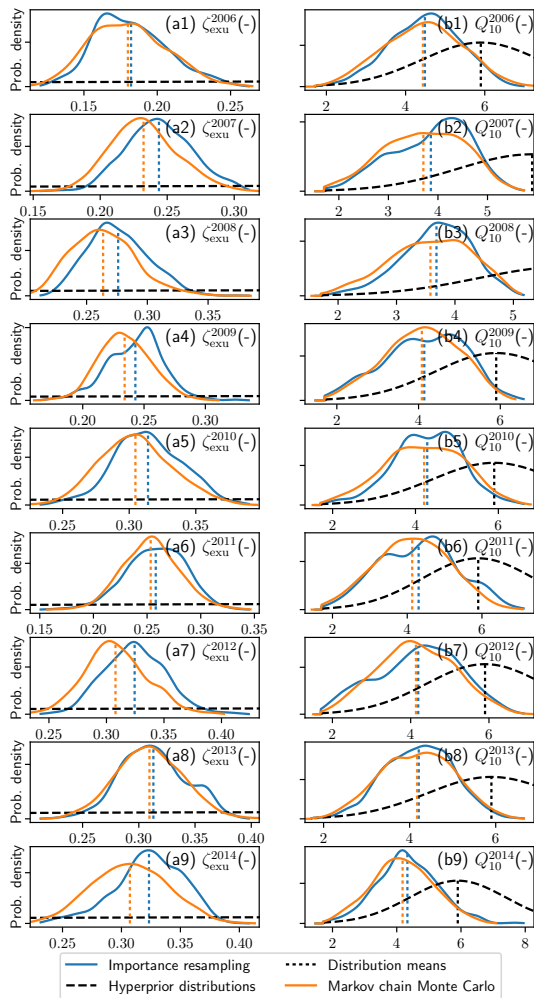


Figure 14. Posterior marginal distributions of the hierarchical parameters from both MCMC and importance sampling, along with the hyperpriors. Panels (a1)–(a9) are for the parameters ζ , and (b1)–(b9) for Q_{10} . The curves shown are smoothed slightly using Gaussian kernel estimates for readability. To make these figures, 70 % from the start of the MCMC chain was discarded as a warm up. The dotted vertical lines show the default parameter values and the mean values of the posterior distributions. Importance resampling had the tendency of moving the posteriors of the ζ parameters slightly higher, despite the weaker prior used for that step.

0.736. From these results, we can conclude that a relatively large portion of the photosynthesized sugar is respired into methane.

The parameter $f_{\text{exu}}^{\text{CH}_4}$ is only affecting the part of the anaerobic respiration generated from root exudates. The two

sources of anaerobic respiration (peat decomposition and root exudates) are in sqHIMMELI controlled by two different processes having different sets of parameters. The parameter controlling the peat decomposition, $g_{\text{CH}_4}^{Q_{10}}$, appearing in Eq. (16) and functioning analogously to $f_{\text{exu}}^{\text{CH}_4}$, is set at the value 0.4 based on prior information, and this parameter was not part of the calibration. The discrepancy between the $g_{\text{CH}_4}^{Q_{10}}$ and $f_{\text{exu}}^{\text{CH}_4}$ parameters is after the optimization rather large, and therefore, in any future calibration of the sqHIMMELI model with flux data from another site or with data from several sites, including this parameter could be also considered. If the value of 0.4 for $g_{\text{CH}_4}^{Q_{10}}$ is an underestimate, the model produces too much carbon dioxide and too little methane from the peat decomposition component. However, since the production processes are correlated in the posterior distribution, as shown in Fig. 5, increasing the value of $g_{\text{CH}_4}^{Q_{10}}$ would also be reflected in decreasing the production of methane from root exudates and increasing the production of carbon dioxide correspondingly. According to Fig. 5, methane oxidation would also be affected by changes to methane production from the root exudate component. Hence, excluding the parameter $g_{\text{CH}_4}^{Q_{10}}$ from the optimization does not effect the total CO_2 and CH_4 fluxes in a major way, but the balance of the production processes and methane oxidation can be slightly affected.

The year-to-year variation of the posterior distributions of the ζ_{exu} parameter, shown in Fig. 14, is large, and this difference has an important role in driving the annual CH_4 production. Especially for the years 2007, 2008, 2012, and 2014, the importance resampling has the effect of increasing the value of the parameter, correspondingly increasing the production of methane. This effect is not visible for the other hierarchically modeled production-related parameter, Q_{10} , whose posterior is not affected by the resampling despite the more permissive prior.

The methane produced by the action of ζ_{exu} is distributed according to the root distribution, whose form is determined by λ_{root} (m). The posterior means reveal that the contribution of the prior component of λ_{root} to the cost function is large. Its values might well be larger with a wider prior and more permissive prior, but in regard to how root distributions are in reality (Fig. 1), larger values for the parameter would make its interpretation difficult. This parameter affects both how exudates are allocated in the column and how deep the fast plant transportation reaches. Clearly, there is a need to reach further down, implying that the model performs more optimally when it transports CH_4 faster to the atmosphere.

The exudate pool size follows the net primary production in Fig. 9 with a delay, as one could expect. According to the modeling, the pool sizes are up to 0.5 mol m^{-2} , and the exudate pool is depleted from December until the start of the growing season.

The methane production from decomposition of peat in anaerobic conditions is aided by the rather strongly corre-

lated parameters Q_{10} (–) and the catotelm carbon decay half-life τ_{cato} (y) as seen in Fig. 4. The prior means of Q_{10} (–) are mostly inside the 1σ bounds of the hyperprior, and the temperature dependence of the anaerobic respiration from peat decomposition is close to what was a priori expected. The MCMC utilized a rather strict prior, which constrained the parameter exploration somewhat. Despite this, also very low values were proposed.

Methane oxidation is quite steady between the different estimates as can be seen in Fig. 13 – except for the default parameters values, with which the amount of oxidation is several tens of percent more. However, there is considerable interannual variability, which seems to be related to the varying production from exudates, as seems to be suggested in Fig. 5 and also in Fig. 13.

The stronger oxidation with the default parameter values can be for its part also linked to the larger V_{O0} ($\text{mol m}^{-3} \text{s}^{-1}$) parameter, despite the other parameter determining oxidation in Eq. (7), ΔE_{oxid} , being slightly lower (50 000 vs. 53 580 for MAP and 55 750 for posterior mean).

The process correlation figure (Fig. 5) also shows that the exudate- and peat-decomposition-based methane production terms are negatively correlated, and that the exudate-based production is roughly 50 % stronger than the peat decay source.

The hard prior bounds of V_{O0} ($\text{mol m}^{-3} \text{s}^{-1}$) were tight; for example, Segers (1998) reports that potential CH_4 oxidation can vary across 3 orders of magnitude. Hence, also lower proportions of CH_4 oxidation could have been seen with a more permissive prior. This would have then also altered the posteriors of the weakly co-varying parameters, most notably λ_{root} .

The parameter V_{R0} ($\text{mol m}^{-3} \text{s}^{-1}$) controlling heterotrophic respiration correlates positively with CH_4 production via τ_{exu} (s) (smaller value enhances methane production), but the correlations with Q_{10} and τ_{cato} seem to cancel out each other. The correlations of ζ_{exu} are weak, implying that process is well constrained by the combined CO_2 and CH_4 data. There is also a weak anticorrelation between V_{R0} and ΔE_R , which is to be expected based on Eq. (6).

5.3.2 Plant transport

The amount of plant transport in the calibrated models, shown in Fig. 7a, is between 75 and 95 %, which is just slightly higher than the range of 68–85 % reported in Wania et al. (2010) in a study simulating CH_4 emissions for seven boreal peatlands. The high optimized share of plant transport is mainly due to the high values of the root depth controlling parameter λ_{root} (m) and some of the difference between the MAP and posterior mean estimates in Fig. 7a may be explained by the higher root-ending cross-sectional area in the MAP estimate, controlled by parameter ρ ($\text{m}^2 \text{kg}^{-1}$). Wania et al. (2010) used the parameterization from Eq. (1) with $\lambda_{\text{root}} = 0.2517$, and the root distribution from the pos-

terior mean estimate is shown alongside that distribution in Fig. 1. Compared with measurements from Saarinen (1996), the amount of roots at 20–60 cm is exaggerated by all of the optimized parameter values. The model provides a better fit to the data when the root conductance is high. However, the posterior distribution of the root tortuosity parameter in Fig. 6 is almost identical to the prior, so obviously there is no need to maximize plant transport at any cost.

Since the parameters ρ ($\text{m}^2 \text{kg}^{-1}$) and τ (m m^{-1}) both affect plant transport and are included in Eq. (8), one could expect them to be tightly coupled. In the posterior, however, they are only slightly correlated, with the correlation coefficient of only 0.12 in Fig. 4. This might be due to ρ having the tendency to be close to its the lower limit. The root-ending area parameter ρ has a notable negative correlation with the air diffusion coefficient $f_{D,a}$ (–). This follows directly from the fact that increased root-ending area increases root conductance, as does faster diffusion through the air-filled aerenchyma cells, via Eq. (8).

5.3.3 Diffusion

The masses of the diffusion coefficient parameters $f_{D,a}$ (–) and $f_{D,w}$ (–) in the posterior distributions (Fig. 6j and k) are within the rather permissive priors having the value of 0.8. The parameter $f_{D,w}$ is optimized close to the upper limit of 1. Kaiser et al. (2016) note that these parameters are not well known and use for both of them the value of 0.8. Constraining the model with the CO_2 flux measurements results in the diffusion component not correlating with the amount of methane produced via anaerobic peat decomposition.

5.3.4 Ebullition

Ebullition is very strongly tied to diffusion in the flux estimates with parameters from the posterior, as is shown in Fig. 5. The flux component time series in Fig. 12 shows that ebullition to the surface is a small fraction (approximately 0–3 % with optimized parameters) of the total flux. Similarly, Wania et al. (2010) report almost virtually no ebullition to the surface. This result is highly dependent on the type of the wetland; for instance, Kaiser et al. (2016) report high ebullition fluxes for a polygonal tundra in the Siberian permafrost region, where the ice-free soil layer reaches only about 30 cm depth during summer. Variation between different sites is very large and depends on whether the water reaches the surface at times of high CH_4 emission.

Contrasting with this, in the simulations with the non-hierarchically optimized parameters, a major part of the diffusive flux, which comprises around 30 % of the total flux for most years, is transported by ebullition (Fig. 8) and diffusion is a major flux component, even though ebullition to the surface accounts for only 5 % of the total flux. Since ebullition is a fast timescale process, it was not directly constrained in the optimization with parameters, as preliminary tests re-

vealed that daily data resolution would not be sufficient for this. While finer time resolution data would have been available, using them would not have been feasible, as there is not enough knowledge about the fine structure of the wetland and micrometeorological conditions affecting the footprint area of the flux tower. It is reasonable to believe that the deviations from the daily averaged fluxes at a finer time resolution would only look like noise in the residuals, not improving our parameter posterior. Despite this, ebullition is controlled indirectly by letting CH_4 production and transport parameters control when the water column has enough CH_4 available for ebullition. This happens when the sum of the partial pressures of dissolved gases is larger than the sum of atmospheric and hydrostatic pressures as shown in Eq. (23). The high ebullition-related proportion of the diffusive flux strengthens the argument that the likelihood formulation results in the model optimizing towards parameter values that support rapid CH_4 transport.

5.4 Parameter and process identifiability

The priors of the hierarchical CH_4 production-related parameters Q_{10} (–) and ζ_{exu} (–) in Fig. 6b and d are constrained by the data, as are the hierarchical parameters themselves, shown in Fig. 14. The priors of these distributions are wider than their posteriors, which is also the case for the other production-related parameters τ_{exu} (s) and τ_{cato} (y). Both process descriptions for obtaining the anaerobic respiration are clearly needed for a good model fit, because the parameter posteriors do not have remarkable mass in the regions minimizing either of these processes (hierarchical parameters at the lower bounds or turnover rate parameters τ_{exu} and τ_{cato} at the upper bound). The covariances in Figs. 4 and 5 show that the two production processes covary slightly, with correlation coefficient -0.32 , and hence they are to that extent interchangeable. Reasonable identifiability of the Q_{10} parameters is not obvious; for example, Müller et al. (2015) optimized a corresponding parameter to end up with the parameter at the lower bound of their prescribed range. However, half of the mass of the production terms in the process correlation plot, Fig. 5, lies within a region that for production from exudates is roughly 10 % of the total production and for the production from peat decay of the order of 35 %, and hence the production processes can be said to be well constrained.

The posterior distributions of V_{R0} ($\text{mol m}^{-3} \text{s}^{-1}$) show that sqHIMMELI performs better when the heterotrophic respiration is close to being minimized, which is also aided by a posterior mean value of ΔE_R (J mol^{-1}) that is lower than the prior mean. For the oxidation parameters V_{O0} ($\text{mol m}^{-3} \text{s}^{-1}$) and ΔE_{oxid} , the situation is different: the former has the tendency of being very small, but the temperature response has the tendency of being stronger with posterior mean and MAP values above the prior mean.

Whereas the fraction of plant transport is stable and high, but still constrained, not all the parameters affecting root con-

ductivity are constrained by the data, as the root tortuosity posterior distribution follows very closely the prior form. The root-ending cross-sectional area, however, is constrained to its lower side despite there being mass also above the prior mean value. For this parameter, the importance resampling resulted in a changed posterior in that there is a lot more mass at the higher end of the distribution, as can be seen in Fig. 6h. In addition to this difference, the effects of the resampling were mostly minor. Still, the resampling informed that the roots should reside slightly higher in the peat column than suggested by the MCMC, and that the $f_{\text{CH}_4}^{\text{exu}}$ is constrained to a higher value by the data than suggested by the initial MCMC run.

The transport pathways are well identified, as can be seen in the ranges of variation in the transport characteristics in Fig. 5. Notably, the transport processes do not strongly anticorrelate implying that they are not obviously interchangeable with each other. The correlation between oxidation and plant transport suggests that uncertainty in oxidation is a major part of the uncertainty in the plant transport portion. On the other hand, there is uncertainty in the absolute magnitude of the total flux (in terms of the posterior uncertainty) and this is reflected in the strong positive correlation between plant transport and the total flux. Similar but weaker positive correlations exist between the total flux and diffusion and ebullition, which is to be expected. The variation of oxidation is around 10 % of the total flux.

5.5 Low WTD in 2006, 2010, and 2011

The calibrated sqHIMMELI model is able to describe the CH_4 flux correctly in times of low water table, which is not obvious, as other studies have indicated the challenges in parameterizations of emission models with respect to the water table depth (e.g., Zhu et al., 2014). Figure 11 shows how the model processes are described under water stress. In times of a very low water table, the plant transport component and methane production from root exudates are decreased somewhat, as is methane oxidation. This results directly from how the model is constructed, as exudate deposition to the peat column is allocated depth-wise according to the root density profile. The fact that the model continues to perform well during these years implies that this method of regulating methane emissions during dry seasons is realistic. The residuals in Fig. 11h further show that there is only a slight positive emission bias at the times of the lowest water table levels.

5.6 Predicting emissions with sqHIMMELI

Modeled CH_4 flux estimates may have large errors, as was shown in Fig. 8 with the default parameter set. The negative biases in the calibration phase that were found with the maximum a posteriori and posterior mean estimates are reasonable since the quality of the modeled input data from, e.g., a

Table 5. The p and r^2 values of the regressions of the Q_{10} (–) parameters against the mean soil temperature of the first 10 weeks of the year at the depth of 35 cm, and the ζ_{exu} parameters against the sum of the net primary production of the first 130 days of the year.

$p_{Q_{10}}$	$r^2_{Q_{10}}$	$p_{\zeta_{\text{exu}}}$	$r^2_{\zeta_{\text{exu}}}$
0.0185	0.571	$4.8\text{e} \times 10^{-6}$	0.957

land surface scheme will also contribute to the uncertainty in the model predictions. Additionally, a known constant bias can be relatively easily accounted for if the interannual variability is correctly modeled.

Compared to the estimate with the optimized annual variations of the methane-production-related parameters, the non-hierarchical posterior mean estimate produces reasonable flux estimates over the assessment period, with twice the variability in fluxes compared to the posterior mean estimate, even though the average of the errors is closer to zero. The variability is seen in Fig. 7. The hierarchical posterior mean, on the other hand, does produce very steady estimates of the CH_4 flux compared with observations even though there is a downward bias of 23 %, and the smaller interannual variance implies better predictive skill. The same is true to a lesser extent also for the maximum a posteriori estimate.

In order to be able to utilize the information regarding the annual variability in the posterior mean estimate for the future prediction of CH_4 emissions, the values of the hierarchical parameters need to be estimated for the simulation years. A simple regression analysis of the hierarchical variables with respect to relevant input data was performed in order to find out if such estimation is possible. As the explaining variables, means, minimums, and maximums of NPP, water table depth, and soil temperature at different depths and over different periods of time were looked at. These time periods were June, July, August, and various different amounts of days from the start of the year.

The analysis revealed that the mean soil temperature of the first 10 weeks (70 days) of the year at the depth of 30–40 cm, denoted here by T_{30-40}^{70} , is the best single-variable predictor of the Q_{10} value for that year, and for ζ_{exu} , it is the sum of NPP from the first 130 days of the year, denoted by NPP^{130} . This is hardly surprising, since the peat decomposition process regulated by the parameter Q_{10} is driven by soil temperature, and the anaerobic respiration from exudates controlled by the parameter ζ_{exu} is driven by the NPP input. These variables also indicate that the timing of the start of the growing season might play a role in determining the parameters. Possible mechanisms could include, e.g., effects of the start of growing season on development of the microbe populations in the spring. However, further analysis would be needed to confirm this.

The p values summarizing the reliabilities of the regressions and the r^2 values, which are the coefficients of determination of the fit, are presented in Table 5. The r^2 values explain what fraction of the variance of the dependent (predicted) variable is explained by the independent (explaining) variables. The p and r^2 values uncover that the hierarchical modeling reveals a clear-cut reliable relationship between the early NPP and the optimal ζ_{exu} parameter ($p = 5 \times 10^{-6}$, $r^2 = 0.957$). This provides new insight into future model development and exemplifies why such a hierarchical description of variables is valuable in Bayesian optimization in a geophysical model context.

For the other interannually changing parameter, Q_{10} , the soil temperatures explain only slightly over half of the variation ($p = 0.0185$, $r^2 = 0.571$). Since the effect of this parameter is very important for the total methane flux, this result leaves lots of room for further analysis. The hierarchical parameters Q_{10} and ζ_{exu} for each year can be estimated with

$$Q_{10} = 3.86 \overline{T_{30-40}^{70}} + 1.76 \quad (25)$$

$$\zeta_{\text{exu}} = -46500\text{NPP}^{130} + 0.431, \quad (26)$$

where the temperatures are in $^{\circ}\text{C}$, and the units of NPP are $\text{mol m}^{-2} \text{s}^{-1}$.

A leave-one-out cross validation (LOO-CV; see, e.g., Gelman et al., 2013) of the regression modeling was performed by optimizing the hierarchical parameters with respect to the cost function in Eq. (24) leaving one year at a time out, calculating the estimates for the hierarchical parameters based on the results obtained for other years, and predicting the CH_4 emissions for the year that was left out. The results of the cross validation are shown in Figs. 7b and 8. The cross-validated results are comparable in terms of annual performance to the non-hierarchical posterior mean. Despite the relatively good performance of the non-hierarchical posterior mean simulation, we note that the cross-validated result should be relied on more for prediction, since the well-predictable ζ_{exu} parameters contain useful information that is not available in the non-hierarchical posterior mean estimate. A hybrid between these approaches could be also used, using the regression-modeled values for the ζ_{exu} parameters and the mean for Q_{10} , to minimize the risk of major annual biases due to unsuccessful prediction of the Q_{10} parameters.

As Fig. 7b shows, much of the error in the cross validation actually comes from challenges estimating the year 2007, which is missing the peak season observations, and therefore the error percentage (in terms of the annual observed flux) is easily high, especially as the start of season is modeled with a delay, which is readily apparent in Fig. 9, and in this sense the negative bias in Fig. 7 gives an unnecessarily pessimistic view of the model performance. For the CO_2 fluxes, it can be noted that there is a persistent positive bias of some tens of percent, but the observations are very noisy and due to the processing for the use in the cost function, they might have biases. The effect of a small bias on the parameter posterior

distribution is, however, minor, since the carbon dioxide observations were given less weight in the cost function than the methane observations. Hence, given their uncertainty, the optimized fit to the measurement data can, also in the cross validation as in the other experiments, be seen as acceptable.

6 Conclusions

In this study, Bayesian calibration of a new process-based wetland CH₄ emission model, sqHIMMELI, was performed using Markov chain Monte Carlo methods, hierarchical statistical modeling of methane production related parameters, Box–Jenkins-type time series modeling, and importance resampling against daily methane and carbon dioxide flux data from the Siikaneva flux measurement site in Finland. The results show that the modeled processes and the estimated parameters are identifiable with the flux data. The parameter correlations and process correlations from random sampling the posterior reveal that there are no redundant processes in the model description. However, a few strong correlations between parameters exist, reminding about the difficulty of strictly interpreting parameter values to be connected to isolated physical processes. The optimized model fits well to the data in that the modeled fluxes fit within a range from the data that is expected based on the error modeling.

Preliminary results obtained also suggest that estimation of the annual variation of the parameters controlling methane production from anaerobic respiration of root exudates is feasible and may help to improve the future estimates of the boreal wetland CH₄ emissions.

For future studies, combining observations from several sites and optimizing them together with the methods presented here in conjunction with independent validation can provide valuable information about the uncertainties related to wetland emission modeling and about how to best improve the quality of predicting wetland methane emissions in land surface schemes of climate models.

Code and data availability. The HIMMELI source code is available as a supplement to the publication of Raivonen et al. (2017). The sqHIMMELI model code is available as a supplement to this publication.

The model input data and the flux measurement data are available upon a reasonable request to the lead author.

Appendix A: Error model for residuals

In Sect. 4.2.1, we described the error models as AR(1)/ARMA(2,1) models where the residuals are Laplace distributed. Intuitively, these models can be thought of as characterizing the “inertia” or “memory” in the model–observation discrepancy. Formally, the observation equation for our statistical inference problem can be written as

$$y_t = \mathbf{x}_t + \mathbf{r}_t^* \quad (\text{A1})$$

$$\mathbf{x}_t = M(\mathbf{x}_{t-1}, \mathbf{z}_{t-1}, \boldsymbol{\theta}). \quad (\text{A2})$$

The vector notation for \mathbf{y} and \mathbf{r}^* in Eq. (A1) refers to the fact that at each time t there can be observations of both methane and carbon dioxide, and M in Eq. (A2) denotes the model (sqHIMMELI) advancing the model state \mathbf{x}_{t-1} forward in time. The term \mathbf{z}_{t-1} is the external model forcing data. In this context, the “error model” that is referred to in the text refers to how the \mathbf{r}_t^* terms are modeled. The modeling is different for the MCMC and importance resampling steps.

Residuals terms for MCMC

For both CO₂ and CH₄, let $y'_t = \max(c_t, y_t)$, where c_t is the 14-day running mean of the gap-filled flux observations y_t . Due to the heteroscedasticity of the model error, we scale the residuals for error modeling by dividing each model prediction and observation with $\alpha|y'_t| + \gamma$, where α and γ are predetermined constants. The error-scaled residual at time t is then

$$r_t = \frac{r_t^*}{\alpha|y'_t| + \gamma}. \quad (\text{A3})$$

Let ϕ denote the lag-1 autocorrelation coefficient, meaning the correlation of the residual time series with the same residual time series 1 day later. The AR(1)-corrected residual for time t then becomes

$$r_t = \phi r_{t-1} + \epsilon_t. \quad (\text{A4})$$

The reason for the way of constructing \mathbf{y}' above was to allow for a reasonable amount of error both in the case when there is an emission spike upwards and when the same happens downwards, avoiding the problems if in the summer there is suddenly a day with zero CH₄ emissions, and the observation would be taken to be extremely precise (as αy_t would be small) even though the low value is rather due to noise.

