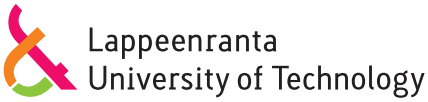


Janne Hakkarainen

ON STATE AND PARAMETER ESTIMATION IN CHAOTIC SYSTEMS

Thesis for the degree of Doctor of Philosophy to be presented with due permission for public examination and criticism in the Auditorium 2310 at Lappeenranta University of Technology, Lappeenranta, Finland on the 22nd of November, 2013, at 2 pm.

Acta Universitatis
Lappeenrantaensis **545**

Supervisors Professor Heikki Haario
Department of Mathematics and Physics
Lappeenranta University of Technology, Finland

Adjunct Professor Johanna Tamminen
Atmospheric Remote Sensing, Earth Observation
Finnish Meteorological Institute, Finland

Reviewers Associate Professor Youssef M. Marzouk
Uncertainty Quantification Group
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology, U.S.A.

Dr. J. Andrés Christen
Department of Probability and Statistics
Centro de Investigación en Matemáticas, Mexico

Opponent Professor Peter Jan van Leeuwen
Data Assimilation Research Centre
Department of Meteorology
University of Reading, United Kingdom

ISBN 978-952-265-499-1
ISBN 978-952-265-500-4 (PDF)
ISSN 1456-4491

Lappeenrannan teknillinen yliopisto
Digipaino 2013

Preface

I would like to thank my supervisors, colleagues, collaborators and academic friends. Without your contributions this would have not been possible. Thank you all. In particular, I would like to mention the following persons who have had a major impact on this dissertation: Heikki Haario, Johanna Tamminen, Antti Solonen, Marko Laine, Alexander Ilin, Heikki Järvinen, Jouni Susiluoto, Randy Moore, Curtis Wood, Peter Jones, and Erkki Kyrölä.

I would like to acknowledge the funding from the Lastu research program (Academy of Finland) and thank the whole NOVAC team for showing what scientific collaboration is at its best. I would like to thank the members of the Atmospheric Remote Sensing research group for providing a pleasant working environment at Finnish Meteorological Institute.

In addition, I would like to thank Youssef Marzouk and Andrés Christen for reviewing the manuscript and especially Peter Jan van Leeuwen, who agreed to challenge publicly the ideas presented in my dissertation.

Finally, I would like to dedicate this dissertation to my wife, Iolanda Ialongo.

Helsinki, November 2013

Janne Hakkarainen

Abstract

Janne Hakkarainen
On state and parameter estimation in chaotic systems
Lappeenranta, 2013
55 p.

Acta Universitatis Lappeenrantaensis 545
Diss. Lappeenranta University of Technology

ISBN 978-952-265-499-1
ISBN 978-952-265-500-4 (PDF)
ISSN 1456-4491

State-of-the-art predictions of atmospheric states rely on large-scale numerical models of chaotic systems. This dissertation studies numerical methods for state and parameter estimation in such systems. The motivation comes from weather and climate models and a methodological perspective is adopted. The dissertation comprises three sections: state estimation, parameter estimation and chemical data assimilation with real atmospheric satellite data.

In the state estimation part of this dissertation, a new filtering technique based on a combination of ensemble and variational Kalman filtering approaches, is presented, experimented and discussed. This new filter is developed for large-scale Kalman filtering applications.

In the parameter estimation part, three different techniques for parameter estimation in chaotic systems are considered. The methods are studied using the parameterized Lorenz 95 system, which is a benchmark model for data assimilation. In addition, a dilemma related to the uniqueness of weather and climate model closure parameters is discussed.

In the data-oriented part of this dissertation, data from the Global Ozone Monitoring by Occultation of Stars (GOMOS) satellite instrument are considered and an alternative algorithm to retrieve atmospheric parameters from the measurements is presented. The validation study presents first global comparisons between two unique satellite-borne datasets of vertical profiles of nitrogen trioxide (NO_3), retrieved using GOMOS and Stratospheric Aerosol and Gas Experiment III (SAGE III) satellite instruments. The GOMOS NO_3 observations are also considered in a chemical state estimation study in order to retrieve stratospheric temperature profiles.

The main result of this dissertation is the consideration of likelihood calculations via Kalman filtering outputs. The concept has previously been used together with stochastic differential equations and in time series analysis. In this work, the concept is applied to

chaotic dynamical systems and used together with Markov chain Monte Carlo (MCMC) methods for statistical analysis. In particular, this methodology is advocated for use in numerical weather prediction (NWP) and climate model applications. In addition, the concept is shown to be useful in estimating the filter-specific parameters related, e.g., to model error covariance matrix parameters.

Keywords: Chaotic system, Lorenz, Kalman filter, filter likelihood, data assimilation, state estimation, parameter estimation, GOMOS, SAGE III, NO_3, retrieval algorithm

UDC 51.001.57:519.246:530.18

DART	data assimilation research testbed
EAKF	ensemble adjustment Kalman filter
ECHAM	European climate model
ECMWF	European Centre for Medium-range Weather Forecasts
EKF	extended Kalman filter
EnKF	ensemble Kalman filter
FL	filter likelihood
GLIM	generalized linear modeling
GOMOS	Global Ozone Monitoring by Occultation of Stars
IRLS	iterative re-weighted least squares algorithm
LBFGS	limited memory Broyden-Fletcher-Goldfarb-Shanno
LETKF	local ensemble transform Kalman filter
MAP	maximum *a posteriori*
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
NWP	numerical weather prediction
RMSE	root mean squared error
SA	state augmentation
SAGE III	Stratospheric Aerosol and Gas Experiment III
SMW	Sherman-Morrison-Woodbury
UTLS	upper troposphere and lower stratosphere
VEnKF	variational ensemble Kalman filter
VKF	variational Kalman filter

$x \in \mathbb{R}^n$	a state vector
$y \in \mathbb{R}^p$	an observation vector
$\theta \in \mathbb{R}^{np}$	a parameter vector
C	an error covariance matrix
$Q \in \mathbb{R}^{n \times n}$	a model error covariance matrix
$R \in \mathbb{R}^{p \times p}$	an observation error covariance matrix
$\mathcal{H}$	an observation operator
$\mathcal{M}$	a state space model
$J(\theta)$	a cost function with parameters as arguments
$p(\theta \mid y)$	a conditional PDF of a random variable θ given y

Chapter I

List of original articles and the author's contribution

This thesis consist of an introductory part and five scientific articles. The articles and the author's contributions are summarized below.

I Solonen, A., Haario, H., Hakkarainen, J., Auvinen, H., Amour, I., and Kauranne, T. Variational ensemble Kalman filtering using limited memory BFGS, *Electron. Trans. Numer. Anal. 39* (2012), 271–285.

II Hakkarainen, J., Ilin, A., Solonen, A., Laine, M., Haario, H., Tamminen, J., Oja, E., and Järvinen, H. On closure parameter estimation in chaotic systems, *Nonlin. Processes Geophys. 19*, 1 (2012), 127–143, doi:10.5194/npg-19-127-2012.

III Hakkarainen, J., Solonen, A., Ilin, A., Susiluoto, J., Laine, M., Haario, H., and Järvinen, H. A dilemma of the uniqueness of weather and climate model closure parameters, *Tellus A 65* (2013), doi:10.3402/tellusa.v65i0.20147.

IV Hakkarainen, J., Tamminen, J., Moore, J. R., and Kyrölä, E. Direct comparisons of GOMOS and SAGE III NO_3 vertical profiles, *Atmos. Meas. Tech. 5*, 7 (2012), 1841–1846, doi:10.5194/amt-5-1841-2012.

V Hakkarainen, J., Laine, M., and Tamminen, J. GOMOS one-step retrieval algorithm, *Proc. SPIE 8890*, Remote Sensing of Clouds and the Atmosphere XVIII (2013), doi: 10.1117/12.2027109.

In Paper **I**, J. Hakkarainen participated in algorithm development, code implementation and helped in writing the paper.

J. Hakkarainen is the principal author of Paper **II**. He participated in code implementation, computer simulations, analyzing the results and with writing the paper.

J. Hakkarainen is the principal author of Paper **III**. He did the computer simulation, analyzed the results and wrote the paper together with co-authors.

J. Hakkarainen is the principal author of Paper **IV**. He conducted the analysis and wrote the paper together with co-authors.

In Paper **V**, J. Hakkarainen designed the case study, conducted the analysis using the algorithm developed by M. Laine and wrote the paper together with co-authors.