The MCMC experiment was performed with a cost function that permissively allowed for exploration of the parameter space. The α and γ were 0.4 and 0.00075 for CH₄ and 1.0 and 0.029 for CO₂, respectively, and the lag-1 autocorrelation coefficient used was 0.6. Uncertainties motivating such a permissive error description include uncertainties in the NPP model, inadequacies in the model description of the peat column and lack of spatial heterogeneity in the model description, filled gaps in the water table depth data, errors

from interpolation of the soil temperature data and heat transfer, and other unknown error sources. The same model error description was used for all MCMC model simulations.

Residuals for importance resampling

The sum of the absolute values of the ϵ_t terms appears in the objective function, Eq. (24), but the AR(1)-modeled values are in the end not independent and do not accurately follow the Laplace distribution, in part because generous values were chosen for α and γ that allowed for easier exploration of the parameter space. The objective function used for importance resampling fixes these problems.

For choosing the order of autoregressive moving average model (the ARMA(p, q) model), the different models up to order $p = q = 4$ were fitted, and the one whose fitting yielded the lowest Bayesian information criterion was picked. After making sure that the fitted residuals are independent by calculating the Durbin–Watson statistic, the order of (p, q) = (2, 1) was chosen. In place of Eq. (A4), the error model for the residuals is then written as

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + \theta \epsilon_{t-1} + \epsilon_t, \quad (\text{A5})$$

where the ϕ parameters are the AR model parameters and the θ is the MA part.

The scaling of the model residuals for choosing the ARMA parameters and the values for α and γ above (separately for the CH₄ and CO₂ time series) was done by effectively calculating the 2-week running mean of the variances of the flux from observations for each day of the year. More explicitly, let

$$\hat{y}_t = \sqrt{(\nabla_{\text{doi}=\tau}[y_t])} \quad (\text{A6})$$

denote the standard deviation of the observed fluxes for a given day of the year over the whole observation dataset. Then, the residuals are scaled as before by

$$r_t = \frac{r_t^*}{\boldsymbol{\alpha} \mathbf{h}^T \widehat{\mathbf{y}}_t + \gamma}, \quad (\text{A7})$$

where \mathbf{h}^T is a vector of length 14 with each element having a value of $\frac{1}{14}$, and $\widehat{\mathbf{y}}_t$ is the vector with elements $\widehat{y}_{t-7}, \dots, \widehat{y}_{t+6}$. Let $\Psi(b_i)$ denote the value of a discretization of the standard Laplace distribution at point $b_i \in \{b_1, \dots, b_{N_b}\}$, and let $S_{\alpha, \beta}^{\widehat{\phi}_1, \widehat{\phi}_2, \widehat{\theta}}(b_i)$ denote the empirical probability density function of the set of the transformed residual terms, the ϵ_t terms in Eq. (A5), again at point b_i . The parameters $\widehat{\phi}_1, \widehat{\phi}_2$, and $\widehat{\theta}$ are the optimized ARMA model parameters from fitting the model.

The ARMA(2,1) model parameters and the parameters α and γ are determined for the importance resampling by minimizing the Kullback–Leibler divergence,

$$D_{\text{KL}} \left(\Psi \parallel S_{\alpha, \gamma}^{\widehat{\phi}_1, \widehat{\phi}_2, \widehat{\theta}} \right) = - \sum_{i=1}^{i=N_b} \log \Psi(b_i) \frac{S_{\alpha, \gamma}^{\widehat{\phi}_1, \widehat{\phi}_2, \widehat{\theta}}(b_i)}{\Psi(b_i)}, \quad (\text{A8})$$

which is a measure of similarity between distributions. Effectively, we fit the error model parameters to make sure that the modeled residuals really are Laplace distributed and independent. The parameters α and γ are then chosen to be

$$\alpha, \gamma = \arg \min_{\alpha, \gamma} D_{\text{KL}} \left(\Psi \parallel S_{\alpha, \gamma}^{\tilde{\phi}_1, \tilde{\phi}_2, \tilde{\theta}} \right), \quad (\text{A9})$$

and the ARMA parameters are chosen to be the ones from the model fit with those parameters α and γ minimizing the KL divergence. The bound optimization by quadratic approximation (BOBYQA) optimization algorithm (Powell, 2009) was used to carry out the minimization. The procedure was performed for 50 parameters vectors randomly sampled from the posterior of the MCMC run and the medians of these values, which were for all parameters narrowly distributed, were the final ones picked for the likelihood used in importance resampling. The actual values of these parameters for methane were $\alpha^{\text{CH}_4} = 0.594$, $\gamma^{\text{CH}_4} = 1.38 \times 10^{-6}$, $\phi_1^{\text{CH}_4} = 1.30$, $\phi_2^{\text{CH}_4} = -0.325$, and $\theta^{\text{CH}_4} = -0.770$; correspondingly for carbon dioxide $\alpha^{\text{CO}_2} = 0.443$, $\gamma^{\text{CO}_2} = 3.96 \times 10^{-3}$, $\phi_1^{\text{CO}_2} = 1.21$, $\phi_2^{\text{CO}_2} = -0.242$, and $\theta^{\text{CO}_2} = -0.738$ were used. The histograms of the ϵ_t values and the autocorrelation functions are shown in Fig. 2.

Appendix B: A basic outline of MCMC

MCMC methods are a class of Bayesian methods that can be used for obtaining the probability distribution $p(\theta|\mathbf{y})$ for a parameter vector $\theta \in \mathbf{R}^n$ given data $\mathbf{y} \in \mathbf{R}^k$. According to Bayes' theorem, this can be written as

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})}, \quad (\text{B1})$$

where $p(\mathbf{y}|\theta)$ is the likelihood (in this work, the first two terms on the right-hand side of Eq. 24), and $p(\theta)$ is the prior (the last term). The evidence, $p(\mathbf{y})$, is often very difficult to evaluate, but in MCMC this is not needed, because MCMC algorithms evaluate ratios of successive evaluations of $p(\theta|\mathbf{y})$, making the denominators cancel out, and hence the evidence term can be dropped.

MCMC sampling starts by taking some starting value θ and calculating the objective function (also known as “cost function”) value $J(\theta) \in \mathbb{R}$ – the notation here is the same as in Eq. (24). The algorithm then draws a new sample of the parameter vector, θ' from a prescribed “proposal distribution” $q(\theta)$, and evaluates $J(\theta')$. It accepts the new parameter vector with a probability that depends on the value of $J(\theta')$ and the objective function value of the previous accepted parameter, $J(\theta)$. If the value is accepted, the chain will move to position θ' (setting $\theta \leftarrow \theta'$), and if θ' is rejected, the value θ will be repeated in the chain. After this, a new value, sampled from $q(\theta)$ (which is possibly a different distribution from the one used at the previous iteration as θ may have changed),

will be proposed and the whole process is repeated. In the end, the procedure will produce a chain of parameter values.

According to Markov chain theory, the sampled parameter values will eventually follow the “target distribution” $p(\theta|\mathbf{y})$ meaning that in such a case picking a random element from the chain amounts to drawing a sample directly from the target distribution. As real-life Markov chains are of finite length, the “posterior distribution” obtained from the chain is an approximation of the underlying target distribution.

In practice, this means that with MCMC it is possible to find a good approximation of the probability density function of the parameter vector θ in the cases where the model is not suitable for analytical treatment. From this probability density function, valuable information such as modes, variances, and correlations of the parameters can be analyzed. The posterior also reveals which parameters are constrained by the data and which are not.

For efficient convergence of the chain to the posterior distribution, a good estimate of $q(\theta)$ is needed. The adaptive Metropolis algorithm automatically calibrates the proposal during the MCMC.

Appendix C: Metropolis within Gibbs sampling of the parameters

The hierarchical parameters Q_{10}^{year} and $\zeta_{\text{exu}}^{\text{year}}$ are denoted here generically by θ^i , where i refers to the different years. The priors of these parameters are defined by the hyperparameters μ_i and σ_i that determine the prior of θ^i by

$$\theta^i \sim N(\mu_i, \sigma_i^2). \quad (\text{C1})$$

The unknown hyperparameters μ_i and σ_i^2 have probabilistic models:

$$\mu_i \sim N(\mu_0, \tau_0^2) \quad (\text{C2})$$

$$\sigma_i^2 \sim \text{Inv-}\chi^2(n_0, \sigma_0^2), \quad (\text{C3})$$

where μ_0 and τ_0^2 define the mean and variance of the hyper-prior of μ_i , $n_0 \in \mathbb{N}$ defines the number of degrees of freedom of the $\text{Inv-}\chi^2$ distribution, and σ_0^2 is the expected value of the scaled $\text{Inv-}\chi^2$ distribution.

In Gibbs sampling, the full conditional posterior distributions of the hyperparameters and the parameters θ_i are sampled in turns. Due to the conjugacy of the normal distribution and the scaled $\text{Inv-}\chi^2$ distribution, closed-form expressions exist for sampling from $p(\mu_i|\sigma^2, \mu_0, \sigma_0^2, \theta^i)$ and $p(\sigma_i^2|\sigma_0^2, n_0, \theta^i)$, where μ is the current mean of the parameters θ_i and σ^2 is their variance. The Gibbs sampling therefore consists of three steps:

1. Draw μ_i from

$$\mu_i | \mu, \sigma^2 \sim N \left(\frac{\frac{n_i \bar{\theta}^i}{\sigma^2} + \frac{\mu_0}{\tau_0^2}}{\frac{n_i}{\sigma^2} + \frac{1}{\tau_0^2}}, \frac{1}{\frac{n_i}{\sigma^2} + \frac{1}{\tau_0^2}} \right), \quad (\text{C4})$$

2. draw σ_i^2 from

$$\sigma_i^2 | \theta, \quad (C5)$$

$$\mu \sim \text{Inv} - \chi^2 \left(n_0 + n_i, \frac{\sigma_0^2 n_0 + \sum_{j=1}^{n_i} (\theta_j^i - \mu_j)^2}{n_0 + n_i} \right),$$

and

3. draw the parameters θ_i (and the non-hierarchical parameters) with MCMC, since closed-form expression for $p(\theta | \phi, y)$, where ϕ denotes all the different hyperparameters, is not available.

In this work, the value of the parameter τ_0^2 was set to the value of σ_0^2 , n_i is the number of years, and the value of n_0 was set to 9.

The means and variances obtained this way describe the interannual variability of the parameters, and not including them as parameters in the MCMC sampling reduces the dimension of space that the MCMC sampler needs to explore, speeding up convergence of the posterior distribution.

Appendix D: Importance resampling

Importance resampling is a method for obtaining samples from a desired (unnormalized) distribution $q(\theta)$ by reevaluating samples from a similar distribution from which it is known how samples are generated, $p(\theta)$. It is usually remarkably faster than, for instance, re-performing an MCMC experiment.

The samples $\theta_1 \dots \theta_N$ are first drawn from $p(\theta)$ (in our case, randomly picked from the MCMC chain), and at these points the new posterior density $q(\theta)$ is evaluated. For each of these, the “weights” are defined by $w(\theta_i) = \frac{q(\theta_i)}{p(\theta_i)}$. The samples from the distribution $q(\theta)$ are then generated by sampling according to the set of normalized weights, $\tilde{w}(\theta_i) = \frac{w(\theta_i)}{\sum_{j=1}^N w(\theta_j)}$. The sampling is performed without replacement. For further details, see, e.g., Gelman et al. (2013).

Appendix E: NPP and LAI

We estimated the net photosynthesis rate, P_n , of vascular plants of Siikaneva for the years 2005–2014 by utilizing regression models of gross photosynthesis, P_g , and autotrophic respiration R_a formulated for peatland vegetation (Riutta et al., 2007a, b; Raivonen et al., 2015). The model of the P_g of sedge and dwarf shrub canopy (Riutta et al., 2007a) simulates the carbon uptake driven by photosynthetically active radiation (PAR), WTD, and air temperature. The model of R_a (Raivonen et al., 2015) simulates the respiration rate driven by air temperature and WTD, and was parameterized for sedges only.

Both P_g and R_a models simulate the carbon fluxes per soil surface area and the rate depends on the LAI. We simulated the LAI using a lognormal function presented by Wilson et al. (2007). Parameter values of the LAI model were obtained by averaging the values reported by Wilson et al. (2007) for the vascular species abundant at Siikaneva. For the growing season peak LAI, we used the maximum LAI observed at the eddy covariance footprint area, viz. approximately $0.4 \text{ m}^2 \text{ m}^{-2}$ (Riutta et al., 2007b). We also included a constant wintertime LAI since a significant green sedge biomass may overwinter, approximately 15 % of the maximum (Saarinen, 1998; Bernard and Hankinson, 1979). The overwintering LAI at Siikaneva would thus be $0.05 \text{ m}^2 \text{ m}^{-2}$. The same LAI was used for all the years, and this LAI also was given as the input for the CH_4 transport model.

The daily averages of P_n were calculated by subtracting R_a from P_g . The models were run with measured meteorological data. We determined the photosynthetically active seasons based on snowmelt dates in spring or arrival of snow cover in autumn from the reflected PAR data or based on air temperature (permanently greater than 5°C assumed to be the growing season). After the calculation, we compared the resulting P_n of vascular vegetation of the year 2005 to eddy covariance CO_2 fluxes from Siikaneva. We used the gross primary production (GPP) derived from the measured net ecosystem exchange (NEE) by Aurela et al. (2007). The GPP was on average 4.5-fold compared with our P_n , with a R^2 of 0.9. GPP also includes the photosynthesis of *Sphagnum* mosses as well as CO_2 released in autotrophic respiration. *Sphagnum* accounted for 20–40 % of the GPP in the study by Riutta et al. (2007a) and autotrophic respiration has been observed to be roughly 50 % of GPP (Gifford, 1994). Consequently, the NPP of vascular vegetation can be estimated by multiplying the GPP with 0.7×0.5 . This estimate was still 1.56-fold compared with the P_n for the year 2005. Since the P_n also was lower than generally reported for peatlands, we chose to trust the eddy covariance measurement and scaled the P_n of all the years upwards by multiplying it by 1.56. For further details, please consult Raivonen et al. (2017).

The Supplement related to this article is available online at <https://doi.org/10.5194/gmd-11-1199-2018-supplement>.

Author contributions. JS designed the study with help from the co-authors, programmed the algorithms, performed the model simulations, analyzed the results, and prepared the manuscript and the figures. MR provided and validated the input data and helped with the interpretation of the results. LB contributed several model sub-routines and helped to interpret the results. ML provided assistance with getting the technical aspects of the Bayesian analysis right. OP provided insight into the data used. JM, TV, and TA provided helpful critical comments and suggestions that helped to improve the manuscript substantially.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to thank the University of Helsinki researchers Pavel Alekseychik and Ivan Mammarella, and Janne Rinne from Lund University for valuable comments and input regarding the Siikaneva measurement site data. We would also like to thank Heikki Haario from Lappeenranta University of Technology, Janne Hakkarainen from Finnish Meteorological Institute, and Samuli Siltanen from University of Helsinki for comments regarding the mathematical aspects of the study.

This work has been supported by the EU LIFE+ project MONIMET LIFE12 ENV/FI/000409, and the EU FP7 project EMBRACE. We additionally acknowledge funding from the RED platform of the Lappeenranta University of Technology and thank the Academy of Finland Center of Excellence (272041), The Strategic Research Council at the Academy of Finland project (312932), CARB-ARC (285630), ICOS Finland (281255), ICOS-ERIC (281250), NCoE eSTICC (57001), EU-H2020 CRESCENDO (641816), EU-H2020 VERIFY (776810), and Academy Professor projects (284701 and 282842).

Edited by: David Lawrence

Reviewed by: two anonymous referees

References

- Arah, J. R. M. and Stephen, K. D.: A Model of the Processes Leading to Methane Emission from Peatland, *Atmos. Environ.*, 32, 3257–3264, 1998.
- Aurela, M., Tuovinen, J.-P., and Laurila, T.: Net CO₂ exchange of a subarctic mountain birch ecosystem, *Theor. Appl. Climatol.*, 70, 135–148, <https://doi.org/10.1007/s007040170011>, 2001.
- Aurela, M., Riutta, T., Laurila, T., Tuovinen, J.-V., Vesala, T., Tuittila, E.-S., Jinne, J., Haapanala, S., and Laine, J.: CO₂ exchange of a sedge fen in southern Finland—the impact of a drought period, *Tellus B*, 59, 826–837, <https://doi.org/10.1111/j.1600-0889.2007.00309.x>, 2007.
- Bellisario, L. M., Bubier, J. L., Moore, T. R., and Chanton, J. P.: Controls on CH₄ emissions from a northern peatland, *Global Biogeochem. Cy.*, 13, 81–91, <https://doi.org/10.1029/1998GB900021>, 1999.
- Bergman, I., Klarqvist, M., and Nilsson, M.: Seasonal variation in rates of methane production from peat of various botanical origins: effects of temperature and substrate quality, *FEMS Microbiol. Ecol.*, 33, 181–189, <https://doi.org/10.1111/j.1574-6941.2000.tb00740.x>, 2000.
- Bernard, J. M. and Hankinson, G.: Seasonal Changes in Standing Crop, Primary Production, and Nutrient Levels in the Carex rostrata Wetland, *Oikos*, 32, 328–336, 1979.
- Bohn, T. J., Melton, J. R., Ito, A., Kleinen, T., Spahni, R., Stocker, B. D., Zhang, B., Zhu, X., Schroeder, R., Glagolev, M. V., Maksyutov, S., Brovkin, V., Chen, G., Denisov, S. N., Eliseev, A. V., Gallego-Sala, A., McDonald, K. C., Rawlins, M. A., Riley, W. J., Subin, Z. M., Tian, H., Zhuang, Q., and Kaplan, J. O.: WETCHIMP-WSL: intercomparison of wetland methane emissions models over West Siberia, *Biogeosciences*, 12, 3321–3349, <https://doi.org/10.5194/bg-12-3321-2015>, 2015.
- Chatfield, C.: The analysis of time series: an introduction, Chapman and Hall, 4th edn., 352 pp., 1989.
- Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., LeQuere, C., Myneni, R., Piao, S., and Thornton, P.: Carbon and Other Biogeochemical Cycles, chap. 6, Cambridge University Press, Cambridge, UK, New York, NY, USA, 465–570, <https://doi.org/10.1017/CBO9781107415324.015>, 2013.
- Cresto Aleina, F., Runkle, B. R. K., Brücher, T., Kleinen, T., and Brovkin, V.: Upscaling methane emission hotspots in boreal peatlands, *Geosci. Model Dev.*, 9, 915–926, <https://doi.org/10.5194/gmd-9-915-2016>, 2016.
- Davidson, E. A., Janssens, I. A., and Luo, Y.: On the variability of respiration in terrestrial ecosystems: moving beyond Q₁₀, *Glob. Change Biol.*, 12, 154–164, <https://doi.org/10.1111/j.1365-2486.2005.01065.x>, 2006.
- Frolking, S., Talbot, J., Jones, M. C., Treat, C. C., Kauffman, J. B., Tuittila, E.-S., and Roulet, N.: Peatlands in the Earth's 21st century climate system, *Environ. Rev.*, 19, 371–396, <https://doi.org/10.1139/a11-014>, 2011.
- Gedney, N., Cox, P. M., and Huntingford, C.: Climate feedback from wetland methane emissions, *Geophysical Research Letters*, 31, L20503, <https://doi.org/10.1029/2004GL020919>, 2004.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D.: Bayesian Data Analysis, Chapman and Hall/CRC, 3rd edn., 675 pp., 2013.
- Gifford, R. M.: The Global Carbon Cycle: a Viewpoint on the Missing Sink, *Funct. Plant Biol.*, 21, 1–15, <https://doi.org/10.1071/PP9940001>, 1994.
- Grant, R. F. and Roulet, N. T.: Methane efflux from boreal wetlands: Theory and testing of the ecosystem model Ecosys with chamber and tower flux measurements, *Global Biogeochem. Cy.*, 16, 1054, <https://doi.org/10.1029/2001GB001702>, 2002.
- Haario, H., Saksman, E., and Tamminen, J.: An Adaptive Metropolis Algorithm, *Bernoulli*, 7, 223–242, 2001.
- Hararuk, O., Xia, J., and Luo, Y.: Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method, *J. Geophys. Res.-Biogeo.*, 119, 403–417, <https://doi.org/10.1002/2013JG002535>, 2014.

- Hornibrook, E. R., Longstaffe, F. J., and Fyfe, W. S.: Spatial distribution of microbial methane production pathways in temperate zone wetland soils: Stable carbon and hydrogen isotope evidence, *Geochim. Cosmochim. Ac.*, 61, 745–753, [https://doi.org/10.1016/S0016-7037\(96\)00368-7](https://doi.org/10.1016/S0016-7037(96)00368-7), 1997.
- IPCC: Summary for Policymakers, book section SPM, Cambridge University Press, Cambridge, UK, New York, NY, USA, 1–30, <https://doi.org/10.1017/CBO9781107415324.004>, 2013.
- Joabsson, A. and Christensen, T. R.: Methane emissions from wetlands and their relationship with vascular plants: an Arctic example, *Glob. Change Biol.*, 7, 919–932, <https://doi.org/10.1046/j.1354-1013.2001.00044.x>, 2001.
- Juottonen, H.: Archaea, Bacteria, and methane production along environmental gradients in fens and bogs, PhD thesis, University of Helsinki, 2008.
- Kaiser, S., Göckede, M., Castro-Morales, K., Knoblauch, C., Ekici, A., Kleinen, T., Zubrzycki, S., Sachs, T., Wille, C., and Beer, C.: Process-based modelling of the methane balance in periglacial landscapes (JSBACH-methane), *Geosci. Model Dev.*, 10, 333–358, <https://doi.org/10.5194/gmd-10-333-2017>, 2017.
- Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., Cameron-Smith, P., Castaldi, S., Chevallier, F., Feng, L., Fraser, A., Heimann, M., Hodson, E. L., Houweling, S., Josse, B., Fraser, P. J., Krummel, P. B., Lamarque, J. F., Langenfelds, R. L., Le Quere, C., Naik, V., O'Doherty, S., Palmer, P. I., Pison, I., Plummer, D., Poulter, B., Prinn, R. G., Rigby, M., Ringeval, B., Santini, M., Schmidt, M., Shindell, D. T., Simpson, I., Spahni, R., Steele, L. P., Strode, S. A., Sudo, K., Szopa, S., van der Werf, G. R., Voulgarakis, A., van Weele, M., Weiss, R. F., Williams, J. E., and Zeng, G.: Three decades of global methane sources and sinks, *Nat. Geosci.*, 6, 813–823, <https://doi.org/10.1038/ngeo1955>, 2013.
- Kullback, S. and Leibler, R. A.: On Information and Sufficiency, *Ann. Math. Statist.*, 22, 79–86, <https://doi.org/10.1214/aoms/1177729694>, 1951.
- Lai, D.: Methane Dynamics in Northern Peatlands: A Review, *Pedosphere*, 19, 409–421, [https://doi.org/10.1016/S1002-0160\(09\)00003-4](https://doi.org/10.1016/S1002-0160(09)00003-4), 2009a.
- Lai, D. Y. F.: Modelling the effects of climate change on methane emission from a northern ombrotrophic bog in Canada, *Environ. Geol.*, 58, 1197–1206, <https://doi.org/10.1007/s00254-008-1613-5>, 2009b.
- Laine, M.: Adaptive MCMC methods with applications in environmental and geophysical models, PhD thesis, Lappeenranta University of Technology, 2008.
- Lehtonen, I., Ruosteenoja, K., and Jylhä, K.: Projected changes in European extreme precipitation indices on the basis of global and regional climate model ensembles, *Int. J. Climatol.*, 34, 1208–1222, <https://doi.org/10.1002/joc.3758>, 2014.
- Mäkelä, J., Susiluoto, J., Markkanen, T., Aurela, M., Järvinen, H., Mammarella, I., Hagemann, S., and Aalto, T.: Constraining ecosystem model with adaptive Metropolis algorithm using boreal forest site eddy covariance measurements, *Nonlin. Processes Geophys.*, 23, 447–465, <https://doi.org/10.5194/npg-23-447-2016>, 2016.
- Mammarella, I., Peltola, O., Nordbo, A., Järvi, L., and Rannik, Ü.: Quantifying the uncertainty of eddy covariance fluxes due to the use of different software packages and combinations of processing steps in two contrasting ecosystems, *Atmos. Meas. Tech.*, 9, 4915–4933, <https://doi.org/10.5194/amt-9-4915-2016>, 2016.
- Mastepanov, M., Sigsgaard, C., Tagesson, T., Ström, L., Tamstorf, M. P., Lund, M., and Christensen, T. R.: Revisiting factors controlling methane emissions from high-Arctic tundra, *Biogeosciences*, 10, 5139–5158, <https://doi.org/10.5194/bg-10-5139-2013>, 2013.
- Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., Eliseev, A. V., Denisov, S. N., Hopcroft, P. O., Lettenmaier, D. P., Riley, W. J., Singarayer, J. S., Subin, Z. M., Tian, H., Zürcher, S., Brovkin, V., van Bodegom, P. M., Kleinen, T., Yu, Z. C., and Kaplan, J. O.: Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP), *Biogeosciences*, 10, 753–788, <https://doi.org/10.5194/bg-10-753-2013>, 2013.
- Müller, J., Paudel, R., Shoemaker, C. A., Woodbury, J., Wang, Y., and Mahowald, N.: CH₄ parameter estimation in CLM4.5bgc using surrogate global optimization, *Geosci. Model Dev.*, 8, 3285–3310, <https://doi.org/10.5194/gmd-8-3285-2015>, 2015.
- Nedwell, D. B. and Watson, A.: CH₄ production, oxidation and emission in a U.K. ombrotrophic peat bog: Influence of SO₄²⁻ from acid rain, *Soil Biol. Biochem.*, 27, 893–903, [https://doi.org/10.1016/0038-0717\(95\)00018-A](https://doi.org/10.1016/0038-0717(95)00018-A), 1995.
- Nilsson, M. and öquist, M.: Partitioning Litter Mass Loss into Carbon Dioxide and Methane in Peatland Ecosystems, 131–144, American Geophysical Union, <https://doi.org/10.1029/2008GM000819>, 2013.
- Petrescu, A. M. R., Lohila, A., Tuovinen, J.-P., Baldocchi, D. D., Desai, A. R., Roulet, N. T., Vesala, T., Dolman, A. J., Oechel, W. C., Marcolla, B., Friborg, T., Rinne, J., Matthes, J. H., Merbold, L., Meijide, A., Kiely, G., Sottocornola, M., Sachs, T., Zona, D., Varlagin, A., Lai, D. Y. F., Veenendaal, E., Partenier, F.-J. W., Skiba, U., Lund, M., Hensen, A., van Huissteden, J., Flanagan, L. B., Shurpali, N. J., Grünwald, T., Humphreys, E. R., Jackowicz-Korczynski, M., Aurela, M. A., Laurila, T., Grünig, C., Corradi, C. A. R., Schrier-Uijl, A. P., Christensen, T. R., Tamstorf, M. P., Mastepanov, M., Martikainen, P. J., Verma, S. B., Bernhofer, C., and Cescatti, A.: The uncertain climate footprint of wetlands under human pressure, *P. Natl. Acad. Sci. USA*, 112, 4594–4599, <https://doi.org/10.1073/pnas.1416267112>, 2015.
- Powell, M. J. D.: The BOBYQA algorithm for bound constrained optimization without derivatives, *Tech. rep.*, 2009.
- Prather, M. J., Holmes, C. D., and Hsu, J.: Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, *Geophys. Res. Lett.*, 39, L09803, <https://doi.org/10.1029/2012GL051440>, 2012.
- Raivonen, M., Mäkiranta, P., Lohila, A., Juutinen, S., Vesala, T., and Tuittila, E.-S.: A simple CO₂ exchange model simulates the seasonal leaf area development of peatland sedges, *Ecol. Model.*, 314, 32–43, <https://doi.org/10.1016/j.ecolmodel.2015.07.008>, 2015.
- Raivonen, M., Smolander, S., Backman, L., Susiluoto, J., Aalto, T., Markkanen, T., Mäkelä, J., Rinne, J., Peltola, O., Aurela, M., Lohila, A., Tomasic, M., Li, X., Larmola, T., Juutinen, S., Tuittila, E.-S., Heimann, M., Sevanto, S., Kleinen, T., Brovkin, V., and Vesala, T.: HIMMELI v1.0: Helsinki Model of METHane buiLd-up and emIssion for peatlands, *Geosci. Model Dev.*, 10, 4665–4691, <https://doi.org/10.5194/gmd-10-4665-2017>, 2017.

- Rezanezhad, F., Price, J. S., Quinton, W. L., Lennartz, B., Milojevic, T., and Cappellen, P. V.: Structure of peat soils and implications for water storage, flow and solute transport: A review update for geochemists, *Chem. Geol.*, 429, 75–84, <https://doi.org/10.1016/j.chemgeo.2016.03.010>, 2016.
- Ricciuto, D. M., Davis, K. J., and Keller, K.: A Bayesian calibration of a simple carbon cycle model: The role of observations in estimating and reducing uncertainty, *Global Biogeochem. Cy.*, 22, GB2030, <https://doi.org/10.1029/2006GB002908>, 2008.
- Richardson, A. D., Hollinger, D. Y., Burba, G. G., Davis, K. J., Flanagan, L. B., Katul, G. G., Munger, J. W., Ricciuto, D. M., Stoy, P. C., Suyker, A. E., Verma, S. B., and Wofsy, S. C.: A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, *Agr. Forest Meteorol.*, 136, 1–18, <https://doi.org/10.1016/j.agrformet.2006.01.007>, 2006.
- Riley, W. J., Subin, Z. M., Lawrence, D. M., Swenson, S. C., Torn, M. S., Meng, L., Mahowald, N. M., and Hess, P.: Barriers to predicting changes in global terrestrial methane fluxes: analyses using CLM4Me, a methane biogeochemistry model integrated in CESM, *Biogeosciences*, 8, 1925–1953, <https://doi.org/10.5194/bg-8-1925-2011>, 2011.
- Rinne, J., Riutta, T., Pihlatie, M., Aurela, M., Haapanala, S., Tuovinen, J.-P., Tuittila, E.-S., and Vesala, T.: Annual cycle of methane emission from a boreal fen measured by the eddy covariance technique, *Tellus B*, 59, 449–457, <https://doi.org/10.1111/j.1600-0889.2007.00261.x>, 2007.
- Riutta, T., Laine, J., Aurela, M., Rinne, J., Vesala, T., Laurila, T., Haapanala, S., Pihlatie, M., and Tuittila, E.-S.: Spatial variation in plant community functions regulates carbon gas dynamics in a boreal fen ecosystem, *Tellus B*, 59, 838–852, <https://doi.org/10.1111/j.1600-0889.2007.00302.x>, 2007a.
- Riutta, T., Laine, J., and Tuittila, E.-S.: Sensitivity of CO₂ Exchange of Fen Ecosystem Components to Water Level Variation, *Ecosystems*, 10, 718–733, <https://doi.org/10.1007/s10021-007-9046-7>, 2007b.
- Roberts, G. O., Gelman, A., and Gilks, W. R.: Weak convergence and optimal scaling of random walk Metropolis algorithms, *Ann. Appl. Probab.*, 7, 110–120, <https://doi.org/10.1214/aoap/1034625254>, 1997.
- Ruosteenoja, K., Jylhä, K., and Kämäräinen, M.: Climate Projections for Finland Under the RCP Forcing Scenarios, *Geophysica*, 51, 17–50, 2016.
- Saarinén, T.: Biomass and production of two vascular plants in a boreal mesotrophic fen, *Can. J. Botany*, 74, 934–938, <https://doi.org/10.1139/b96-116>, 1996.
- Saarinén, T.: Demography of *Carex rostrata* in boreal mesotrophic fen: shoot dynamics and biomass development, *Ann. Bot. Fenn.*, 203–209, 1998.
- Sanso, B., Forest, C. E., and Zantedeschi, D.: Inferring climate system properties using a computer model, *Bayesian Anal.*, 3, 1–37, <https://doi.org/10.1214/08-BA301>, 2008.
- Saunio, M., Bousquet, P., Poulter, B., Peregón, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S., Janssens-Maenhout, G., Tubiello, F. N., Castaldi, S., Jackson, R. B., Alexe, M., Arora, V. K., Beerling, D. J., Bergamaschi, P., Blake, D. R., Brailsford, G., Brovkin, V., Bruhwiler, L., Crevoisier, C., Crill, P., Covey, K., Curry, C., Frankenberg, C., Gedney, N., Höglund-Isaksson, L., Ishizawa, M., Ito, A., Joos, F., Kim, H.-S., Kleinen, T., Krümmel, P., Lamarque, J.-F., Langenfelds, R., Locatelli, R., Machida, T., Maksyutov, S., McDonald, K. C., Marshall, J., Melton, J. R., Morino, I., Naik, V., O'Doherty, S., Parmentier, F.-J. W., Patra, P. K., Peng, C., Peng, S., Peters, G. P., Pison, I., Prigent, C., Prinn, R., Ramonet, M., Riley, W. J., Saito, M., Santini, M., Schroeder, R., Simpson, I. J., Spahni, R., Steele, P., Takizawa, A., Thornton, B. F., Tian, H., Tohjima, Y., Viovy, N., Voulgarakis, A., van Weele, M., van der Werf, G. R., Weiss, R., Wiedinmyer, C., Wilton, D. J., Wiltshire, A., Worthly, D., Wunch, D., Xu, X., Yoshida, Y., Zhang, B., Zhang, Z., and Zhu, Q.: The global methane budget 2000–2012, *Earth Syst. Sci. Data*, 8, 697–751, <https://doi.org/10.5194/essd-8-697-2016>, 2016.
- Schuldt, R. J., Brovkin, V., Kleinen, T., and Winderlich, J.: Modelling Holocene carbon accumulation and methane emissions of boreal wetlands – an Earth system model approach, *Biogeosciences*, 10, 1659–1674, <https://doi.org/10.5194/bg-10-1659-2013>, 2013.
- Segers, R.: Methane production and methane consumption: a review of processes underlying wetland methane fluxes, *Biogeochemistry*, 41, 23–51, <https://doi.org/10.1023/A:1005929032764>, 1998.
- Sriskantharajah, S., Fisher, R. E., Lowry, D., Aalto, T., Hatakka, J., Aurela, M., Laurila, T., Lohila, A., Kuitunen, E., and Nisbet, E. G.: Stable carbon isotope signatures of methane from a Finnish subarctic wetland, *Tellus B*, 64, 18818, <https://doi.org/10.3402/tellusb.v64i0.18818>, 2012.
- Stephen, K., Arah, J., Thomas, K., Benstead, J., and Lloyd, D.: Gas diffusion coefficient profile in peat determined by modelling mass spectrometric data: implications for gas phase distribution, *Soil Biol. Biochem.*, 30, 429–431, [https://doi.org/10.1016/S0038-0717\(97\)00118-1](https://doi.org/10.1016/S0038-0717(97)00118-1), 1998.
- Szafrańek-Nakonieczna, A. and Stepniewska, Z.: Aerobic and anaerobic respiration in profiles of Polesie Lubelskie peatlands, *Int. Agrophys.*, 28, 219–229, <https://doi.org/10.2478/intag-2014-0011>, 2014.
- Tang, J., Zhuang, Q., Shannon, R. D., and White, J. R.: Quantifying wetland methane emissions with process-based models of different complexities, *Biogeosciences*, 7, 3817–3837, <https://doi.org/10.5194/bg-7-3817-2010>, 2010.
- van Huissteden, J., Petrescu, A. M. R., Hendriks, D. M. D., and Rebel, K. T.: Sensitivity analysis of a wetland methane emission model based on temperate and arctic wetland sites, *Biogeosciences*, 6, 3035–3051, <https://doi.org/10.5194/bg-6-3035-2009>, 2009.
- Vile, D., Garnier, E., Shipley, B., Laurent, G., Navas, M.-L., Roumet, C., Lavorel, S., Díaz, S., Hodgson, J. G., Lloret, F., Midgley, G. F., Poorter, H., Rutherford, M. C., Wilson, P. J., and Wright, I. J.: Specific Leaf Area and Dry Matter Content Estimate Thickness in Laminae Leaves, *Ann. Bot.-London*, 96, 1129–1136, <https://doi.org/10.1093/aob/mci264>, 2005.
- Walker, T. S., Bais, H. P., Grotewold, E., and Vivanco, J. M.: Root Exudation and Rhizosphere Biology, *Plant Physiol.*, 132, 44–51, <https://doi.org/10.1104/pp.102.019661>, 2003.
- Walter, B. P. and Heimann, M.: A process-based, climate-sensitive model to derive methane emissions from natural wetlands: Application to five wetland sites, sensitivity to model parameters, and climate, *Global Biogeochem. Cy.*, 14, 745–765, <https://doi.org/10.1029/1999GB001204>, 2000.

- Wania, R.: Modelling northern peatland land surface processes, vegetation dynamics and methane emissions, PhD thesis, University of Bristol, 2007.
- Wania, R., Ross, I., and Prentice, I. C.: Implementation and evaluation of a new methane model within a dynamic global vegetation model: LPJ-WHyMe v1.3.1, *Geosci. Model Dev.*, 3, 565–584, <https://doi.org/10.5194/gmd-3-565-2010>, 2010.
- Watson, A., Stephen, K. D., Nedwell, D. B., and Arah, J. R.: Oxidation of Methane in Peat: Kinetics of CH₄ and O₂ Removal and the Role of Plant Roots, *Soil Biol. Biochem.*, 29, 1257–1267, 1997.
- Whiting, G. J. and Chanton, J. P.: Primary production control of methane emission from wetlands, *Nature*, 364, 794–795, <https://doi.org/10.1038/364794a0>, 1993.
- Wilson, D., Alm, J., Riutta, T., Laine, J., Byrne, K. A., Farrell, E. P., and Tuittila, E.-S.: A high resolution green area index for modelling the seasonal dynamics of CO₂ exchange in peatland vascular plant communities, *Plant Ecol.*, 190, 37–51, <https://doi.org/10.1007/s11258-006-9189-1>, 2007.
- Zhu, Q., Liu, J., Peng, C., Chen, H., Fang, X., Jiang, H., Yang, G., Zhu, D., Wang, W., and Zhou, X.: Modelling methane emissions from natural wetlands by development and application of the TRIPLEX-GHG model, *Geosci. Model Dev.*, 7, 981–999, <https://doi.org/10.5194/gmd-7-981-2014>, 2014.

PAPER III

III

Jouni Pulliainen, Mika Aurela, Tuomas Laurila, Tuula Aalto, Matias Takala, Miia Salminen, Markku Kulmala, Alan Barr, Martin Heimann, Anders Lindroth, Ari Laaksonen, Chris Derksen, Annikki Mäkelä, Tiina Markkanen, Juha Lemmetyinen, Jouni Susiluoto, Sigrid Dengel, Ivan Mammarella, Juha-Pekka Tuovinen, Timo Vesala

Early snowmelt significantly enhances boreal springtime carbon uptake

In *Proceedings of the National Academy of Sciences of the United States of America*,
October 17 2017, vol. 114, no. 42, pp. 11081-11086
www.pnas.org/cgi/doi/10.1073/pnas.1707889114

Copyright © Authors 2017. Reprinted with permission.



Early snowmelt significantly enhances boreal springtime carbon uptake

Jouni Pulliainen^{a,1}, Mika Aurela^a, Tuomas Laurila^a, Tuula Aalto^a, Matias Takala^a, Miia Salminen^a, Markku Kulmala^b, Alan Barr^{c,d}, Martin Heimann^{e,f}, Anders Lindroth^g, Ari Laaksonen^a, Chris Derksen^c, Annikki Mäkelä^h, Tiina Markkanen^a, Juha Lemmetyinen^a, Jouni Susiluoto^a, Sigrid Dengelⁱ, Ivan Mammarella^j, Juha-Pekka Tuovinen^a, and Timo Vesala^{f,h,j}

^aFinnish Meteorological Institute, FIN-00101 Helsinki, Finland; ^bDepartment of Physics, University of Helsinki, FI-00014 Helsinki, Finland; ^cClimate Research Division, Environment and Climate Change Canada, Toronto, ON M3H 5T4, Canada; ^dGlobal Institute for Water Security, University of Saskatchewan, Saskatoon, SK S7N 3H5, Canada; ^eMax Planck Institute for Biogeochemistry, 07701 Jena, Germany; ^fDepartment of Physics, University of Helsinki, FI-00014 Helsinki, Finland; ^gDepartment of Physical Geography and Ecosystems Science, Lund University, SE-22362 Lund, Sweden; ^hDepartment of Forest Sciences, University of Helsinki, FI-00014 Helsinki, Finland; ⁱClimate & Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94721; and ^jViikki Plant Science Centre, University of Helsinki, FI-00014 Helsinki, Finland

Edited by F. Stuart Chapin III, University of Alaska, Fairbanks, AK, and approved September 5, 2017 (received for review May 11, 2017)

We determine the annual timing of spring recovery from spaceborne microwave radiometer observations across northern hemisphere boreal evergreen forests for 1979–2014. We find a trend of advanced spring recovery of carbon uptake for this period, with a total average shift of 8.1 d (2.3 d/decade). We use this trend to estimate the corresponding changes in gross primary production (GPP) by applying in situ carbon flux observations. Micrometeorological CO₂ measurements at four sites in northern Europe and North America indicate that such an advance in spring recovery would have increased the January–June GPP sum by 29 g·C·m⁻² [8.4 g·C·m⁻² (3.7%)/decade]. We find this sensitivity of the measured springtime GPP to the spring recovery to be in accordance with the corresponding sensitivity derived from simulations with a land ecosystem model coupled to a global circulation model. The model-predicted increase in springtime cumulative GPP was 0.035 Pg/decade [15.5 g·C·m⁻² (6.8%)/decade] for Eurasian forests and 0.017 Pg/decade for forests in North America [9.8 g·C·m⁻² (4.4%)/decade]. This change in the springtime sum of GPP related to the timing of spring snowmelt is quantified here for boreal evergreen forests.

carbon uptake | earth observation | snowmelt

High-latitude warming and an associated reduction in spring snow cover are expected to have complex impacts on regional climate patterns and ecological responses (1, 2). The timing of spring recovery of photosynthesis in boreal evergreen forest following snowmelt is one of the major factors affecting the carbon balance across high latitudes (3–5). The integrated effect of anthropogenic impacts on climate causes the total radiative forcing to be positive, which increases the heat balance of the atmosphere. The largest individual cause of warming is the anthropogenic increase in CO₂ concentration (6), which is controlled by the response of the global carbon cycle to anthropogenic CO₂ emissions. Recent studies have documented, with very high confidence, that the world's forests constitute an important carbon sink (4, 6, 7). The annual total boreal forest net carbon sink is estimated to be 0.5 ± 0.1 Pg·C·y⁻¹ for the 2000–2007 period (4). Based on coupled carbon cycle–climate modeling, this sink is estimated to be increasing by 0.014 Pg·C·y⁻¹ for boreal North America including all land areas and by 0.018 Pg·C·y⁻¹ for boreal Asia, respectively (7). However, considerable uncertainty remains regarding the magnitude of this terrestrial sink, particularly how it changes with time due to external climate drivers including the timing of spring snowmelt. It is vital for future climate scenarios to reduce the forest sink uncertainty and obtain information on the spatiotemporal variability and trends. High-latitude warming over the land surface is associated with observed reductions in spring snow extent through earlier snowmelt (2, 8–10). Advanced snowmelt across the boreal zone has a potentially major impact on the carbon balance (11–14). Here, we combine spatially continuous

time series of satellite-derived snowmelt data (clearance from the landscape, that is, the time when fractional snow cover reaches zero) with point-wise carbon flux observations to address this open question. This is performed by quantifying the relationship between the observed declines in spring snow cover extent (due to earlier snowmelt) and the carbon balance in springtime.

Results

We define the dynamics of boreal forest carbon uptake by using the change of snow clearance day (SCD) as a proxy indicator for changes in evergreen boreal forest spring recovery (SR) of photosynthesis, defined here as the timing in spring when the ecosystem CO₂ uptake exceeds 15% of its summertime maximum. Timing of SR is a major factor influencing the springtime carbon balance of high-latitude boreal forests (15–17). The advance of SR is associated with an increase in the spring carbon sink that can be quantified in terms of net ecosystem production (NEP) or gross primary production (GPP). CO₂ flux-derived NEP and GPP time series obtained for a Canadian Jack Pine forest demonstrate the connection between the advance of SR and the increase in carbon uptake (Fig. 1). Note that NEP is equivalent to GPP minus ecosystem respiration. We develop a unique approach for monitoring SR over the Earth's boreal forest zone by exploiting recent Earth Observation (EO) time series. We compare hemispheric long-term

Significance

We quantified a 36-y trend of advanced spring recovery of carbon uptake across the northern hemisphere boreal evergreen forest zone. From this trend, we estimated the corresponding change in global gross primary production (GPP) and further quantified the magnitude and spatiotemporal variability of spring GPP, that is, the cross-photosynthetic carbon uptake by forest. Our main findings are the following: (i) We developed a proxy indicator for spring recovery from in situ flux data on CO₂ exchange and recent satellite snowmelt products and (ii) we established a relation between spring recovery and carbon uptake to assess changes in springtime carbon exchange showing a major advance in the CO₂ sink.

Author contributions: J.P., M.A., T.A., and T.V. designed research; J.P., M.A., T.A., C.D., A.M., T.M., and J.S. performed research; M.T. and J.L. contributed new analyses tools; J.P., M.A., T.L., T.A., M.S., T.M., J.S., S.D., I.M., and J.-P.T. analyzed data; J.P., M.A., T.A., M.S., M.K., A.B., M.H., A. Lindroth, A. Laaksonen, C.D., A.M., T.M., and T.V. wrote the paper; and A.B., M.H., A. Lindroth, S.D., and I.M. provided flux data.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. Email: jouni.pulliainen@fmi.fi.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1707889114/-DCSupplemental.