In addition, J. Hakkarainen contributed to the paper [51] referred to in this work, but not included as a part of this dissertation.

Chapter II

Introduction

Chaotic systems have been of academic interest since the 1880s, when Henri Poincaré discovered the phenomenon of chaos while studying the three-body problem related to planetary movement [1]. Since then, chaotic behavior has been found in various forms in nature as well as in man-made systems [25]. Probably the most well-known example is the problem related to numerical weather prediction (NWP). In the 1960s, Edward Lorenz found the so-called butterfly effect while making weather simulations using computers. He discovered that a simulated weather system is extremely sensitive to initial values, i.e., two simulations from very similar initial values result in completely different forecast trajectories. To emphasize this extreme sensitivity, in the title of his 1972 conference presentation [1], he asked

> Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?

Although there is no universally recognized definition for chaotic systems, three properties are typically required [9]. A continuous model from a metric space to itself is said to be chaotic [13], if

1. it is sensitive to initial conditions,
2. it is transitive, and
3. its periodic orbits are dense.

While the first property is the central idea of chaos, properties 2 and 3 imply the first one [9]. The phenomenon related to the sensitivity of the initial conditions is illustrated in Fig. 2.1, where several forecasts have been launched near the optimal initial values. After some days the predictability is lost.

In common language, terms like "randomness" are often associated with chaos. Chaotic systems however are fully deterministic described by the mathematical models. Even

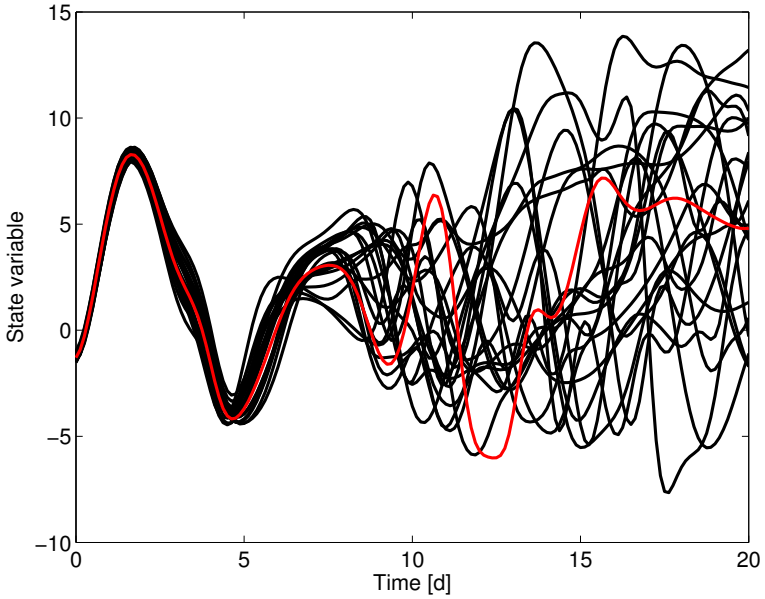


Figure 2.1: An illustration of sensitivity to initial conditions in a chaotic system: Lorenz 95 simulations with perturbed initial conditions. The red trajectory indicates the run from the optimal initial values. Predictability in the Lorenz 95 system [32], like numerical weather prediction, is generally lost after one week.

if the chaotic systems are sensitive to the initial conditions, when two simulations are launched from identical initial values then they produce precisely the same forecast trajectory. Hence, chaotic system are not to be confused with stochastic systems where random effects do take place.

Mathematical models describing dynamical chaotic systems consist of *state* and *parameters*. States are dynamically changing conditions in the physical space and parameters are static variables controlling the system. The main challenge in chaotic systems is the problem related to the predictability, i.e., long-range forecasts are impossible to make. This predictability problem also makes model validation and parameter estimation challenging; after the predictability is lost, direct model–observation comparisons become meaningless.

The aim of this dissertation is to study state and parameter estimation techniques in chaotic systems. The motivation comes from large-scale models like weather and climate models; however, only low-order tests are performed and a methodological perspective is adopted.