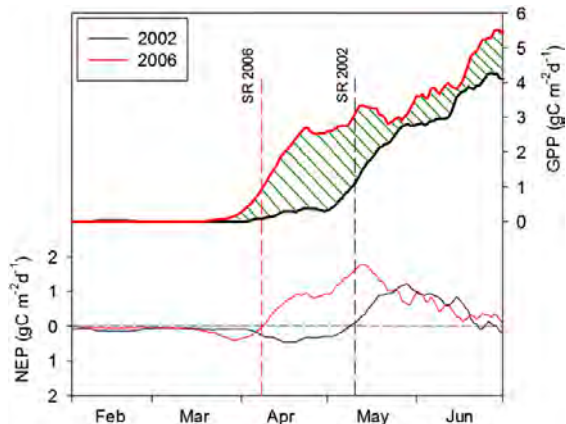


Fig. 1. Relation between SR date and carbon uptake of boreal forests in terms of GPP (thick lines, *Upper*) and NEP (thin lines, *Lower*). Daily (15-d running mean) GPP and NEP during two contrasting springs at an old-growth Jack Pine forest site in Canada demonstrate the effect of the advanced SR on the carbon uptake for a single site. Vertical dashed lines depict the flux-data-derived estimate of SR while green shading illustrates the increase in GPP during earlier spring (shaded area corresponding to change in carbon sink/uptake).

passive microwave satellite retrievals of SCD with in situ SR estimates from CO₂ flux-tower measurements. The method facilitates the use of SCD as an indicator for evergreen boreal forest spring recovery. This is necessary because no method has yet been developed to directly retrieve the recovery of photosynthesis or carbon uptake by forests from satellite observations for long periods across the northern hemisphere boreal zone. Annual spring recovery maps are generated for the 36-y microwave satellite data record. The uniqueness of this time series is its temporal precision achieved through complete passive microwave satellite coverage of boreal and high-latitude land areas every 1–2 d.

SCD information is derived solely from analysis of the space-borne passive microwave radiometer time series (10) (Fig. 2), which provides continuous coverage regardless of cloud and illumination conditions. Through the comparison of satellite retrievals with in situ data (SR derived from eddy covariance tower measurements of CO₂ fluxes in Finland, Sweden, Canada, and Russia), a linear regression model is established to describe SR as a function of SCD (Fig. 3). Based on the regression formula, the spatial patterns of SR can be mapped (Fig. 2) and temporal variability and trends in SR determined (Figs. 2 and 4 and Table 1). For comparison, SR is also analyzed using an independent landscape freeze/thaw Earth system data record (FT-ESDR) also determined from space-borne microwave radiometer data (18). Our SCD estimates and the FT-ESDR spring thaw estimates are highly correlated despite different retrieval approaches. We place added confidence in our approach because comparison with SCD derived from weather station observations indicates that the SCD dataset has a higher correlation with observed snowmelt than FT-ESDR. Note that FT-ESDR is highly correlated with the near-surface air temperature, as reanalysis-based air temperature is used for calibration of freeze/thaw retrievals (18).

We estimated the in situ SR through CO₂ net ecosystem exchange (NEE, equivalent to -NEP) measured at eddy covariance flux towers; see *Materials and Methods* for details. The 10 stations employed for regression analysis are located in Finland (4), Sweden (2), Canada (2), and Russia (2), representing conifer-dominated northern, central, and southern boreal forests. Flux observations from the stations cover different time periods between 1996 and 2014 and collectively provide 84 SR dates. The comparison of

satellite data with in situ SR (Fig. 3) shows that SCD retrievals can be used as a proxy indicator for the spatial patterns of SR for coniferous forests. The coefficient of determination (R^2) of the

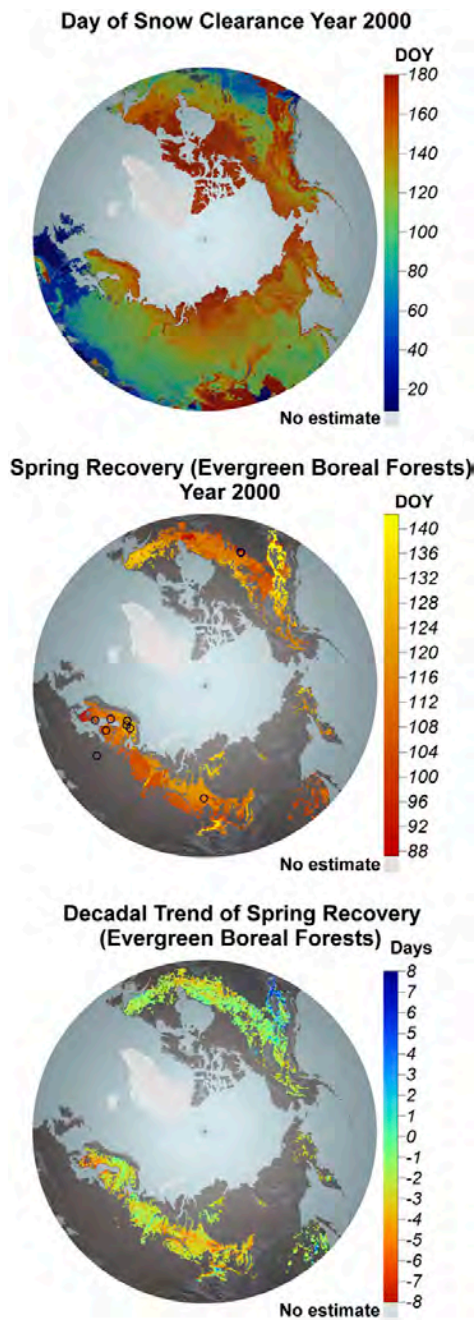


Fig. 2. SCD used as a proxy indicator for SR. (*Top*) SCD for the year 2000, retrieved from space-borne microwave radiometry. (*Middle*) The spatial behavior of SR for the year 2000 [day of year (DOY) for SR based on Eq. 1]. CO₂ flux measurement stations (10) are depicted by circles. (*Bottom*) Estimated decadal trend (days per decade) of SR determined over the period 1979–2014.

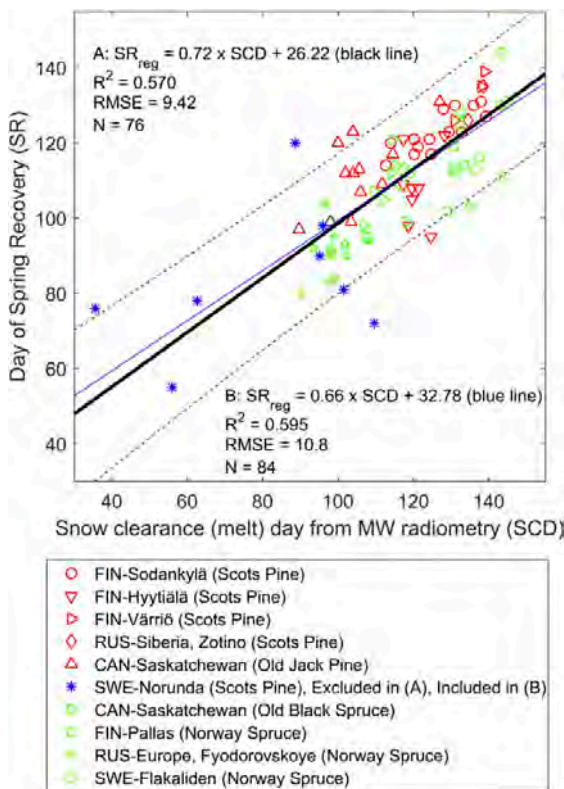


Fig. 3. Proxy for SR. Relation between CO₂ flux-tower data-derived SR and SCD from microwave radiometry for 10 conifer-dominated sites representing southern, central, and northern boreal forests. The flux stations include treated forests (harvested in the past) as well as natural forests (including forests with uniform age structure suggesting regeneration following wildfire). The regression equation between SCD and SR, used as a proxy indicator for SR, is also shown: The line of regression is shown by a black line and the 95% prediction band of the regression equation by dashed lines (case A), whereas the regression including Norunda data from the ephemeral snow region is depicted by a blue line (case B).

linear regression was 0.57 and the root mean-square error (statistical accuracy) was 9.4 d. FT-ESDR data on landscape freeze/thaw state were employed similarly to establish the regression relation between SR and the landscape thaw estimate. A slightly lower coefficient of determination was obtained ($R^2 = 0.52$). Table 1 shows the decadal change of SR for evergreen boreal forests obtained using both SCD and FT-ESDR as proxies. Both data sources provide SR estimates and trends that agree with each other within statistical error margins.

SCD can be used as a proxy for SR because of the tight coupling between snowmelt, soil thaw, and the onset of transpiration and photosynthesis (17, 19, 20). While soil thaw can be a critical factor affecting SR via the availability of water to roots, earlier work has demonstrated that photosynthesis can begin across a range of soil temperatures, following the rise of near-surface air temperature through a dynamic delay response (13, 16, 19). In regions with seasonal snow cover, the onset of snowmelt immediately precedes soil thaw that starts from the top of and bottom of the frozen layer (21, 22). The soil begins to thaw well before snow clearance is completed, triggered by the infiltration of water from the melting snowpack. Snow melt and clearance are

also related to daily air temperature and are often described in hydrology with a simple degree day model (21), similar to the dynamic delay response of photosynthesis (23). Thus, our basic

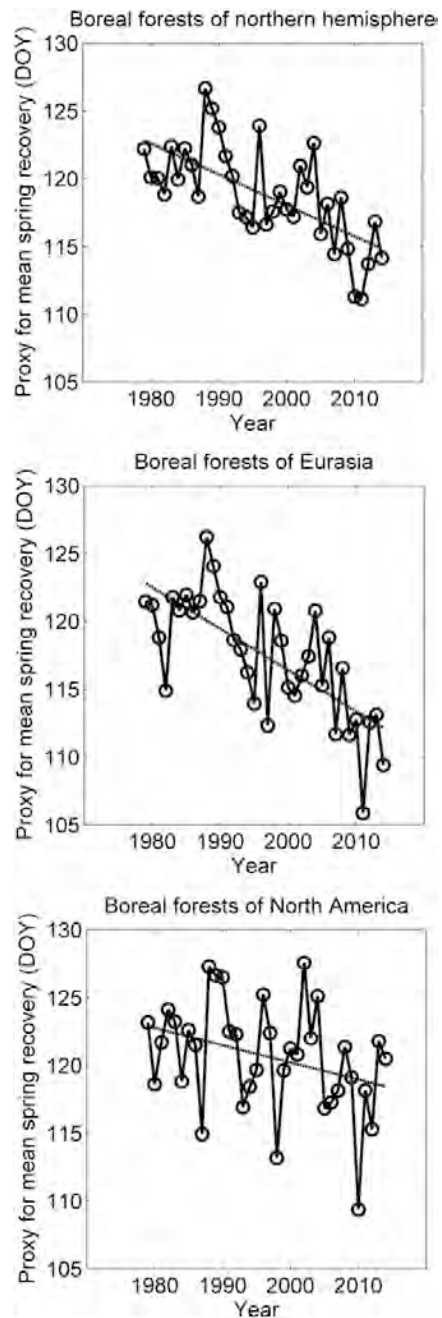


Fig. 4. SR trends in northern hemisphere. Estimated yearly time series of mean SR are shown for conifer boreal forests of northern hemisphere, Eurasia, and North America based on the regression equation of Fig. 3. Trend lines by linear regression are also shown (see Table 1 for numerical results).

Table 1. The extent of the evergreen boreal forest used in the spring recovery (SR) analysis, the change of SR based on trend lines estimated by linear regression (\pm intervals based on 95% confidence bounds from Eq. 1), *P* values of trend lines, and the corresponding decadal change of carbon sink (GPP)

Region	Forested area, 10 ⁶ km ²	Δ SR, change of spring recovery,* d/decade	<i>P</i> value for SR trend line	Δ GPP _{spring} , increase of springtime GPP [†]	
				Pg C/decade	g·C·m ⁻² /decade
Northern hemisphere	3.95	-2.28 \pm 0.46 (-1.99 \pm 0.46)	1.0·10 ⁻⁵	0.052 [‡] (5.8%)	13.1 [‡]
Eurasia	2.26	-3.04 \pm 0.61 (-2.40 \pm 0.55)	1.4·10 ⁻⁶	0.033 \pm 0.004 [‡]	8.4 \pm 1.7 [‡]
North America	1.69	-1.27 \pm 0.25 (-1.44 \pm 0.33)	0.05	0.022 \pm 0.005 [‡]	9.9 \pm 2.0 [‡]
				0.017 [‡] (4.4%)	9.8 [‡]
				0.009 \pm 0.002 [‡]	5.2 \pm 1.0 [‡]

In parentheses is the change of SR estimated from the FT-ESDR dataset, using the same regression method.

* Δ SR estimated from the satellite observations.

[†]Global model-simulated Δ GPP_{spring} (6-mo sum of daily GPP).

[‡]Increase of GPP estimated using Δ SR values of the three regions and average δ GPP_{spring}/ δ SR values determined for (i) two flux stations in Eurasia, (ii) two stations in North America, and (iii) all four stations in the hemispheric case: Δ GPP_{spring} = Δ SR \times δ GPP_{spring}/ δ SR.

assumption that the SCD derived from microwave data may be used as a proxy for boreal forest SR has a strong biophysical justification.

Our results indicate a trend toward earlier snow clearance from the landscape, with a related shift to earlier SR (Fig. 4 and Table 1). This hemispheric trend is driven almost entirely by Eurasia, where the start of the growing season is \sim 11 d earlier by the end of the time series (1979–2014). For North America the change is smaller and the interannual variability is larger. The influence of advanced SR on springtime carbon uptake was quantified by analyzing the flux measurement-based 6-mo GPP sum starting from the beginning of the year (GPP_{spring}) at four sites having a sufficient number of observation years; two forests in northern Finland [Kenttäröva (N67.98, E24.25) and Sodankylä (N67.37, E26.63) sites] and two forests in Canada [old Jack Pine (OJP) (N53.92, W104.69) and old Black Spruce (OBS) (N53.99, W105.12) sites]. When combined with the hemispheric advance of SR, the average observed sensitivity of GPP_{spring} to satellite-derived SR (δ GPP_{spring}/ δ SR) suggests an increase of 8.4 g·C·m⁻² (3.7%)/decade in GPP_{spring} for such boreal forests (Table 1).

Additionally, global climate model simulations provided the change of GPP_{spring} for all of the model grid cells covered by boreal forest. These estimates confirm that the earlier SR affects the decadal carbon sink (uptake) during spring within the whole region (Table 1). It is well known that the representation of vegetation phenology could be improved in many ecosystem models (24). However, testing a different temperature response for GPP in our model retained the sensitivity of GPP to SR, confirming that the result is robust. The correlation between early onset of SR and the level of midsummer GPP was also investigated by comparing satellite data retrievals of SR with the model-predicted July–August GPP sum. The analysis was carried out for all pixels representing evergreen boreal forests (over 11,000 pixels with a size of 625 km²). The results show a weak positive correlation between the early onset of uptake and the higher level of GPP during the midsummer, even though a small negative correlation was found for some regions (Fig. S1).

We investigated the validity of the modeling approach by comparing the modeled springtime GPP sums with satellite-derived SR (sensitivity δ GPP_{spring}/ δ SR) through Eurasia and North America and by comparing δ GPP_{spring}/ δ SR values with those observed for Canadian and Finnish flux sites (Fig. 5). Flux data analysis provided δ GPP_{spring}/ δ SR values ranging from -3.0 to -4.1 g·C·m⁻²·d⁻¹. The sensitivities obtained from the model predictions showed a mean value of -2.2 g·C·m⁻²·d⁻¹ with a SD of 1.4 g·C·m⁻²·d⁻¹ for Eurasian forests and mean of -2.3 g·C·m⁻²·d⁻¹

with a SD of 1.6 g·C·m⁻²·d⁻¹ for North American forests, respectively (Fig. 6). This indicates that the applied flux stations represent typical boreal forest in terms of δ GPP_{spring}/ δ SR.

Earlier work applying satellite data for boreal forests has primarily used channel fraction indexes such as normalized difference water index (NDWI) and normalized difference vegetation index (NDVI), determined from optical satellite data (25, 26) to investigate the start of the growing season in relation to phenological observations (e.g., timing of bud burst) or modeled phenological indexes (26, 27). Especially in the case of evergreen coniferous forests, only a small number of investigations have directly used CO₂ flux measurements as reference (28, 29). A previous case study on conifer-dominated boreal forests in Finland showed that snowmelt information derived from optical satellite data provides better estimates of SR than the normalized difference snow index (NDSI) or NDWI (29).

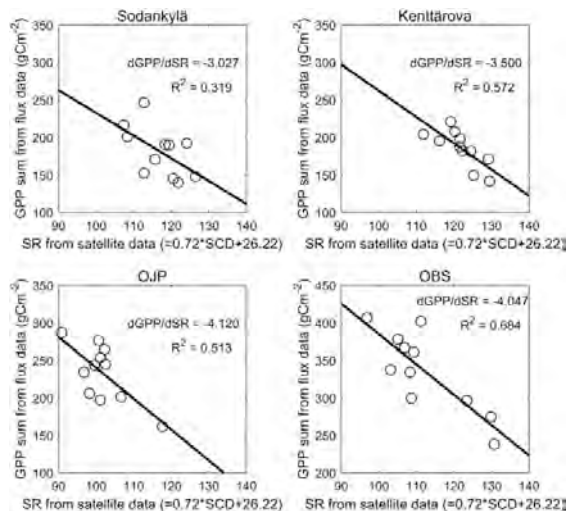


Fig. 5. Relation between satellite data (microwave radiometry)-derived SR date and carbon uptake (6-mo GPP sum, January–June) determined from CO₂ observations at flux towers. Four sites in Canada and Finland provided 11 annual observations of GPP_{spring}, enabling the estimation of sensitivity δ GPP_{spring}/ δ SR through linear regression.

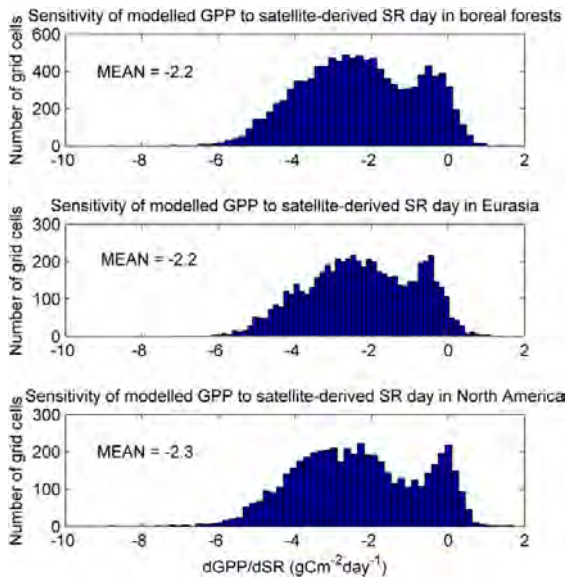


Fig. 6. Sensitivity of modeled GPP_{spring} to satellite data-derived SR for different regions. The sensitivity values are obtained by linear least-squares fitting of SR data to modeled annual springtime GPP sums (January–June) for each grid cell for 1979–2014. The slope of the regression line provides the sensitivity $\delta GPP_{spring}/\delta SR$.

Although our regression results indicate some systematic differences between the flux stations, the overall relationship between SR and SCD is strong (Fig. 3). There is no indication that the differences among stations arise from differences in tree species or region, that is, southern, central, or northern boreal forest. The SR of photosynthesis has also been found to follow environmental drivers according to a general pattern across different types of boreal coniferous forest (30). An obvious constraint of the methodology is the requirement of the presence of persistent seasonal snow cover, as demonstrated by the results for Norunda, Sweden. The Norunda data show more scatter (Fig. 3) because of its ephemeral (transitory) snow conditions. Fortunately, the proportion of boreal forest with ephemeral snow is very small. Another factor that weakens the correlation between SR and SCD is a large difference in the size of their respective footprints: ~ 25 km spatial resolution for SCD from space-borne microwave radiometer data, compared with a few hundred hectares for eddy-covariance measurements above forests.

Discussion

Our results show that passive microwave satellite-derived estimates of SCD can be combined with continuous CO_2 flux measurements to retrieve the trends of boreal forest SR (Fig. 3). The trend over 36 y is statistically significant for both Eurasia and North America and particularly strong for Eurasia (Table 1). This trend results in a significant increase in the springtime carbon uptake for Eurasia over the investigated period (Table 1). Here we affirm the important role of EO in producing spatial and temporal information on variability in the carbon cycle not available from flux-tower measurements alone. Thus, combining EO and in situ flux data is a powerful tool to move from direct geophysical retrievals (snow clearance) to high-order parameters (SR and carbon uptake). The numbers obtained here for the advancement of SR (0.23 d/y for the whole region) are consistent with the observed longer-term advance of the seasonal cycle of atmospheric CO_2 in

high latitudes, such as 0.17 d/y measured at Barrow, AK (31). An increase in the equivalent photosynthetic active period of 0.48 d/y has been estimated for the boreal zone (figure 3d in ref. 32). This estimate is for the entire year, but it is consistent with our estimates, which are for springtime (6 mo) only. Our estimates of increasing GPP in spring for the boreal forest are comparable to model predictions of annual net primary production reported elsewhere (7). This is apparent since our results suggest that there is typically a slight positive correlation between the early onset of carbon uptake and the level of July–August GPP.

The recent boreal warming trend causes earlier SR, which increases the carbon uptake during spring. This negative feedback loop reduces radiative forcing, in part counteracting the positive feedback of the earlier snowmelt (shown here) that reduces the albedo. Concerning the annual carbon cycle, earlier studies suggest that the increased soil respiration due to autumn warming may offset 90% of the increased CO_2 uptake during spring (33). The results obtained here on springtime uptake may be used to revise the trends in annual carbon balance of boreal forests.

Materials and Methods

Evergreen boreal forest SR dates from in situ CO_2 flux measurements are compared with microwave satellite retrievals of the SCD across the northern hemisphere boreal forest. The analysis focuses on 10 eddy-covariance flux sites in Eurasia and North America for 1996–2014. SR was determined from flux data based on the night-day difference in NEE. (Note that $NEE = -NEP$, where NEE is typically used for instantaneous exchange while NEP is used for longer-term balances.) Daily δNEE ($= NEE_{night} - NEE_{day}$) was obtained as a difference of 7-d running mean of nighttime [photosynthetic photon flux density (PPFD) $< 20 \mu mol \cdot m^{-2} \cdot s^{-1}$] NEE and 3-d running mean of daytime [photosynthetic photon flux density (PPFD) $> 60 \mu mol \cdot m^{-2} \cdot s^{-1}$] NEE. The summer maximum daily δNEE (δNEE_{max}) across all measurement years at each site (e.g., 2001–2010 in Sodankylä) was estimated as the 90th percentile of the daily δNEE from the 30-d period with the highest uptake on average. The SR for different years was then defined as the date when daily δNEE first exceeded 15% of site specific δNEE_{max} . The data from all stations were analyzed in the same manner, providing an unbiased dataset.

The obtained linear least-squares model between SR and SCD (Fig. 3) is

$$SR_{reg} = \hat{\beta}_1 SCD + \hat{\beta}_0, \quad [1]$$

where $\hat{\beta}_1 = 0.72 \pm 0.15$ d and $\hat{\beta}_0 = 26.2 \pm 17.4$ d (with 95% CIs). The coefficient of determination for Eq. 1 is 0.57. This equation holds for evergreen conifers in regions of seasonal snow cover. Of the 10 flux-tower sites, only Norunda (Sweden), at the southern border of the boreal forest zone, has ephemeral snow conditions and is thus excluded from the determination of Eq. 1. Nevertheless, the overall behavior of the Norunda data agrees reasonably well with Eq. 1, and the regression including Norunda data points does not differ statistically significantly from Eq. 1 and Fig. 3. The method according to Eq. 1 was also applied to FT-ESDR data by replacing SCD with the corresponding FT-ESDR landscape freeze-to-thaw transition date.

Based on long-term hemispheric satellite observations of snow cover, derived from daily passive microwave radiometer observations, we derive spatial maps of SCD for each year (Fig. 2, Top). The time series analysis algorithm indicates the timing of snow clearance for all hemispheric grid cells with seasonal snow cover (10). This snowmelt dataset has been also applied to construct the European Space Agency (ESA) GlobSnow daily snow water equivalent (SWE) and SCD climate data record (CDR) extending from 1979 to the present (34, 35). The spatial information on SR is generated by applying Eq. 1 to the SCD retrievals (Fig. 2, Middle). The boreal forest extent is extracted using a criterion that each grid cell of size 625 km^2 includes conifer evergreen forests for more than 30% of the total area. The forest mask is determined according to ESA GlobCover and ESA Climate Change Initiative (CCI) land cover information (36), with the latter used only to filter out larch-dominated regions of Siberian forests.

The reliability of trends in Fig. 4 was analyzed by adding a random noise (SD 9.42 d from Fig. 3) to every data-point time series of Fig. 4 (a Monte Carlo simulation). This resulted in a worst-case scenario assuming that the confidence of the regression algorithm of Fig. 3 is limited by the interannual variability (i.e., fluctuations arise from the year-to-year variability in the relation between the SR and SCD). This worst-case scenario suggests that there is a likelihood $P > 0.93$ that the trend is negative for the boreal forests of the northern hemisphere, $P > 0.97$ for Eurasian forests and $P > 0.79$ for North American forests, respectively (Figs. S2 and S3).

Carbon cycle–climate model simulations were carried out using the JSBACH ecosystem model (37, 38) coupled with the ECHAM6 global circulation model (39). The simulations for past decades (1957–2014) were performed with the coupled ECHAM/JSBACH model nudged toward observed climate [ERA40 and ERA Interim data (40, 41)], sea surface temperature, and atmospheric CO₂ concentration data. Springtime GPP was estimated for boreal coniferous evergreen forest in all model grid cells with significant coverage of that plant functional type, using the first 180 d of each year over the period of 1979–2014. The change of GPP for the simulation period was estimated in each grid cell by linear regression.

ACKNOWLEDGMENTS. This work is supported by the Academy of Finland CARBARC (285630); ESA GlobSnow; Academy of Finland Centre of Excellence (118780); Academy Professor projects (1284701 and 1282842); Integrated Carbon Observation System (ICOS) (271878) and ICOS-Finland (281255); ICOS-ERIC (281250); and the Nordic Centres of Excellence, DEFROST and CRAICC. Flux-tower measurements for the Russian sites were obtained as part of the TCOS-Siberia European Union (EU) Project (EU Project NEVK2-2001-00143) and supported by the Max-Planck Institute for Biogeochemistry. M.H. was supported in part by the KrasFlux Russian Science Fund project (Grant 14-24-00113).

1. Flanner MG, Shell KM, Barlage M, Perovich DK, Tschudi MA (2011) Radiative forcing and albedo feedback from the northern hemisphere cryosphere between 1979 and 2008. *Nat Geosci* 4:151–155.
2. Brown RD, Mote PW (2009) The response of northern hemisphere snow cover to a changing climate. *J Clim* 22:2124–2145.
3. Liu Yi Y, et al. (2015) Recent reversal in loss of global terrestrial biomass. *Nat Clim Chang* 5:470–474.
4. Pan Y, et al. (2011) A large and persistent carbon sink in the world's forests. *Science* 333:988–993.
5. Grogan P, Jonasson S (2006) Ecosystem CO₂ production during winter in a Swedish subarctic region: The relative importance of climate and vegetation type. *Glob Chang Biol* 12:1479–1495.
6. Ciais P, et al. (2013) *Climate Change 2013: The Physical Science Basis*, eds Stocker TF, et al. (Cambridge Univ Press, Cambridge, UK), pp 465–570.
7. Sitch S, et al. (2015) Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences* 12:653–679.
8. Derksen C, Brown R (2012) Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections. *Geophys Res Lett* 39:L19504.
9. Choi G, Robinson DA, Kang S (2010) Changing northern hemisphere snow seasons. *J Clim* 23:5305–5310.
10. Takala M, Pulliainen J, Metsämäki S, Koskinen J (2009) Detection of snow melt using spaceborne microwave radiometer data in Eurasia from 1979–2007. *IEEE Trans Geosci Remote Sens* 47:2996–3007.
11. Aurela M, Laurila T, Tuovinen J-P (2004) The timing of snow melt controls the annual CO₂ balance in a subarctic fen. *Geophys Res Lett* 31:L16119.
12. Black TA, et al. (2000) Increased carbon sequestration by a boreal deciduous forest in years with a warm spring. *Geophys Res Lett* 27:1271–1274.
13. Tanja S, et al. (2003) Air temperature triggers the recovery of evergreen boreal forest photosynthesis in spring. *Glob Chang Biol* 9:1410–1426.
14. Schwartz MD, Ahas R, Aasa A (2006) Onset of spring starting earlier across the northern hemisphere. *Glob Chang Biol* 12:343–351.
15. Suni T, et al. (2003) Interannual variability and timing of growing-season CO₂ exchange in a boreal forest. *J Geophys Res* 108:4265.
16. Thum T, et al. (2009) Spring initiation and autumn cessation of boreal coniferous forest CO₂ exchange assessed by meteorological and biological variables. *Tellus B Chem Phys Meteorol* 61:701–717.
17. Barr AG, Black TA, McCaughey JH (2009) Climatic and phenological controls of the carbon and energy balances of three contrasting boreal forest ecosystems in western Canada. *Phenology of Ecosystem Processes*, ed Noormets A (Springer, New York), pp 3–34.
18. Kim Y, Kimball JS, McDonald KC, Glassy J (2011) Developing a global data record of daily landscape freeze/thaw status using satellite passive microwave remote sensing. *IEEE Trans Geosci Remote Sens* 49:949–960.
19. Mäkelä A, Hari P, Berninger F, Hänninen H, Nikinmaa E (2004) Acclimation of photosynthetic capacity in Scots pine to the annual cycle of temperature. *Tree Physiol* 24:369–376.
20. Bergh J, Linder S (1999) Effects of soil warming during spring on photosynthesis recovery in boreal Norway spruce stands. *Glob Chang Biol* 5:245–253.
21. DeWalle R, Rango A (2008) *Principles of Snow Hydrology* (Cambridge Univ Press, Cambridge, UK).
22. Rautiainen K, et al. (2014) Detection of soil freezing from L-band passive microwave observations. *Remote Sens Environ* 147:206–218.
23. Peltoniemi M, et al. (2015) A semi-empirical model of boreal forest gross primary production, evapotranspiration, and soil water—Calibration and sensitivity analysis. *Boreal Environ Res* 20:151–171.
24. Richardson AD, et al. (2012) Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American carbon program site synthesis. *Glob Chang Biol* 18:566–584.
25. Zhang L, Friedl M, Schaaf C, Strahler A (2004) Climate controls on vegetation patterns in northern mid- and high latitudes inferred from MODIS data. *Glob Chang Biol* 10:1133–1145.
26. Delbart N, Le Toan T, Kergoat L, Fedotova V (2006) Remote sensing of spring phenology in boreal regions: A free of snow-effect method using NOAA-AVHRR and SPOT-VGT data (1982–2004). *Remote Sens Environ* 101:52–62.
27. Jönsson A, Eklundh L, Hellström M, Bähring L, Jönsson P (2010) Annual changes in MODIS vegetation indices of Swedish coniferous forests in relation to snow dynamics and tree phenology. *Remote Sens Environ* 114:2719–2730.
28. Gonsamo A, Chen J, Price D, Kurz W, Wu C (2012) Land surface phenology from optical satellite measurement and CO₂ eddy covariance technique. *J Geophys Res* 117:G03032.
29. Böttcher K, et al. (2014) MODIS time-series-derived indicators for the beginning of the growing season in boreal coniferous forest—A comparison with CO₂ flux measurements and phenological observations in Finland. *Remote Sens Environ* 140:625–638.
30. Minunno F, et al. (2016) Calibration and validation of a semi-empirical flux ecosystem model for coniferous forests in the boreal region. *Ecol Modell* 341:37–52.
31. Graven HD, et al. (2013) Enhanced seasonal exchange of CO₂ by northern ecosystems since 1960. *Science* 341:1085–1089.
32. Xu L, et al. (2013) Temperature and vegetation seasonality diminishment over northern lands. *Nat Clim Chang* 3:581–586.
33. Piao S, et al. (2008) Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature* 451:49–52.
34. Takala M, et al. (2011) Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sens Environ* 115:3517–3529.
35. Pulliainen J (2006) Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. *Remote Sens Environ* 101:257–269.
36. Bontemps S, et al. (2011) GLOBCOVER: Products Description and Validation Report (European Space Agency, ESRI, Italy), Report 2.2.
37. Raddatz TJ, et al. (2007) Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? *Clim Dyn* 29:565–574.
38. Reick CH, Raddatz T, Brovkin V, Gayler V (2013) Representation of natural and anthropogenic land cover change in MPI-ESM. *J Adv Model Earth Syst* 5:459–482.
39. Stevens B, et al. (2013) Atmospheric component of the MPI-M Earth system model: ECHAM6. *J Adv Model Earth Syst* 5:146–172.
40. Uppala SM, et al. (2005) The ERA40 reanalysis. *Q J R Meteorol Soc* 131:2961–3012.
41. Dee DP, et al. (2011) The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137:553–597.

PAPER IV

Jarmo Mäkelä, Jouni Susiluoto, Tiina Markkanen, Mika Aurela, Heikki Järvinen, Ivan Mammarella, Stefan Hagemann, Tuula Aalto

Constraining ecosystem model with adaptive Metropolis algorithm using boreal forest site eddy covariance measurements

In *Nonlinear Processes in Geophysics*, 23, 447-465, 2016,
www.nonlin-processes-geophys.net/23/447/2016/

Copyright © Authors 2016. CC Attribution 3.0 License.
Reprinted with permission.



Constraining ecosystem model with adaptive Metropolis algorithm using boreal forest site eddy covariance measurements

Jarmo Mäkelä¹, Jouni Susiluoto¹, Tiina Markkanen¹, Mika Aurela¹, Heikki Järvinen², Ivan Mammarella², Stefan Hagemann³, and Tuula Aalto¹

¹Finnish Meteorological Institute, P.O. Box 503, 00101 Helsinki, Finland

²Department of Physics, P.O. Box 48, University of Helsinki, 00014 Helsinki, Finland

³Max Planck Institute for Meteorology, Bundesstraße 53, 20146 Hamburg, Germany

Correspondence to: Jarmo Mäkelä (jarmo.makela@fmi.fi)

Received: 29 February 2016 – Published in Nonlin. Processes Geophys. Discuss.: 11 April 2016

Revised: 17 October 2016 – Accepted: 21 November 2016 – Published: 9 December 2016

Abstract. We examined parameter optimisation in the JSBACH (Kaminski et al., 2013; Knorr and Kattge, 2005; Reich et al., 2013) ecosystem model, applied to two boreal forest sites (Hyytiälä and Sodankylä) in Finland. We identified and tested key parameters in soil hydrology and forest water and carbon-exchange-related formulations, and optimised them using the adaptive Metropolis (AM) algorithm for Hyytiälä with a 5-year calibration period (2000–2004) followed by a 4-year validation period (2005–2008). Sodankylä acted as an independent validation site, where optimisations were not made.

The tuning provided estimates for full distribution of possible parameters, along with information about correlation, sensitivity and identifiability. Some parameters were correlated with each other due to a phenomenological connection between carbon uptake and water stress or other connections due to the set-up of the model formulations. The latter holds especially for vegetation phenology parameters. The least identifiable parameters include phenology parameters, parameters connecting relative humidity and soil dryness, and the field capacity of the skin reservoir. These soil parameters were masked by the large contribution from vegetation transpiration.

In addition to leaf area index and the maximum carboxylation rate, the most effective parameters adjusting the gross primary production (GPP) and evapotranspiration (ET) fluxes in seasonal tuning were related to soil wilting point, drainage and moisture stress imposed on vegetation. For daily and half-hourly tunings the most important parameters were the ratio of leaf internal CO₂ concentration to exter-

nal CO₂ and the parameter connecting relative humidity and soil dryness. Effectively the seasonal tuning transferred water from soil moisture into ET, and daily and half-hourly tunings reversed this process.

The seasonal tuning improved the month-to-month development of GPP and ET, and produced the most stable estimates of water use efficiency. When compared to the seasonal tuning, the daily tuning is worse on the seasonal scale. However, daily parametrisation reproduced the observations for average diurnal cycle best, except for the GPP for Sodankylä validation period, where half-hourly tuned parameters were better. In general, the daily tuning provided the largest reduction in model–data mismatch.

The models response to drought was unaffected by our parametrisations and further studies are needed into enhancing the dry response in JSBACH.

1 Introduction

Inverse modelling of ecosystem model parameters against in situ observations is an established way to tune model parameters (e.g. Scharnagl et al., 2011). As observation sites have their own characteristics, it is necessary to make local site simulations for model evaluation and calibration as they may reveal new insight into model behaviour and guide further development. Model–data fusion has been applied for boreal forest sites by, e.g., Aalto et al. (2004) Peltoniemi et al. (2015b), Thum et al. (2007, 2008) and Wu et al. (2011).

In this study we perform site level parameter optimisation in the JSBACH model (Kaminski et al., 2013; Knorr and Kattge, 2005; Reick et al., 2013). JSBACH is the land surface component of the Earth system model of the Max Planck Institute for Meteorology (MPI-ESM), used to simulate water and carbon storages and fluxes in the soil–vegetation–atmosphere continuum. The water and carbon fluxes are coupled in the model and thus modification of parameters related to one component usually has an effect on the others as well. The optimisation process and the optimised values are also affected by the assimilation frequency and interval in minimising the model–data mismatch. This effect can be studied in numerous ways; e.g. Santaren et al. (2014) varied the length of assimilation interval while Matheny et al. (2014) focused on the diurnal error patterns.

The motivation for this study comes from results showing that CMIP5 model simulations, one of which is MPI-ESM, have systematic evapotranspiration biases over continental areas (Mueller and Seneviratne, 2014). These kinds of biases not only have significant implications for climate change projections (Boé and Terray, 2008) but also have a distinctive behaviour on a regional scale. In addition, a comparative study of the gross primary production (GPP) of Finnish forests (Peltoniemi et al., 2015a) revealed that JSBACH has an insufficient response to water limitation in Finland – it tends to overestimate GPP and evapotranspiration during dry periods. This is especially apparent in the dry year 2006, as JSBACH is unable to transfer the reduced rainfall into lower levels of GPP.

In this study we apply the JSBACH ecosystem model for Hyytiälä (Kolari et al., 2009; Suni et al., 2003) and Sodankylä (Aurela, 2005; Thum et al., 2008) sites. We identify key parameters in soil hydrology and evapotranspiration-related formulations and test their effectiveness with elementary methods. We study the effect of different timescales (seasonal, daily and half-hourly) on the assimilation process and the effect of this on the optimised parameter values. Several optimisations are performed using the adaptive Metropolis (AM) algorithm over a 5-year calibration period (2000–2004) followed by a 4-year validation period (2005–2008).

The goals of this study are to test the applicability of the AM optimisation method for JSBACH and the impact of different temporal resolutions on the optimisation process, and to improve the models response to environmental drivers, focusing on dryness.

2 Materials and methods

2.1 Measurements, sites and instrumentation

In this study we use site level data from two Finnish measurement sites: Hyytiälä (61°51′N, 24°17′E; 180 m a.s.l.) and Sodankylä (67°22′N, 26°38′E; 179 m a.s.l.). These well-established sites have long continuous measurement data sets

representing the southern and northern boreal Finnish coniferous evergreen forests. The data used in this study are available for the scientific community through various databases such as FLUXNET (re3data.org, 2016).

Hyytiälä site is a Finnish Scots pine (*Pinus sylvestris*) forest (Kolari et al., 2009), planted in 1962 after burning and mechanical soil preparation. The soil type in Hyytiälä is Haplic Podzol on glacial till and the site is of medium fertility (Kolari et al., 2009). The forest also has sparse understorey of Norway Spruce (*Picea abies*) and scattered deciduous trees. The maximum of measured all-sided leaf area index (LAI) is $6.5 \text{ m}^2 \text{ m}^{-2}$ for the Scots pine. The carbon dioxide (CO₂) and water vapour (H₂O) fluxes between vegetation and atmosphere have been measured in Hyytiälä continuously since 1997 (Vesala et al., 2005).

The Sodankylä forest, in Sodankylä at the Finnish Meteorological Institute's Arctic Research Centre, is also a Scots pine forest (*Pinus sylvestris*) with maximum measured LAI of $3.6 \text{ m}^2 \text{ m}^{-2}$ as determined from a forest inventory in early autumn (Thum et al., 2007). The forest on fluvial sandy Podzol has been naturally regenerated after forest fires with tree age ranging approximately from 50 to 100 years. The sparse ground vegetation consists of lichens (73%), mosses (12%) and ericaceous shrubs (15%). The CO₂ and H₂O flux measurements in Sodankylä have been running since 2000 (Aurela, 2005).

The CO₂ and H₂O fluxes were measured by the micrometeorological eddy covariance (EC) method, which provides a direct measurement of the mass and energy exchange between the atmosphere and the biosphere averaged on an ecosystem scale. In the EC method, the flux is obtained as the covariance of the high-frequency (10 Hz) observations of vertical wind speed and the constituent in question (Baldocchi, 2003). The CO₂ fluxes were corrected for the storage change below the measurement height to accurately estimate the net ecosystem CO₂ exchange (NEE). The GPP was derived by subtracting the modelled respiration (*R*) from the NEE observation ($\text{GPP} = \text{NEE} - R$) utilising standard flux partitioning procedures (Reichstein et al., 2005; Kolari et al., 2009). By using the same parametrisations as in the partitioning, the NEE and GPP time series were gap-filled for comparison with the model results. The daily evapotranspiration (ET) sums were calculated from H₂O flux data that were gap-filled based on the mean diurnal cycles or regressions on available radiative energy.

The EC instrumentation in Hyytiälä consisted of a Solent 1012R3 three-axis sonic anemometer (Gill Instruments Ltd., Lymington, UK) and a LI-6262 closed-path CO₂/H₂O gas analyser (Li-Cor Inc., Lincoln, NE, USA), while in Sodankylä a USA-1 (METEK GmbH, Elmshorn, Germany) anemometer and an LI-7000 (Li-Cor, Inc., Lincoln, NE, USA) closed-path gas analyser were used. The EC fluxes were calculated as half-hourly averages taking into account the required corrections. The measurement systems and the post-processing procedures have been presented in

more detail for Hyytiälä by Kolari et al. (2004) and Mammarella et al. (2009), and for Sodankylä by Aurela (2005) and Aurela et al. (2009).

The measurement error in the EC flux data may be classified into two categories: systematic errors and random errors. The main systematic errors (density fluctuations, high-frequency losses, calibration issues) are mostly corrected for as part of the post-processing of the data, and the random errors tend to dominate the uncertainty of the instantaneous fluxes. The random error is often assumed Gaussian but can be more accurately approximated by a symmetric exponential distribution (Richardson et al., 2006). It increases linearly with the magnitude of the flux, with a standard deviation typically less than 20 % of the flux (Richardson et al., 2008; Rannik et al., 2016).

2.2 The JSBACH model

JSBACH is a process-based ecosystem model and the land surface component of the MPI-ESM. We used JSBACH offline using an observational atmospheric data set to force the model. Implications of this one-way coupling with the atmosphere include lack of feedback from the surface energy balance to the atmosphere; i.e. latent and sensible heat fluxes and surface thermal radiation do not directly affect prescribed air temperature or humidity. Similarly, the feedback of surface to the vertical transfer coefficients within the atmospheric surface layer is missing, as the wind speed that drives mixing is prescribed. Furthermore, since we use site level data (a single grid point), the grid resolution does not affect the results (Tesfa et al., 2014; Singh et al., 2015). We give here a general introduction to JSBACH, whereas a more complete model description can be found in Roeckner et al. (2003).

In JSBACH the land surface is a fractional structure where the land grid cells are divided into tiles representing the most prevalent vegetation classes called plant functional types (PFTs) within each grid cell (Reick et al., 2013). The grid cell is first divided into bare soil and vegetative area which is furthermore fractionally divided into PFTs. The model was set up to effectively use only one tile, coniferous evergreen trees. Henceforth, all model and process descriptions are considered in relation to coniferous evergreen trees and no distinction between PFT-specific and general parameters are made in this study.

Coniferous evergreen trees are characterised by a set of parameters that control vegetation-related biological and physical processes accounting for the land–atmosphere interactions. We made use of expert knowledge to set these parameters for our local sites and verified that they are in line with those presented by Hagemann (2002) and Hagemann and Stacke (2015).

The seasonal development of LAI is regulated by air temperature and soil moisture with a specific maximum LAI as a limiting value. The cycle is driven by a pseudo soil tem-

perature that is a weighted running mean of air temperature. The predictions of phenology are produced by the Logistic Growth Phenology (LoGro-P) model of JSBACH.

Photosynthesis is described by the biochemical photosynthesis model (Farquhar et al., 1980). Following Kattge et al. (2009), we set the maximum carboxylation rate at 25 °C to 1.9 times the maximum electron transport rate at 25 °C.

The photosynthetic rate is resolved in two steps. First the stomatal conductance under conditions with no water stress is assumed to be controlled by photosynthetic activity (Schulze et al., 1994). Here the leaf internal CO₂ concentration is assumed to be a constant fraction of ambient concentration, which allows for an explicit resolution of the photosynthesis (Knorr, 1997). Then the impact of soil water availability is accounted for by a soil moisture-dependent multiplier that is identical for each canopy layer (Knorr, 1997).

Radiation absorption is estimated by a two-stream approximation within a three-layer canopy (Sellers, 1985). Especially in the sparse canopies, the radiation absorption is affected by clumping of the leaves, which is here taken into account according to the formulation by Knorr (1997).

2.3 The JSBACH model spin-up and runs

Before tuning the JSBACH model, some of the more slowly changing variables (e.g. LAI) need to be equilibrated in order to bring the model into a (semi-)steady state. We achieve this by running the model through a spin-up period generated by looping the measurement interval over itself. During this period the necessary variables are equilibrated and their values become acceptable for the tuning process. At the end of the spin-up a restart file is generated so that the model can be restarted from this state.

We use half-hourly measurements from 1999 to 2008 for Hyytiälä. The spin-up finishes at the end of 1999 and is followed by a calibration period (abbreviated as HC for Hyytiälä calibration) of 2000–2004 and a validation period (HV for Hyytiälä validation) of 2005–2008, including an exceptionally dry summer in 2006. The set-up for Sodankylä is similar but we use measurements from 2000 to 2008, where the spin-up finishes at the end of 2008. The model is then restarted from the start of 2000, but we only examine the Sodankylä validation period (SV) of 2005–2008. The main reason to exclude the Sodankylä calibration period is that essentially we do not calibrate (or tune) the model for Sodankylä and we do not want to appear to do so.

The meteorological data used to drive the climate were air temperature, air pressure, atmospheric CO₂ concentration, precipitation, specific humidity, short- and longwave radiation, potential shortwave radiation and wind speed.

2.4 The parameters

The JSBACH model was modified to fit our custom-built test bed so that all parameters of interest could be read from an

Table 1. Parameter descriptions with references to equations in Appendix A.