State estimation is at the center of numerical weather prediction (NWP) problems, and it is considered in numerous engineering applications, too. In NWP, state estimation techniques are used to obtain optimal initial values for the next weather prediction and, in this way, the model is kept on-track and the effect of chaos is alleviated. In addition, state estimation is used in re-analysis studies to find the best combination of noisy data

and the model to obtain the optimal state of the atmosphere. NWP is considered to be more an initial value problem than a parameter estimation problem [6]. Quite the opposite is true for climate models that are designed for long-term climate simulations, where the outcome is determined much more by the details of its parameterizations rather than the initial state.

General circulation models, like climate and weather models, contain atmospheric processes which are not affordable for explicit modeling. These sub-grid-scale processes are modeled with physical parameterization schemes. These schemes contain the so-called closure parameters. Traditionally, the optimal values of these closure parameters are set by the best expertise knowledge and only a relatively small number of model simulations is used for model validation. This process is somewhat subjective and therefore open to criticism. In this dissertation, numerical algorithms are presented for this kind of parameter estimation.

A relatively recent application area of state estimation methodology is chemical data assimilation, pioneered in the 1990s [29]. Chemical data assimilation can be used, e.g., to assimilate long-lived species—where the transport timescale is much smaller than chemistry timescales—like ozone to improve the quality of the wind forecasts in NWP systems or in stand-alone chemistry transport applications for ozone monitoring. Chemical state estimation can also be used to assimilate short-lived species—where chemistry timescales are much smaller than the transport timescale—as done in this dissertation.

In this work, a novel method for state estimation, variational ensemble Kalman filter (VEnKF), is presented in Paper **I**. The technique combines the ideas of the variational Kalman filtering [8] and ensemble Kalman filtering [15] approaches. This assimilation algorithm is aimed at large-scale Kalman filtering applications. The method, unlike many other ensemble-based methods, allows the direct use of the model error covariance matrix. In the numerical experiments of Paper **I**, the method showed improved performance in comparison with the traditional ensemble Kalman filtering approach, especially when the number of ensembles was small.

In Paper **II**, three numerical techniques for parameter estimation in chaotic systems are presented, experimented and discussed. The most promising method is based on the likelihood computations via Kalman filtering outputs. The method itself is not new; previously, it has been used together with stochastic differential equations and in time series analysis [49, 14]. In this work, it is advocated for use in NWP and climate model applications together with Markov chain Monte Carlo (MCMC) methods. In addition, this methodology can be used to estimate filter-specific parameters that are related, e.g., to model error covariance matrix. In the case study of Paper **II**, the filter likelihood technique showed excellent agreement with the validation metrics that use the true state of the system.

In Paper **III**, a dilemma related to the uniqueness of the weather and climate model closure parameters is demonstrated. It is shown that the different state estimation components yield different optimal model parameters. The main message of Paper **III** is that the parameters are no longer only the property of the model alone, but that the whole prediction system—including, e.g., the data assimilation component—should be considered together. In particular, if the data assimilation component of the prediction system is changed, the model parameters should be re-estimated too.

In Paper **IV**, atmospheric data from Global Ozone Monitoring by Occultation of Stars (GOMOS) and Stratospheric Aerosol and Gas Experiment III (SAGE III) satellite instruments are compared. This is the first time that global comparison of satellite-borne datasets of vertical profiles of nitrogen trioxide (NO_3) are presented. In Chapter 6, GOMOS NO_3 data are used to retrieve temperature profiles in the stratosphere via a chemical state estimation system.

In Paper **V**, an alternative algorithm for retrieving atmospheric parameters from the GOMOS measurements is presented, experimented and discussed. The algorithm is aimed at targeted case studies, e.g., applications in the upper troposphere and lower stratosphere (UTLS) region.

The summary part of this thesis is organized as follows. In Chapter 3, a general mathematical background of this work, based on Bayesian calculus, is presented. In Chapter 4, standard state estimation methods are reviewed and the VEnKF algorithm is discussed. Chapter 5 discusses parameter estimation in chaotic systems. In Chapter 6, GOMOS data are presented, compared against SAGE III and used in a chemical state estimation system. Finally, Chapter 7 concludes this dissertation.

Chapter III

Mathematical background

The mathematical background of the methods and techniques presented in this dissertation are based heavily on the Bayesian calculus and hence a short introduction to the Bayesian paradigm is in order.

In the Bayesian calculus, all variables are modeled as random variables with probability distributions. This means that uncertainty in the models and physical observations are taken into account, and the estimates are considered as approximations of the true solution with given error estimates.