Parameter	Units	Class	Description	
Δ_{\max}	–	I	Maximum all-sided leaf area index that vegetation can reach.	Eq. (A1)
$V_{C,\max}$	\diamond	I	Farquhar model maximum carboxylation rate at 25 °C of the enzyme Rubisco (coupled with maximum electron transport rate at 25 °C with a factor of 1.9) [$\diamond = \mu\text{mol}(\text{CO}_2)\text{m}^{-2}\text{s}^{-1}$].	Eq. (A8)
veg_{\max}	–	I	Fraction of vegetative soil in a grid cell. The rest is bare soil.	–
α_q	–	II	Farquhar model efficiency for photon capture at 25 °C.	Eq. (A11)
c_b	–	II	Adjustment parameter used in stability functions for momentum and heat (Louis, 1979).	–
f_{C3}	–	II	Ratio of C3-plant internal/external CO_2 concentration.	Eq. (A9)
p_{int}	–	II	Fraction of precipitation intercepted by the canopy.	Eq. (A12)
w_{dr}	–	II	Critical fraction of field capacity above which fast drainage occurs for soil water content.	–
w_{hum}	–	II	Fraction depicting relative humidity based on soil dryness.	Eq. (A17)
w_{pwp}	–	II	Fraction of soil moisture at permanent wilting point.	Eq. (A15)
w_{skin}	m	II	Maximum water content of the skin reservoir of bare soil.	–
w_{tsp}	–	II	Fraction of soil moisture above which transpiration is not affected by soil moisture stress.	Eq. (A15)
s_{sm}^*	m	II	Depth for correction of surface temperature for snowmelt.	–
T_{alt}	°C	III	LoGro phenology: alternating temperature. Cut-off temperature used for calculating heat sum to determine the spring event and the number of chill days since the last autumn event.	Eqs. (A2), (A3)
C_{decay}^*	–	III	LoGro phenology: memory loss parameter for chill days.	Eq. (A4)
S_{min}	°C	III	LoGro phenology: minimum value of critical heat sum.	Eq. (A4)
S_{range}^*	°C	III	LoGro phenology: maximal range of critical heat sum.	Eq. (A4)
T_{ps}	°C	III	LoGro phenology: memory loss parameter for calculating pseudo soil temperature.	Eq. (A6)

* These parameters were tested but yielded no or only a minimal response to cost functions and were thus removed from the trial.

external file. We examined 15 parameters (Table 1) that are for convenience separated into three classes. The class I parameters are used differently from those of class II and III – namely, class I parameters are only tuned in the seasonal tuning (explained in detail in Sect. 3.1). Additionally, the only distinction between class II and III parameters is that the latter belong to a specific part of JSBACH called the LoGro-P – there is no difference in how these parameters are used. We also note that the only parameter (of those examined) that can vary from site to site is veg_{\max} (the vegetative fraction of a grid cell).

2.5 Parameter sampling

The parameter sampling in this study was done with the AM algorithm. The AM algorithm is an adaptive Markov chain Monte Carlo (MCMC) process described below (it is not strictly Markovian but satisfies the necessary ergodicity requirements). AM is based on the classical Metropolis algorithm, extended with the adaptation of the parameter proposal distribution. Due to the adaptive nature of AM, it does not rely on the choice of the initial proposal distribution. AM is a sampling method that produces estimates of the full distribution of possible parameter values (unlike straightforward optimisation methods), thus enabling, e.g., the study of parameter identifiability, sensitivity and (nonlinear) correlation – this information is paramount to understanding the optimisation process in contrast to merely receiving the optimised parameter values. The rigorous mathematical presentation of the AM algorithm is given in Haario et al. (2001).

The AM algorithm draws samples (sets of parameters) from the parameter space to generate probability distributions for the parameters. The consecutive draws form an MCMC chain. We used the algorithm simultaneously for several independent chains that are parallel adaptations of the algorithmic process (see e.g. Craiu et al., 2009; Solonen et al., 2012) – we take several random starting points and launch the algorithm for each of these simultaneously. The history of all chains is used for updating the proposal covariance matrix that describes how the parameters relate to one another. Our initial proposal covariance matrix had diagonal elements corresponding to 1/200 of the respective parameter's range. The first sample for each chain was chosen at random within this range. The algorithmic process can be described by a few steps:

1. Draw a new sample (x') of the parameter space from the vicinity of the current sample (x) using the current proposal covariance matrix.
2. Calculate the acceptance ratio (a) for the drawn sample. This is the value of a likelihood function (f) that is proportional to the desired probability distribution, at the drawn sample divided by the value at the current sample ($a = f(x')/f(x)$).
3. Accept the new candidate (x') with the probability a (if $a \geq 1$, we always accept). If the candidate was rejected, the current sample (x) is reused as a basis of the next draw and repeated in the chain. Update the covariance matrix and draw a new sample.

We obtain the likelihood function (f) from the cost functions (cf) described below by assuming Gaussian error statistics and setting $f = e^{-cf}$. In general to estimate the distribution of parameters of any model based on some data, we require some information about the underlying measurement and modelling errors. We treat the JSBACH model as described by the equation $y = M(x, \theta) + e$. Here y are the observations, x is the model state vector, θ are the current parameters and e is the model–data mismatch. Since we only have a robust estimate for the measurement errors and no true error statistics for the model, the full error (e) is treated as Gaussian white noise.

The cost function (Eq. 1) used in this study for seasonal tuning is based on summary statistics of GPP and evapotranspiration (ET) along with the maximum of LAI. The cost function (Eq. 1) calculates the relative error in total GPP, ET and growing season maximum of LAI against observations (these are respectively denoted as G_1 , E_1 and L_1) and sums them up. Overlined variables refer to the mean value of that variable for a given period (calibration or validation), subscripts denote observation or model results.

$$cf_1 = \underbrace{\left(\frac{\overline{GPP_{mod}} - \overline{GPP_{obs}}}{\overline{GPP_{obs}}} \right)^2}_{G_1} + \underbrace{\left(\frac{\overline{ET_{mod}} - \overline{ET_{obs}}}{\overline{ET_{obs}}} \right)^2}_{E_1} + \underbrace{\left(\frac{\max(LAI_{mod}) - \max(LAI_{obs})}{\max(LAI_{obs})} \right)^2}_{L_1} \quad (1)$$

The second cost function (Eq. 2) is a slightly modified mean squared error estimate used for daily (cf_2) and half-hourly (cf_3) tuning. With multiple variables there is always the problem of having one variable dominating over the others. Since no true errors were available, it was decided to normalise the residuals using the mean of observations in the cost function (Eq. 2). This way the resulting function is sensitive to changes in both variables – AM is used as a noise-resistant optimiser and sampling is done in the spirit of studying the identifiability and correlations of the parameters. The components are denoted as G_2 , E_2 for daily and G_3 , E_3 for half-hourly tuning.

$$cf_{2,3} = \underbrace{\frac{1}{N_{GPP}} \sum \left(\frac{GPP_{mod} - GPP_{obs}}{\overline{GPP_{obs}}} \right)^2}_{G_{2,3}} + \underbrace{\frac{1}{N_{ET}} \sum \left(\frac{ET_{mod} - ET_{obs}}{\overline{ET_{obs}}} \right)^2}_{E_{2,3}} \quad (2)$$

As noted previously, JSBACH was used uncoupled from the other components of the full MPI-ESM. This has a tendency to lead to biased results in the model runs as has been recently studied by Dalmonech et al. (2015). Especially in the high

latitudes, evapotranspiration can be unrealistic during winter since night-time is longer and temperatures low. In order to improve the credibility of our results, we masked the evapotranspiration values of the coldest periods, and only took into account those from May to September for each year in both cost functions.

2.6 Parameter analysis

The optimised parameter values are taken as the mean values of all chains in the sampling process. In the case that the parameter chains converge to a bound of an a priori prescribed range of allowed values, the maximum a posteriori (MAP) value is used instead. After tuning the model, we analysed different aspects of this process. Class I parameters are excluded from this analysis since they are used to bring the model to an “acceptable initial state”; hence, we regard them as a part of the model initialisation (our motivation is explained in Sect. 3.1).

We calculated the correlations and correlation matrices between different parameters for different tunings using the tested parameter vectors in the AM process. Then we performed a principal component analysis (PCA) on the correlation matrices to get the eigenvectors (v_i) and eigenvalues (e_i) of the least identifiable parameters in the tuning process with the given data. The PCA transforms the correlation matrix into an orthogonal form where the eigenvector related to the greatest eigenvalue is the least identifiable with the given data. We then calculate the weight ($w_i = \sqrt{\frac{e_i^2}{\sum_i e_i^2}}$) for

each component (or vector v_i ; note that the squared weights sum up to one). We also determine the most dominant parameters for each component (v_i) by similarly dividing the length of the vector towards that parameter by the length of the whole vector (weight of vector components).

The information derived with PCA could be extracted by analysing the parameters posterior probability distributions, but PCA yields a simple, straightforward method for the same purpose. The main caveat of the standard PCA method is that it is not applicable to cases with strong nonlinear correlations. Therefore, we also calculate kernel density estimates (KDE) for the parameters to show that there are no nonlinear correlations. The KDE method places a Gaussian distribution (kernels) centred at each parameter of the MCMC chain and then sums these kernels to produce an estimate for the whole distribution. The bandwidth is calculated using the Scott’s rule (Scott, 2004).

We also wanted to examine which parameters contributed the most to the change in the cost function values as we switched from one parameter set to another. This was done by calculating the change in the cost function values of the tuned parameter set and a set where one parameter has been reverted to the value the tuning started with (henceforth, the reference values are for seasonal tuning the default values

and for daily and half-hourly tuning the seasonally tuned values). We call this method “relative effectiveness”, since we want to analyse the effect of the parameters to the cost function. For each tuned set of parameter values, the relative effectiveness of a parameter is calculated as follows:

1. change one parameter from the set of tuned parameter values to a reference value and calculate the difference in the cost function for the changed set and the tuned set;
2. return the changed parameter to the tuned value and repeat for all parameters (sum up the differences);
3. the relative effectiveness for each parameter is the difference obtained from step 1 divided by the sum from step 2.

The relative effectiveness is similar to a class of methods commonly referred to as the one-at-a-time (OAT) or one-factor-at-a-time (OFAT) methods. These methods are generally used to acquire robust information about model behaviour when one parameter at a time is changed to a new and hopefully better value (e.g. Murphy et al., 2004). The main difference of our method to classical methods such as the Morris OAT (Morris, 1991) is that in such methods the change in values is (usually) random, whereas we have fixed values. Additionally, our point of view is from the optimised parameters to the original state – we have already optimised the parameters (as a group) and merely want some robust and easily comprehensible information about the effect of changes in parameter values to the cost functions. This method does not reveal information about how well the parameters constrain the cost function (e.g. we could have a highly dominating parameter that would optimise to the default value and hence the relative effectiveness would be zero), rather which parameters contribute most to the change in cost function values.

Lastly, we calculate the root mean squared error (RMSE; $\sqrt{\sum_i \frac{(o_i - m_i)^2}{n}}$), bias ($\sum_i \frac{o_i - m_i}{n}$) and the coefficient of determination ($r^2 = 1 - \frac{\sum_i (o_i - m_i)^2}{\sum_i (o_i - \bar{o}_i)^2}$) for the time series generated by the different tunings (o_i is observed and m_i is modelled).

3 Model tuning

The model was optimised for Hyytiälä with the AM algorithm using three different timescales: seasonal, daily and half-hourly tuning, which are described below. Tuning was done on a powerful laptop with an Intel Core i7-3520M processor. We removed unwanted output streams from the model and tweaked the I/O. With a single core the spin-up generation takes approximately 150 s, the run through calibration period with daily output takes 20 s and with half-hourly output 320 s. We used daily output also for the seasonal tuning.

3.1 Seasonal tuning

The fundamental motivation for the seasonal tuning is to ensure that the model reproduces the observed growing season maximum of LAI, since we have previously noticed that JSBACH underestimates LAI at the site level (even the default value of Δ_{\max} is lower than the measured maximum for Hyytiälä). The reason for this approach was to enhance the vegetation transpiration and to emphasise the model response to precipitation. We also want to ensure that the model performs adequately well in terms of seasonal cumulative GPP and ET. The seasonal tuning was done in three consecutive steps each using the cost function (Eq. 1). The procedure is as follows:

1. All three class I parameters are tuned with four independent chains each consisting of 3000 samples. This step required a 30-year spin-up for each sample separately.
2. Class II and III parameters are each separately tested with 24 evenly separated values for an extensive range and those nine parameters that did not yield a negligible difference in the maximal and minimal values in the objective function are tuned. The consequent tuning was done with eight independent chains each consisting of 10 000 samples. A single spin-up, common for all samples, used optimal parameter values from step 1 and default values for the rest of the parameters.
3. All the previously tuned 12 parameters with eight independent chains each consisting of 10 000 samples are returned. Initial proposal covariance was generated from previous step and spin-up was generated separately for each sample.

At the end of seasonal tuning, class I parameters were fixed and a single spin-up was generated to be used with daily and half-hourly tuning. This approach is computationally justifiable (as we do not have to rerun the spin-up at each iteration of the algorithm) and is also acceptable from a modelling point of view since the robust site level scaling has already been done. The vegetative fraction of a grid cell remained at its default value of 0.52 and the carboxylation rate at 25 °C was lowered to 45.0 (and the electron transport rate to 85.5).

3.2 Daily and half-hourly tuning

The difference in daily and half-hourly tuning is the time interval used in the model output and observations in the cost function (Eq. 2). For both tuning runs we first tested the response of class II and III parameters against the cost function (Eq. 2) and removed those parameters that yielded only negligible or no response (as in step 2 in “Seasonal tuning”). The rest of the parameters (12) were then tuned using eight independent chains each consisting of 10 000 samples.

It should be noted that even though the cost function (Eq. 2) formulation is the same for daily and half-hourly tun-

Table 2. The highest correlations between parameters.

Tuning	Parameters		<i>r</i>
Seasonal	f_{C3}	w_{tsp}	0.49
	T_{alt}	α_q	0.40
Daily	f_{C3}	w_{tsp}	0.52
	w_{dr}	w_{tsp}	0.52
	T_{alt}	T_{ps}	-0.48
	T_{alt}	S_{min}	0.47
Half-hourly	f_{C3}	w_{tsp}	0.68
	p_{int}	w_{skin}	-0.44

ing, the values of the cost function are not directly comparable. Half-hourly tuning uses 48 values per day, and the resulting diurnal pattern resembles the form of the normal distribution. In daily tuning we use an average of these values. In practice, the component and cost function values will be higher for half-hourly tuning.

3.3 Tuning for Sodankylä

After tuning the model for Hyytiälä we took the parameter set from seasonal tuning and re-tuned only the maximum LAI parameter (Δ_{max}) with the cost function (Eq. 1) for Sodankylä. This was done because the measured LAI for Sodankylä is approximately half of that for Hyytiälä. The other parameter values were taken from the respective Hyytiälä tuning runs and spin-ups were generated similarly to Hyytiälä spin-ups so that we could use the Sodankylä results to validate the tuning process.

4 Results and discussion

The parameters and cost function components involved in the different phases of the optimisation process need to be studied before the performance of the optimisation method can be evaluated.

As noted above, we decided to reject the unreliable wintertime ET values. This masking leaves out the start of the growing season, which reduces the reliability of some of the tuned parameters, including all the LoGro phenology model parameters (class III), which mostly affect the timing of the spring event and regulate the development of the LAI towards the peak season. However, as a result of the tuning processes, all the analysed parameters were revealed to have unimodal posterior probability distributions, with different skewness and deviations.

We analysed the correlations and effectiveness of the parameters in the seasonal, daily and half-hourly optimisations on the Hyytiälä site for the calibration period. We also analysed the contributions from the cost function components referring to ET, GPP and LAI and generated the time series

Table 3. Significant components of principal component analysis for the different tunings. The given parameters are the most dominant within the component and the ratio is how many times larger the factor related to the first parameter is when compared to that of the second. Coverage reveals how much of the component is accounted for by the given parameters (sum of the weights of given vector components).

Component	Weight	Parameters		Ratio	Coverage
Seasonal 1	0.996	w_{hum}	w_{skin}	2.1	> 99 %
Daily 1	0.717	T_{ps}	w_{skin}	1.4	> 99 %
Daily 2	0.261	w_{hum}	w_{tsp}	2.3	> 99 %
Half-hourly 1	0.530	T_{ps}	–	–	> 99 %
Half-hourly 2	0.310	w_{skin}	w_{hum}	1.7	96 %
Half-hourly 3	0.121	T_{alt}	–	–	> 99 %

and daily cycles of GPP and ET for both Hyytiälä and Sodankylä sites. For all these examinations, individual spin-ups were generated using the optimised parameter values.

The parameter correlations (Table 2) do not reveal much information, which is common for larger systems where the underlying parameter dependencies are more complex. Usually more sophisticated methods are used to analyse the parameters, but we omit these examinations here since pairwise Kernel density estimates (Fig. 1) did not reveal any new insights.

The strongest correlation was between the ratio of leaf internal CO₂ concentration to external CO₂ (f_{C3}) and fraction of soil moisture above which transpiration is unaffected by soil moisture stress (w_{tsp}) in all the tunings. This positive correlation strengthens as we increase the temporal resolution (and the complexity of the underlying cost function). This is due to the carbon assimilation that is limited not only by the amount of carbon available but also by a linear water stress factor (which takes the value of zero at the wilting point (w_{wilt}) and one at the w_{tsp}), which is checked at each time step. Most of the other parameters with high correlations are those of the LoGro phenology model, where we would expect high correlation since the parameters are intimately connected.

Approximately half of the parameters with high correlation are also the least identifiable (Table 3) with the given data and cost function. This means that the values these parameters acquire, as a result of the tuning process, are the most unreliable – it does not reflect on the parameters contribution to the cost function. The PCA merely highlights where most of the parametric unreliability lies.

The PCA analysis revealed that most of the unreliability is explained by a handful of parameters. Disregarding those of the LoGro phenology model, the two most dominantly unreliable parameters in every tuning were the fraction depicting relative humidity based on soil dryness (w_{hum}) and the maximum field capacity of the skin reservoir (w_{skin}). Both of these parameters affect the amount of water available for

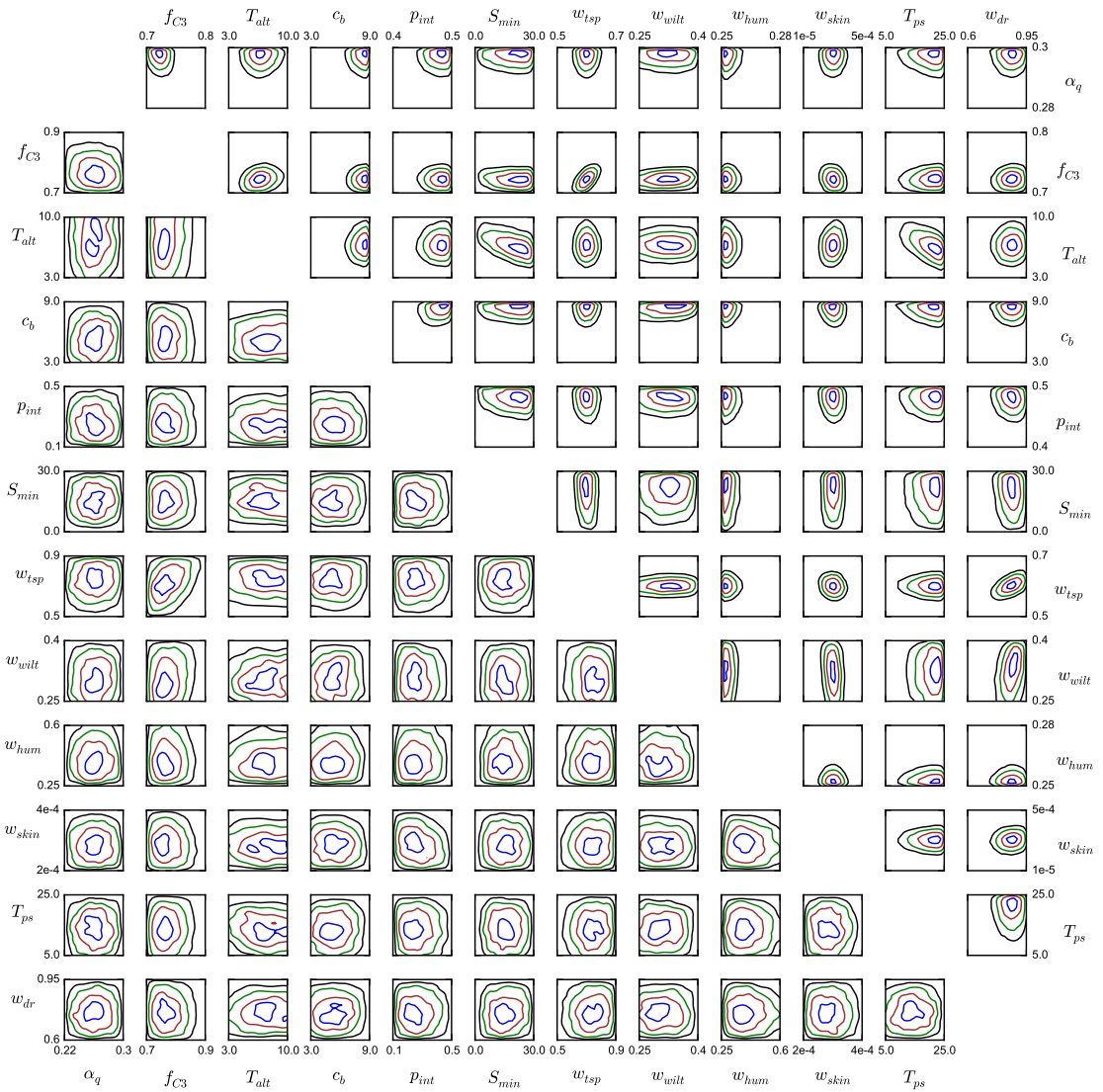


Figure 1. Kernel density estimates of the last 20 000 parameter samples with daily (upper triangle) and half-hourly tunings. The contours correspond to densities in a two-dimensional Gaussian distribution ($\mu_x, \mu_y = 0, \sigma_x, \sigma_y = 1$) with 2σ (black), 1.5σ (green), σ (brown), 0.5σ (blue).

evaporation from bare soil and are both subject to changes in other parameters. Bare soil evaporation is also dominated by vegetative transpiration, which explains why these two parameters are the most unreliable.

4.1 The parameters and their relative effectiveness

The default and optimised parameter values from the different tuning metrics are presented in Table 4 along with their relative effectiveness. The reference values for seasonal tuning are the default values. Since we fixed class I parameters with seasonal tuning, the realistic reference values for daily and half-hourly tunings are the seasonal parameter val-

Table 4. Default and optimised parameter values using the last 20 000 samples (if no value is given, the parameter was not part of that tuning, and the default value was used instead). The percentage next to a parameter value is the effectiveness of that parameter for that tuning. The reference values for seasonal tuning are the default values and for daily and half-hourly tunings the seasonal values.

Parameter	Default	Seasonal		Daily		Half-hourly	
α_q	0.28	0.26	7 %	0.30	3 %	0.27	1 %
c_b	5.0	–	–	8.8	7 %	5.0	0 %
f_{C3}	0.87	0.88	8 %	0.72	70 %	0.76	68 %
p_{int}	0.25	0.27	1 %	0.49	4 %	0.27	0 %
w_{dr}	0.9	0.79	14 %	0.87	1 %	0.75	–1 %
w_{hum}	0.5	0.54	1 %	0.25	14 %	0.37	22 %
w_{pwp}	0.35	0.28	10 %	0.34	0 %	0.31	–1 %
w_{skin} [m]	2.0×10^{-4}	3.1×10^{-4}	6 %	3.0×10^{-4}	0 %	2.2×10^{-4}	6 %
w_{tsp}	0.75	0.64	53 %	0.60	1 %	0.75	3 %
T_{alt} [°C]	4.0	8.1	0 %	6.9	1 %	6.9	2 %
S_{min} [°C]	10.0	–	–	23.0	–0 %	14.7	–0 %
T_{ps} [°C]	10.0	–	–	21.0	–0 %	12.4	–0 %

ues. Here we note that using one spin-up for all daily and half-hourly optimisation runs is computationally justifiable but generates errors as the general spin-up differs from those generated by the optimised parameters. These errors are relatively small but give rise to, e.g., the negative relative effectiveness values in daily and half-hourly parametrisations.

Most seasonally tuned parameters are near their default values and the most effective parameters are the fraction of soil moisture above which transpiration is unaffected by soil moisture stress (w_{tsp}), the fraction of soil moisture at permanent wilting point (w_{pwp}) and the fraction of field capacity above which fast drainage occurs (w_{dr}). For daily and half-hourly tunings the most important parameters are the ratio of leaf internal CO₂ concentration to external CO₂ (f_{C3}) and the fraction depicting relative humidity (w_{hum}). It should be noted that w_{hum} was one of the least identifiable parameters for seasonal tuning. Taking into account the importance of these parameters on transpiration and soil moisture estimations, we took a closer look at modelled soil moisture and evapotranspiration components for the calibration period (taking into account only values from May to September for each year as explained at the end of Sect. 2.5).

When we compare the model output streams with seasonal against those with default parametrisation, we notice that the average evapotranspiration for the calibration period has increased 15 %. Most of this is due to not only added transpiration (18 % increase) but also increased evaporation (6 %). In addition drainage was accelerated by 11 %. These increases are mostly compensated by a 15 % reduction in average soil moisture. In addition soil moisture values that are under the limit when transpiration is affected by soil moisture stress (below the value of w_{tsp}) increased 2.3 %.

The daily and half-hourly tunings lower the average evapotranspiration by 22 and 35 % respectively, when compared to the seasonal values. Transpiration is decreased by 28 and 37 %, whereas evaporation is increased by 0.5 % and de-

creased by 28 % respectively, for daily tuning and half-hourly tuning. Soil moisture is increased by 11 and 8 % and the amount of values below w_{tsp} is decreased by 62 % for daily tuning and increased by 7 % for half-hourly tuning. Out of curiosity, both the adjustment parameter in stability functions (c_b) and the fraction of precipitation intercepted by canopy (p_{int}) have been significantly increased with daily tuning and returned to seasonally tuned values with half-hourly tuning.

4.2 The cost function components

Using the optimised values (parametrisations), we calculated the components of each cost function for Hyttiälä calibration period and Hyttiälä and Sodankylä validation period (Table 5).

First, we note that with the default parameters L_1 dominates cf_1 for Hyttiälä and contributes approximately 90 % to its value. As expected the L_1 for Sodankylä is not as dominant as for Hyttiälä since the measured maximum of LAI for Hyttiälä is roughly half as large as for Sodankylä, which directly lowers the LAI component in cost function (Eq. 1). The L_1 contribution is significantly reduced with the seasonally tuned parameters as was our intention and even though LAI plays no part in daily and half-hourly tunings, the differences in the maximum value are negligible.

Second, the value of the E_1 component (error in seasonal ET) with default parametrisation is significantly increased in daily and especially half-hourly parametrisations. Simultaneously the value of G_1 is significantly lowered. The component values for seasonal parametrisation are better than the default values with the exception of E_1 for Hyttiälä validation period.

Third, for the cost function (Eq. 2) the pairwise ratio of dominating E_i or G_i components in all tunings is 5 : 1. On average E_2/E_3 contributes to approximately 60% of cf_2/cf_3 .

Table 5. Cost function components for each parametrisation for Hyytiälä calibration (HC), Hyytiälä validation (HV) and Sodankylä validation (SV) periods. L_1 , E_1 and G_1 are the LAI, ET and GPP components in cost function (Eq. 1), represented by cf_1 and used for seasonal tuning. Likewise E_2 and G_2 are the components in cost function (Eq. 2) for daily values (cf_2), whereas E_3 and G_3 are for half-hourly values (cf_3). Note that the values of cf_2 and cf_3 are not directly comparable.

		L_1	E_1	G_1	E_2	G_2	E_3	G_3	cf_1	cf_2	cf_3
HC	Default	0.396	0.021	0.036	0.306	0.191	1.126	0.681	0.45	0.50	1.8
	Seasonal	5×10^{-5}	1.7×10^{-4}	5.7×10^{-6}	0.343	0.161	1.326	0.720	2.3×10^{-4}	0.50	2.0
	Daily	7.4×10^{-5}	0.055	1.4×10^{-4}	0.206	0.149	0.906	0.683	0.06	0.36	1.6
	Half-hourly	1.0×10^{-4}	0.128	5.4×10^{-3}	0.276	0.151	0.864	0.661	0.13	0.43	1.5
HV	Default	0.396	0.002	0.028	0.226	0.157	1.027	0.479	0.43	0.38	1.5
	Seasonal	9.3×10^{-5}	0.011	7.5×10^{-4}	0.300	0.134	1.370	0.459	0.01	0.43	1.8
	Daily	1.4×10^{-4}	0.007	3.5×10^{-4}	0.164	0.124	0.981	0.446	7×10^{-3}	0.29	1.4
	Half-hourly	1.1×10^{-4}	0.058	2.9×10^{-3}	0.182	0.118	0.748	0.412	0.06	0.30	1.2
SV	Default	0.108	4.0×10^{-3}	0.140	0.423	0.596	1.660	1.795	0.25	1.02	3.5
	Seasonal	5.9×10^{-3}	1.8×10^{-5}	0.068	0.467	0.411	1.786	1.429	0.07	0.88	3.2
	Daily	6.1×10^{-3}	0.063	0.048	0.289	0.352	1.258	1.294	0.12	0.64	2.6
	Half-hourly	5.9×10^{-3}	0.164	0.022	0.379	0.290	1.246	1.185	0.19	0.67	2.4

This translates to ET being twice as significant as GPP in the cost function (Eq. 2). The main reason for ET dominating GPP is that ET is more erratic in comparison to GPP and the residuals of ET (divided by the mean value) are larger than the residuals of GPP. The daily and half-hourly tunings themselves work as intended as they lower the corresponding cost function value. It is noteworthy to mention that the G_2 component gets its lowest value for both validation periods with the half-hourly parametrisation even though G_2 calculates GPP errors on a daily scale.

Lastly, we examine how the algorithm and cost functions have performed. The best parameter set (the lowest cost function value) for a given cost function, in each of the three different periods (HC, HV, SV), is the same as that used in the corresponding tuning process. For example the lowest value for cf_1 (the cost function for seasonal tuning) in Sodankylä validation period (0.07) coincides with the seasonally tuned parameters. This is expected as the tuning process aims to be the “best” parameter value, which reassures us that no gross mistakes (human errors) have been made. The relation holds true for every cost function with the exception of cf_1 for Hyytiälä validation period, where the lowest value is reached with the daily tuned parameters (we note that the absolute difference between daily and seasonally tuned parameters is small). Hence we can confidently state that the algorithm and cost functions have performed as intended, especially since the optimised parameters work for Sodankylä as well, where no optimisation (besides the site-specific maximum of LAI) was applied.

4.3 Time series

The overall structure of the model time series was not affected by the parametrisations obtained with different

tunings (Figs. 2 and 3). Some time series characteristics have been enhanced and others reduced but the timing of the peaks and dips in GPP and ET are the same as before. The corresponding RMSE and bias estimates are given in Table 6. In comparison we estimated the PRELES model biases for Hyytiälä from Fig. 5 in Peltoniemi et al. (2015b). These estimates give a bias of $0.81 \times 10^{-6} \text{ kg m}^{-2} \text{ s}^{-1}$ ($0.07 \text{ mm m}^{-2} \text{ day}^{-1}$) for ET and $-1.45 \times 10^{-7} \text{ mol [CO}_2\text{] m}^{-2} \text{ s}^{-1}$ ($-0.15 \text{ g(C) m}^{-2} \text{ day}^{-1}$) for GPP. Additionally, the coefficient of determination (r^2) for GPP in Hyytiälä is in the range of 0.74–0.76 for all tunings, whereas the values reported in literature range from 0.68 (Trusilova et al., 2004) to 0.96 (Peltoniemi et al., 2015b) with most above 0.9 (Aalto et al., 2004; Durusma et al., 2009). For additional comparisons see also Abramowitz et al. (2007). Note that our estimates are calculated using only values from the beginning of May to the end of September.

The best seasonal performance was obtained by seasonal tuning as we previously noted from the cost function components (Table 5). Even though the optimisation is done on the seasonal level, especially the GPP cycle is noticeably improved from that generated by the default parameters. This tuning also gives rise to the most stable (least fluctuating) water use efficiency (WUE), when calculated as a pointwise ratio of GPP and ET. We use WUE here only as a diagnostic variable to examine the balance between the GPP and ET.

When compared to the seasonal tuning, the daily tuning is worse on the seasonal scale and lowers both the ET and GPP cycles. WUE follows the observations better but starts to give rise to some fluctuation. With half-hourly tuning, this behaviour is further enhanced and especially ET is lowered to too low levels, which manifests the high WUE values. The

Table 6. RMSE and bias of ET and GPP calculated from half-hourly data for first two summers of the validation period for Hyttiälä (corresponding to Fig. 2) and last two summers of the validation period for Sodankylä (corresponding to Fig. 3).

	ET (kg m ⁻² s ⁻¹)				GPP (mol (CO ₂) m ⁻² s ⁻¹)			
	Hyttiälä		Sodankylä		Hyttiälä		Sodankylä	
	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
Default	2.03 × 10 ⁻⁵	-1.31 × 10 ⁻⁶	2.27E × 10 ⁻⁵	2.31 × 10 ⁻⁶	3.09 × 10 ⁻⁶	8.77 × 10 ⁻⁷	3.16 × 10 ⁻⁶	-9.19 × 10 ⁻⁷
Seasonal	2.37 × 10 ⁻⁵	-4.32 × 10 ⁻⁶	2.35 × 10 ⁻⁵	1.09 × 10 ⁻⁶	3.10 × 10 ⁻⁶	-2.00 × 10 ⁻⁷	2.89 × 10 ⁻⁶	-5.97 × 10 ⁻⁷
Daily	2.03 × 10 ⁻⁵	-0.74 × 10 ⁻⁶	2.06 × 10 ⁻⁵	5.00 × 10 ⁻⁶	3.06 × 10 ⁻⁶	-1.07 × 10 ⁻⁷	2.74 × 10 ⁻⁶	-4.57 × 10 ⁻⁷
Half-hourly	1.69 × 10 ⁻⁵	2.77 × 10 ⁻⁶	2.04 × 10 ⁻⁵	7.14 × 10 ⁻⁶	2.94 × 10 ⁻⁶	3.39 × 10 ⁻⁷	2.67 × 10 ⁻⁶	-2.79 × 10 ⁻⁷

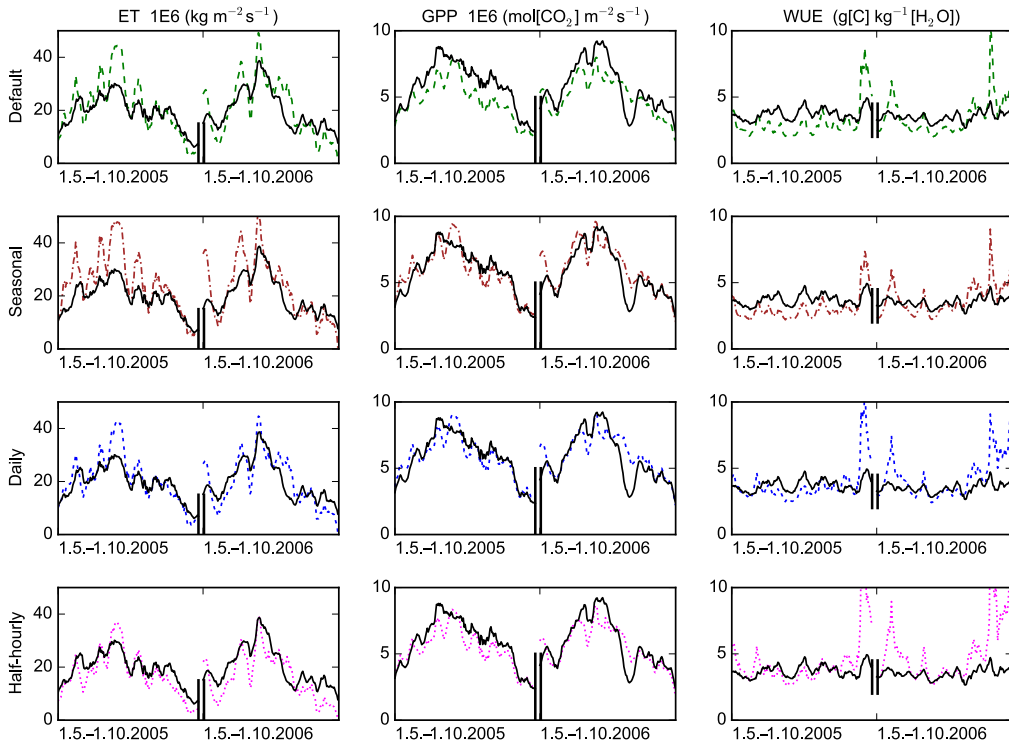


Figure 2. Hyttiälä 7-day-running mean time series for different tunings for the first two summers of the validation period. Solid black line represents the observations.

worsening in the model time series with daily and half-hourly tunings are explained by biases in the diurnal cycle.

4.4 Diurnal cycles

Average diurnal cycles with different parametrisations (Fig. 4) show that modelled night-time ET values are too low when compared to the observed and this behaviour was not affected by the tunings. Low night-time values are compensated by too high midday values in the default and seasonal tuning so that the average daily and seasonal values are on

an acceptable level. For the daily and half-hourly tuning, the algorithm lowers the daytime values, which results in too low average daily and half-hourly values. It is noteworthy to mention that with the default setting we get too low GPP for Hyttiälä but too high GPP for Sodankylä. The unrealistic wintertime and the biased night-time ET values actually have the same origin. Since we do not have the coupling from the land surface model (LSM) back to the atmosphere, we get an erroneous energy balance as we lose the energy released by condensation.

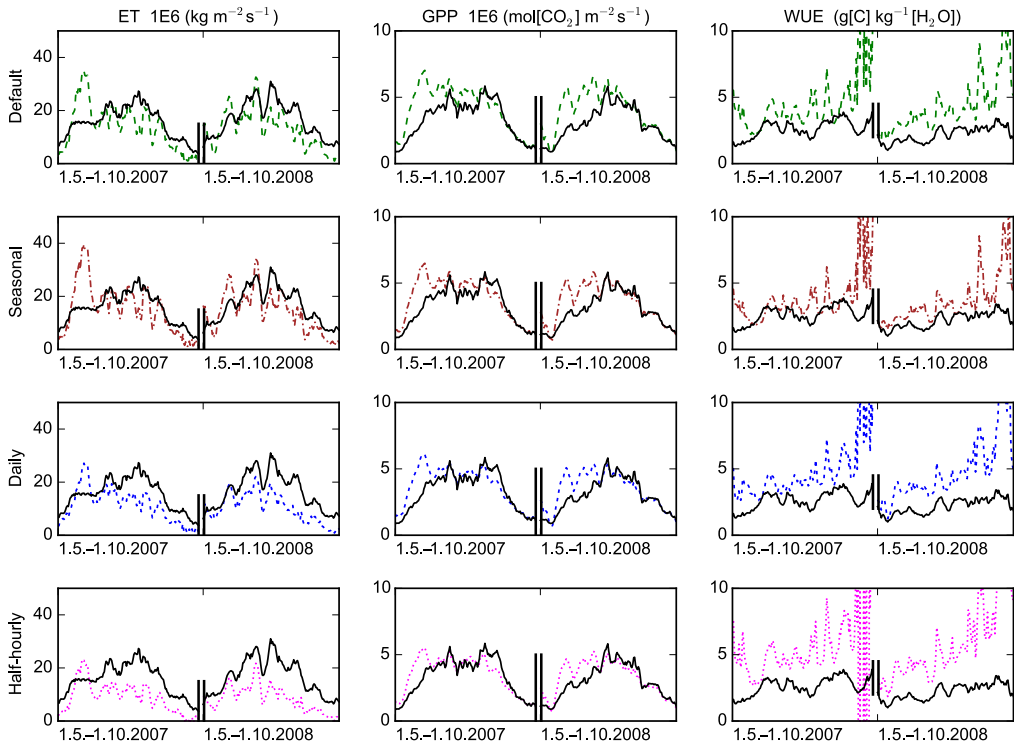


Figure 3. Sodankylä 7-day-running mean time series for different tunings for the last two summers of the validation period. Solid black line represents the observations.

Disregarding the default parametrisation we notice that seasonal parametrisation show the highest values, daily in the middle and half-hourly show the lowest values. Daily parametrisation reproduces the observations for average diurnal cycle better than the others in every occasion except the GPP for Sodankylä, where half-hourly tuning is better (verified by pointwise RMSE from the average diurnal cycle). We also notice that Sodankylä daily patterns, and to some extent Hyytiälä as well, are slightly out of phase. Our current understanding is that this is (at least partly) due to a slightly misaligned sensor (which can cause significant errors on high latitudes), measuring radiation fluxes. Fortunately this affects mainly the cost function for half-hourly tuning since it is the only one operating on the densest half-hourly timescale.

4.5 Dry event

Dry period in the summer 2006 can be clearly located by the massive drawdown in observed GPP, and to a lesser extent in ET, at Hyytiälä (Fig. 2). In a closer look at this event (Fig. 5) it is evident that none of our parametrisation schemes were able to capture it correctly. As it was with the time series,

the overall structure of the daily time series during this event remains the same (there are no divergent aspects in the model output between the different tunings).

During the drought event (defined here as 31 July–15 August 2006), the soil moisture is on average 27 % lower for default, daily and half-hourly tuning and 40 % lower for seasonal tuning when compared to the corresponding values from other years – seasonal tuning has the lowest overall soil moisture. During this event the modelled soil moisture decreases monotonically for all tunings and reaches the lowest values on 13 August, after which it starts to rise. During the period the modelled ET and GPP are predominantly higher than the observations. WUE on the other hand follows the “observations” remarkably well and deviates from the observed only towards the end of the event when modelled ET drops to near-zero values, coinciding with the lowest modelled soil moisture values. Gao et al. (2016) examined deviation in the dependencies of GPP and ET to vapour pressure deficit (VPD) between model and observation results under the most severe soil moisture stress conditions at the end of the prolonged period of soil water scarcity (that occurred in 2006). This can be attributed to the lack of ex-

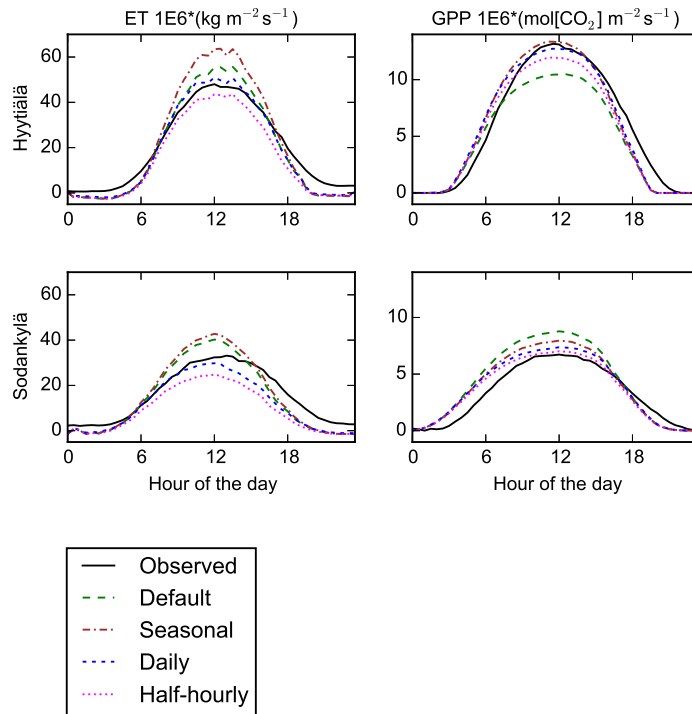


Figure 4. Average diurnal cycle from May to September for the validation period.

PLICIT dependence of the modelled stomatal conductance on the atmospheric humidity.

5 Conclusions

Initially we tuned the model to produce near-measured seasonal ET, GPP and especially maximum LAI to enhance the vegetation transpiration and to emphasise the response to precipitation. This was done successfully with seasonal tuning in the hopes of bringing forth the underlying model responses to dryness. With the consecutive daily and half-hourly tunings, we managed to improve the average diurnal cycles of both ET and GPP, but failed in reproducing the low ET and GPP levels during the dry event in 2006. Effectively we first (seasonal tuning) transferred water from soil moisture into (too high levels of) ET, and later (with daily and half-hourly tunings) transferred some of it back.

In addition to the maximum LAI (Δ_{max}) and maximum carboxylation rate ($V_{C,max}$), the most effective parameters in the seasonal tuning were the fraction of soil moisture above which transpiration is not affected by soil moisture stress (w_{tsp}) and the critical fraction of field capacity above which fast drainage occurs for soil water content (w_{dr}). The

reduction in ET and GPP was mostly accounted for by lowering the approximate ratio of leaf internal CO_2 concentration to external CO_2 (f_{C3}), which reduces the amount of carbon available for photosynthesis. For daily tuning ET was further reduced by the increase of the fraction of precipitation intercepted by canopy (p_{int}) and lower relative humidity fraction (w_{hum} – air humidity is based on soil dryness).

Despite the fact that we were unable to enhance the dry response of the model, we are confident in saying that the algorithm itself worked well and performed as intended with the daily tuning providing the most reduction in model–data mismatch. We optimised 12 parameters simultaneously (with daily and half-hourly tunings) using eight fairly short chains (8000 samples). With daily tuning the resulting estimates are well matured, but with half-hourly tuning the parameter deviations are larger (which is probably due to modelling inefficiencies and noise in measurements). Nevertheless, all optimisation procedures worked well with regard to what was optimised (seasonality, daily averages or diurnal cycle).

Recently, Knauer et al. (2015) found canopy conductance formulation to be a key factor in prescribing the transfer of carbon and water between terrestrial biosphere and the lower atmosphere. Additionally, Gao et al. (2016) found that during a prolonged period of soil water scarcity, the lack of explicit

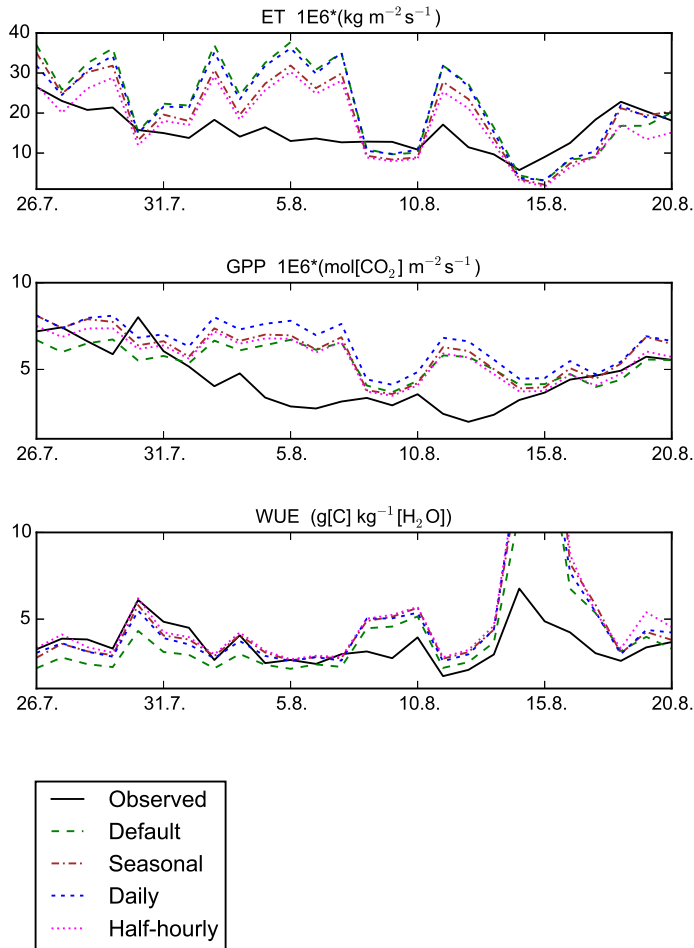


Figure 5. Daily averages for ET, GPP and WUE on a dry event in 2006 for Hyytiälä.

dependence of the stomatal conductance on the atmospheric humidity is one of the contributing factors to this issue. Further studies into enhancing the dry response in JSBACH are needed and these studies should reflect these latest findings.

6 Data availability

The measurement data required to run and tune the model can be procured from the FLUXNET database (doi:10.17616/R36K9X). The JSBACH model is available to the scientific community under a version of the Max Planck Institute for Meteorology Software License Agreement (<http://www.mpimet.mpg.de/en/science/models/license/>). For any questions regarding the simulations data, we encourage you to contact the author at jarmo.makela@fmi.fi.

Appendix A: Parametric equations within JSBACH

In this Appendix we present the main equations that the parameters in this study affect.

A1 Logistic Growth Phenology model

The parameters from the LoGro-P model, which are of interest here, are mainly used to determine the spring event for JSBACH. The maximum all-sided leaf area index (Δ_{\max}) is also part of this model; hence, we introduce this first and then deal with the spring event. Δ_{\max} is used to calculate LAI at each time step by a logistic equation (Eq. A1). Here k is the growth and p the shedding rate, both of which further depend on temperature and soil moisture.

$$\frac{d\Delta}{dt} = k\Delta \left(1 - \frac{\Delta}{\Delta_{\max}}\right) - p\Delta \quad (\text{A1})$$

To determine the date of the spring event, we first introduce a few additional variables, namely, the heat sum ($S_T(d)$), the number of chill days ($C(d)$) and the critical heat sum ($S_{\text{crit}}(d)$). Also $T(d)$ denotes the mean temperature at day d .

$$S_T(d) = \sum_{d'=d_0}^d \max(T(d') - T_{\text{alt}}, 0) \quad (\text{A2})$$

Heat sum $S_T(d)$ cumulates the amount of “heat” above the parameter T_{alt} after the previous growing season. The actual starting date d_0 of the summation need not be known since it is enough to start the summation “reasonably late” after the last growth season.

$$C(d) = \sum_{d'=d_a}^d H(T_{\text{alt}} - T(d)) \quad (\text{A3})$$

The number of chill days is calculated as the number of days when the mean temperature is below T_{alt} . Here $H()$ denotes the Heaviside step function and the summation starts at the day (d_a) of the last autumn event.

$$S_{\text{crit}}(d) = S_{\text{min}} + S_{\text{range}} e^{-C(d)/C_{\text{decay}}} \quad (\text{A4})$$

The critical heat sum (S_{crit}) decreases as the number of chill days $C(d)$ increases. The spring event happens when

$$S_T(d) \geq S_{\text{crit}}(d). \quad (\text{A5})$$

Pseudo soil temperature ($T_s(t)$) at time t is calculated as an average air temperature (T) with an exponential memory loss (T_{ps}). Pseudo soil temperature is used in determining the autumn event (when it falls below a certain threshold). In the equation N is the normalisation constant and τ is the length of a time step.

$$T_s(t) = \frac{1}{N} \sum_{n=-\infty}^t T(n) e^{-(t-n)\frac{\tau}{T_{\text{ps}}}} \quad (\text{A6})$$

A2 Photosynthesis

The Farquhar model is based on the observation that the assimilation rate in the chloroplast is limited either by the carboxylation rate (V_C) or the transport rate (J_E) of two electrons freed during the photoreaction. The total rate of carbon fixation A is given by the following equation, where R_d is the dark respiration:

$$A = \min(V_C, J_E) - R_d. \quad (\text{A7})$$

Oxygenation of the Rubisco molecule reduces the carboxylation rate, which is given as

$$V_C = V_{C,\text{max}} \frac{C_i - \Gamma_*}{C_i + K_C(1 + O_i/K_O)}. \quad (\text{A8})$$

Here C_i and O_i are the leaf internal CO_2 and O_2 concentrations, Γ_* is the CO_2 compensation point, K_C and K_O are Michaelis–Menten constants parametrising the dependence on CO_2 and O_2 concentrations. Furthermore, leaf internal CO_2 concentration depends on the external concentration C_E by

$$C_i = f_{C3} C_E. \quad (\text{A9})$$

Likewise the electron transport rate is given as

$$J_E = J(I) \frac{C_i - \Gamma_*}{4(C_i + 2\Gamma_*)}. \quad (\text{A10})$$

Here $J(I)$ is a function of radiation intensity I in the photosynthetically active band, the maximum electron transport rate J_{max} and the quantum efficiency for photon capture α_q .

$$J(I) = J_{\text{max}} \frac{\alpha_q I}{\sqrt{J_{\text{max}}^2 + \alpha_q^2 I^2}} \quad (\text{A11})$$

A3 Soil water

In JSBACH, the soil water budget is based on several reservoirs (skin, soil, bare soil, rain intercepted by canopy, etc.) and the different formulations are plentiful. We present here only the most crucial of these. Changes in soil water (w_s) due to rainfall (R), evapotranspiration (ET), snowmelt (M), surface runoff (R_s) and drainage (D) are calculated with a geographically varying maximum field capacity (w_{fc}).

$$\rho \frac{\partial w_s}{\partial t} = (1 - p_{\text{int}}) R + \text{ET} + M - R_s - D \quad (\text{A12})$$

The interception parameter (p_{int}) also affects the amount of water intercepted by vegetation and bare soil that further affects evaporation, etc. The skin reservoir is limited by w_{skin} and excess water is transferred to soil water. Likewise when the soil water content (in relation to maximum field capacity) is greater than parameter w_{dr} , the excess water is rapidly

drained (in addition to the limited drainage below this threshold).

Evaporation from wet surfaces (E_{ws}) depends on air density (ρ), specific humidity (q_a), saturation-specific humidity (q_s) at surface temperature (T_s) and pressure (p_s) and aerodynamic resistance ($r_a = C_h |v_h|^{-1}$; these are heat transfer coefficient and horizontal velocity).

$$E_{ws} = \rho \frac{q_a - q_s(T_s, p_s)}{r_a} \quad (\text{A13})$$

Transpiration from vegetation (T_v) is likewise formulated but additionally depends on the stomatal resistance of canopy (r).

$$T_v = \rho \frac{q_a - q_s(T_s, p_s)}{r_a + r} \quad (\text{A14})$$

The stomatal resistance is given as a minimal stomatal resistance of the canopy without water stress (r_{\min} , depends on photosynthetically active radiation and LAI) divided by a water stress factor (f_{ws}). That is $r = r_{\min}/f_{ws}$. The water stress factor depends on how much water is in the soil in relation to the maximum field capacity ($w_f = w_s/w_{fc}$) when compared to the limit when transpiration is no longer affected by soil moisture stress (w_{isp}) and the permanent wilting point (w_{pwp}).

$$f_{ws} = \begin{cases} 1 & w_f \geq w_{isp} \\ \frac{w_f - w_{pwp}}{w_{isp} - w_{pwp}} & w_{pwp} \leq w_f \leq w_{isp} \\ 0 & w_f \leq w_{pwp} \end{cases} \quad (\text{A15})$$

Evaporation from dry bare soil (E_s) is similarly defined as

$$E_s = \rho \frac{q_a - h q_s(T_s, p_s)}{r_a} \quad (\text{A16})$$

Here h is relative humidity at the surface relative to soil dryness:

$$h = \max \left[w_{\text{hum}} (1 - \cos(\pi w_f)), \min \left(1, \frac{q_a}{q_s(T_s, p_s)} \right) \right]. \quad (\text{A17})$$

The total evapotranspiration is a weighted average of E_{ws} , T_v and E_s , where the weights are based on, e.g., fill levels of reservoirs (similar to w_f above) and vegetative fraction of the grid cell (veg_{\max}).

Author contributions. Tuula Aalto, Heikki Järvinen, Tiina Markkanen and Stefan Hagemann chose the parameters in the optimisation process and provided support throughout the experiments. Mika Aurela and Ivan Mammarella provided knowledge on the observations. Jouni Susiluoto provided the algorithm test bed and Jarmo Mäkelä integrated the model into the test bed, ran the experiments and prepared the manuscript with contributions from all co-authors.

Acknowledgements. This work was funded by the European Commission's 7th Framework Programme, under grant agreement no. 282672, EMBRACE project, and the Nordic Centre of Excellence "Tools for Investigating Climate Change at High Northern Latitudes" (eSTICC) under the Nordic Top-Level Research Initiative. This work was also supported by the Academy of Finland Center of Excellence (no. 272041), ICOS-Finland (no. 281255) and ICOS-ERIC (no. 281250) funded by Academy of Finland. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia and USCCC. The FLUXNET eddy covariance data processing and harmonisation was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux and AsiaFlux offices.

Edited by: O. Talagrand

Reviewed by: two anonymous referees

References

- Aalto, T., Ciais, P., Chevallard, A., and Moulin, C.: Optimal determination of the parameters controlling biospheric CO₂ fluxes over Europe using eddy covariance fluxes and satellite NDVI measurements, *Tellus B*, 56, 93–104, doi:10.3402/tellusb.v56i2.16413, 2004.
- Abramowitz, G., Pitman, A., Gupta, H., Kowalczyk, E., and Wang, Y.: Systematic Bias in Land Surface Models, *J. Hydrol.*, 8, 989–1001, doi:10.1175/JHM628.1, 2007.
- Aurela, M.: Carbon dioxide exchange in subarctic ecosystems measured by a micrometeorological technique, *Finnish Meteorol. Inst. Contr.*, 51, 1–39, 2005.
- Aurela, M., Lohila, A., Tuovinen, J., Hatakka, J., Riutta, T., and Laurila, T.: Carbon dioxide exchange on a northern boreal fen, *Boreal Environ. Res.*, 14, 699–710, 2009.
- Baldocchi, D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, *Global Change Biol.*, 9, 479–492, doi:10.1046/j.1365-2486.2003.00629.x, 2003.
- Boé, J. and Terray, L.: Uncertainties in summer evapotranspiration changes over Europe and implications for regional climate change, *Geophys. Res. Lett.*, 35, L05702, doi:10.1029/2007GL032417, 2008.
- Craiu, R., Rosenthal, J., and Yang, C.: Learn From Thy Neighbor: Parallel-Chain and Regional Adaptive MCMC, *J. Am. Stat. Assoc.*, 104, 1454–1466, doi:10.1198/jasa.2009.tm08393, 2009.
- Dalmonech, D., Zaehle, S., Schürmann, G., Brovkin, V., Reick, C., and Schnur, R.: Separation of the Effects of Land and Climate Model Errors on Simulated Contemporary Land Carbon Cycle Trends in the MPI Earth System Model version 1, *J. Climate*, 28, 272–291, doi:10.1175/JCLI-D-13-00593.1, 2015.
- Duursma, R., Kolari, P., Perämäki, M., Pulkkinen, M., Mäkelä, A., Nikinmaa, E., Hari, P., Aurela, M., Berbigier, P., Bernhofer, C., Grünwald, T., Loustau, D., Mölder, M., Verbeek, H., and Vesala, T.: Contributions of climate, leaf area index and leaf physiology to variation in gross primary production of six coniferous forests across Europe: a model-based analysis, *Tree Physiol.*, 29, 621–639, doi:10.1093/treephys/tpp010, 2009.
- Farquhar, G., von Caemmerer, S., and Berry, J.: A Biochemical Model of Photosynthetic CO₂ Assimilation in Leaves of C₃ species, *Planta*, 149, 78–90, doi:10.1007/BF00386231, 1980.
- Gao, Y., Markkanen, T., Aurela, M., Mammarella, I., Thum, T., Tsuruta, A., Yang, H., and Aalto, T.: Response of water use efficiency to summer drought in boreal Scots pine forests in Finland, *Biogeosciences Discuss.*, doi:10.5194/bg-2016-198, in review, 2016.
- Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, *Bernoulli*, 7, 223–242, 2001.
- Hagemann, S.: An improved land surface parameter dataset for global and regional climate models, Max Planck Institute for Meteorology Report 336, 1–28, https://www.mpimet.mpg.de/fileadmin/publikationen/Reports/max_scirep_336.pdf (last access: 7 December 2016), 2002.
- Hagemann, S. and Stacke, T.: Impact of the soil hydrology scheme on simulated soil moisture memory, *Clim. Dynam.*, 44, 1731–1750, doi:10.1007/s00382-014-2221-6, 2015.
- Kaminski, T., Knorr, W., Schürmann, G., Scholze, M., Rayner, P., Zaehle, S., Blessing, S., Dorigo, W., Gayler, V., Giering, R., Gobron, N., Grant, J., Heimann, M., Hooker-Stroud, A., Houweling, S., Kato, T., Kattge, J., Kelley, D., Kemp, S., Koffi, E., Köstler, C., Mathieu, P.-P., Pinty, B., Reick, C., Rödenbeck, C., Schnur, R., Scipal, K., Sebal, C., Stacke, T., Terwisscha van Scheltinga, A., Vossbeck, M., Widmann, H., and Ziehn, T.: The BETHY/JSBACH Carbon Cycle Data Assimilation System: experiences and challenges, *J. Geophys. Res.-Biogeo.*, 118, 1414–1426, doi:10.1002/jgrg.20118, 2013.
- Kattge, J., Knorr, W., Raddatz, T., and Wirth, C.: Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models, *Global Change Biol.*, 15, 976–991, doi:10.1111/j.1365-2486.2008.01744.x, 2009.
- Knauer, J., Werner, C., and Zaehle, A.: Evaluating stomatal models and their atmospheric drought response in a land surface scheme: A multibiome analysis, *J. Geophys. Res.-Biogeo.*, 120, 1894–1911, doi:10.1002/2015JG003114, 2015.
- Knorr, W.: Satellite Remote Sensing and Modelling of the Global CO₂ Exchange of Land Vegetation: A Synthesis Study, Examenarbeit, Max-Planck-Institut für Meteorologie, Hamburg, 1894–1911, 1997.
- Knorr, W. and Kattge, E.: Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling, *Global Change Biol.*, 11, 1333–1351, doi:10.1111/j.1365-2486.2005.00977.x, 2005.

- Kolari, P., Pumpanen, J., Rannik, U., Ilvesniemi, H., Hari, P., and Berninger, F.: Carbon balance of different aged Scots pine forests in Southern Finland, *Global Change Biol.*, 10, 1106–1119, doi:10.1111/j.1529-8817.2003.00797.x, 2004.
- Kolari, P., Kulmala, L., Pumpanen, J., Launiainen, S., Ilvesniemi, H., Hari, P., and Nikinmaa, E.: CO₂ exchange and component CO₂ fluxes of a boreal Scots pine forest, *Boreal Environ. Res.*, 14, 761–783, 2009.
- Louis, J.-F.: A parametric model of vertical eddy fluxes in the atmosphere, *Bound.-Lay. Meteorol.*, 17, 187–202, 1979.
- Mammarella, I., Launiainen, S., Gronholm, T., Keronen, P., Pumpanen, J., Rannik, U., and Vesala, T.: Relative Humidity Effect on the High-Frequency Attenuation of Water Vapor Flux Measured by a Closed-Path Eddy Covariance System, *J. Atmos. Ocean. Tech.*, 26, 1856–1866, doi:10.1175/2009JTECHA1179.1, 2009.
- Matheny, A., Bohrer, G., Stoy, P., Baker, I., Black, A., Desai, A., Dietze, M., Gough, C., Ivanov, V., Jassal, R., Novick, K., Schäfer, K., and Verbeeck, H.: Characterizing the diurnal patterns of errors in the prediction of evapotranspiration by several land-surface models: An NACP analysis, *J. Geophys. Res.-Biogeo.*, 119, 1458–1473, doi:10.1002/2014JG002623, 2014.
- Morris, M.: Factorial Sampling Plans for Preliminary Computational Experiments, *Technometrics*, 33, 161–174, 1991.
- Mueller, B. and Seneviratne, S.: Systematic land climate and evapotranspiration biases in CMIP5 simulations, *Geophys. Res. Lett.*, 41, 128–134, doi:10.1002/2013GL058055, 2014.
- Murphy, J., Sexton, D., Barnett, D., Jones, G., Webb, M., Collins, M., and Stainforth, D.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 430, 768–772, doi:10.1038/nature02771, 2004.
- Peltoniemi, M., Markkanen, T., Härkönen, S., Muukkonen, P., Thum, T., Aalto, T., and Mäkelä, A.: Consistent estimates of gross primary production of Finnish forests – comparison of estimates of two process models, *Boreal Environ. Res.*, 20, 196–212, 2015a.
- Peltoniemi, M., Pulkkinen, M., Aurela, M., Pumpanen, J., Kolari, P., and Mäkelä, A.: A semi-empirical model of boreal-forest gross primary production, evapotranspiration, and soil water – calibration and sensitivity analysis, *Boreal Environ. Res.*, 20, 151–171, 2015b.
- Rannik, Ü., Peltola, O., and Mammarella, I.: Random uncertainties of flux measurements by the eddy covariance technique, *Atmos. Meas. Tech.*, 9, 5163–5181, doi:10.5194/amt-9-5163-2016, 2016.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biol.*, 11, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.
- Reick, C., Raddatz, T., Brovkin, V., and Gayler, V.: Representation of natural and anthropogenic land cover change in MPI-ESM, *J. Adv. Model. Earth Syst.*, 5, 1–24, doi:10.1002/jame.20022, 2013.
- re3data.org: FLUXNET; editing status 2015-05-19; re3data.org – Registry of Research Data Repositories, doi:10.17616/R36K9X, 2016.
- Richardson, A., Hollinger, D., Burba, G., Davis, K., Flanagan, L., Katul, G., Munger, J., Riccitiello, D., Stoy, P., Suyker, A., Verma, S., and Wofsy, S.: A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, *Agr. Forest Meteorol.*, 136, 1–18, doi:10.1016/j.agrformet.2006.01.007, 2006.
- Richardson, A., Mahecha, M., Falge, E., Kattge, J., Moffat, A., Papale, D., Reichstein, M., Stauch, V., Braswell, B., Churkina, G., Kruijt, B., and Hollinger, D.: Statistical properties of random CO₂ flux measurement uncertainty inferred from model residuals, *Agr. Forest Meteorol.*, 148, 38–50, doi:10.1016/j.agrformet.2007.09.001, 2008.
- Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblüeh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U., and Tompkins, A.: The atmospheric general circulation model ECHAM5. PART I: Model description, Max Planck Institute for Meteorology Report 349, 1–127, http://www.mpimet.mpg.de/fileadmin/publikationen/Reports/max_scirep_349.pdf (last access: 7 December 2016), 2003.
- Santaren, D., Peylin, P., Bacour, C., Ciais, P., and Longdoz, B.: Ecosystem model optimization using in situ flux observations: benefit of Monte Carlo versus variational schemes and analyses of the year-to-year model performances, *Biogeosciences*, 11, 7137–7158, doi:10.5194/bg-11-7137-2014, 2014.
- Scharnagl, B., Vrugt, J., Vereecken, H., and Herbst, M.: Inverse modelling of in situ soil water dynamics: investigating the effect of different prior distributions of the soil hydraulic parameters, *Hydrol. Earth Syst. Sci.*, 15, 3043–3059, doi:10.5194/hess-15-3043-2011, 2011.
- Schulze, E., Kelliher, F., Korner, C., Lloyd, J., and Leuning, R.: Relationships among Maximum Stomatal Conductance, Ecosystem Surface Conductance, Carbon Assimilation Rate, and Plant Nitrogen Nutrition: A Global Ecology Scaling Exercise, *Annu. Rev. Ecol. Syst.*, 25, 629–662, 1994.
- Scott, D. W.: Multivariate Density Estimation and Visualization, <http://EconPapers.repec.org/RePEc:zbw:casesp:200416> (last access: 7 December 2016), 2004.
- Sellers, P.: Canopy reflectance, photosynthesis and transpiration, *Int. J. Remote Sens.*, 6, 1335–1372, doi:10.1080/01431168508948283, 1985.
- Singh, R., Reager, J., Miller, N., and Famiglietti, J.: Toward hyper-resolution land-surface modeling: The effects of fine-scale topography and soil texture on CLM4.0 simulations over the Southwestern U.S., *Water Resour. Res.*, 51, 2648–2667, doi:10.1002/2014WR015686, 2015.
- Solonen, A., Ollinaho, P., Laine, M., Haario, H., Tamminen, J., and Järvinen, H.: Efficient MCMC for Climate Model Parameter Estimation: Parallel Adaptive Chains and Early Rejection, *Bayesian Anal.*, 7, 715–736, doi:10.1214/12-BA724, 2012.
- Suni, T., Rinne, J., Reissell, A., Altimir, N., Keronen, P., Rannik, U., Dal Maso, M., Kulmala, M., and Vesala, T.: Longterm measurements of surface fluxes above a Scots pine forest in Hyttiälä, southern Finland, 1996–2001, *Boreal Environ. Res.*, 8, 287–301, 2003.

- Tesfa, T., Li, H.-Y., Leung, L., Huang, M., Ke, Y., Sun, Y., and Liu, Y.: A subbasin-based framework to represent land surface processes in an Earth system model, *Geosci. Model Dev.*, 7, 947–963, doi:10.5194/gmd-7-947-2014, 2014.
- Thum, T., Aalto, T., Laurila, T., Aurela, M., Kolari, P., and Hari, P.: Parametrization of two photosynthesis models at the canopy scale in northern boreal Scots pine forest, *Tellus B*, 59, 874–890, doi:10.1111/j.1600-0889.2007.00305.x, 2007.
- Thum, T., Aalto, T., Laurila, T., Aurela, M., Lindroth, A., and Vesala, T.: Assessing seasonality of biochemical CO₂ exchange model parameters from micrometeorological flux observations at boreal coniferous forest, *Biogeosciences*, 5, 1625–1639, doi:10.5194/bg-5-1625-2008, 2008.
- Trusilova, K., Trembath, J., and Churkina, G.: Parameter Estimation and Validation of the Terrestrial Ecosystem Model BIOME-BGC Using Eddy-Covariance Flux Measurements, <http://EconPapers.repec.org/RePEc:zbw:caseps:200416> (last access: 7 December 2016), 2004.
- Vesala, T., Suni, T., Rannik, Ü., Keronen, P., Markkanen, T., Sevanto, S., Grönholm, T., Smolander, S., Kulmala, M., Ilvesniemi, H., Ojansuu, R., Uotila, A., Levula, J., Mäkelä, A., Pumpanen, J., Kolari, P., Kulmala, L., Altimir, N., Berninger, F., Nikinmaa, E., and Hari, P.: Effect of thinning on surface fluxes in a boreal forest, *Global Biogeochem. Cy.*, 19, GB2001, doi:10.1029/2004GB002316, 2005.
- Wu, S., Jansson, P., and Kolari, P.: Modeling seasonal course of carbon fluxes and evapotranspiration in response to low temperature and moisture in a boreal Scots pine ecosystem, *Ecol. Model.*, 222, 3103–3119, doi:10.1016/j.ecolmodel.2011.05.023, 2011.



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

FINNISH METEOROLOGICAL INSTITUTE

Erik Palménin aukio 1
P.O. Box 503
FI-00560 HELSINKI
tel. +358 29 539 1000
WWW.FMI.FI

FINNISH METEOROLOGICAL INSTITUTE
CONTRIBUTIONS No. 154
ISBN 978-952-336-080-8 (paperback)
ISSN 0782-6117
Edita Prima Oy 2019

ISBN 978-952-336-081-5 (pdf)
<https://doi.org/10.35614/isbn.9789523360815>
Helsinki 2019