Let us consider a very typical case. Let $\mathcal{H}\colon \mathbb{R}^{np} \to \mathbb{R}^p$ be a model for the observations $y \in \mathbb{R}^p$

$$\mathcal{H}(x, \theta) = y + \varepsilon, \tag{3.1}$$

where $\theta \in \mathbb{R}^{np}$ is the vector of the model parameters and $x \in \mathbb{R}^n$ is the model state vector. The uncertainty in the observations and in the observation model is given by the observation error term $\varepsilon \in \mathbb{R}^p$. Here, the aim is to estimate the unknown parameters θ given a set of observations y with the known model state x. In some applications, we could estimate the model state instead of the model parameters and in that case we would write our model as $\mathcal{H}(\theta, x)$. Naturally, the role of the state and parameters in the model can be thought symmetric. This kind of model is very typical, e.g., in atmospheric remote sensing applications. In Kalman filtering, $\mathcal{H}$ is called an observation operator.

Another typical case is where the state x is given by the dynamical prediction model $\mathcal{M}_\theta\colon \mathbb{R}^n \to \mathbb{R}^n$. In this case, it is common that parameter estimation is considered given a set of observations $y_{1:K}$ at different time instances k. The observations at different time instances are often assumed to be conditionally independent of each other given the parameters θ.

Bayes' formula. In Bayesian methodology (e.g., [16]) knowledge about the unknown parameters is inferred from the *posterior* distribution:

$$p(\theta \mid y) \propto p(\theta) p(y \mid \theta), \tag{3.2}$$

which is evaluated using the *prior* $p(\theta)$ and the *likelihood* $p(y \mid \theta)$. The prior contains the information that we have about the parameters based on the accumulated information from the past. The likelihood function specifies how plausible the observed data are given the model parameter values. Therefore, defining a proper likelihood function is the central question in parameter estimation.

Prior. In Bayesian sense, prior information should be the true estimation of the unknown parameters or the state. In practical sense, prior information can be seen as a regularization that constrains the solution when there is not enough information in the data alone. In studies where only a small number of parameters are estimated, an uninformative flat prior is often used. This accounts for the fact that typically only a minimal amount of *a priori* information is wanted and the prior is not needed for constraining the solution. Another typical form of priors are the lower and upper boundaries for the parameters, e.g., the positivity constrains where the modeler *a priori* knows that the value of the estimated quantity is positive. In atmospheric remote sensing, Gaussian profiles around climatological values are often used. The GOMOS profiles presented in Chapter 6 are retrieved using smoothness priors. In state estimation studies where the whole state is estimated, model predictions from the previous state are typically used as priors.

Likelihood. The observation error is often assumed to have zero mean and be Gaussian, and the likelihood is then evaluated as

$$p(y \mid \theta) \propto \exp\left(-\frac{1}{2}(y - \mathcal{H}(x,\theta))^{\mathrm{T}} R^{-1} (y - \mathcal{H}(x,\theta)) \right), \tag{3.3}$$

where $R \in \mathbb{R}^{p \times p}$ is the observation error covariance matrix. In the case of a dynamical model and multiple conditionally independent observations $y_{1:K}$, the likelihood could be written as

$$p(y_{1:K} \mid \theta) \propto \exp\left(-\frac{1}{2} \sum_{k=1}^{K} (y_k - \mathcal{H}(\mathcal{M}(x_k,\theta)))^{\mathrm{T}} R_k^{-1} (y_k - \mathcal{H}(\mathcal{M}(x_k,\theta))) \right). \tag{3.4}$$

However, sometimes the likelihood function is not directly applicable. This is the case, e.g., in chaotic systems where the likelihood can be formed, e.g., using summary statistics of the long model simulations or Kalman filtering outputs.

Predictive distribution. In Bayesian inference, the predictive distribution of future observations y_{K+1}—given the previous observations $y_{1:K}$—can be calculated using the posterior distribution of the previous analysis assuming again that the observations are conditionally independent

$$p(y_{K+1} \mid y_{1:K}) = \int p(y_{K+1} \mid \theta) p(\theta \mid y_{1:K}) \, \mathrm{d}\theta. \tag{3.5}$$

This property is often considered in validation studies of dynamical models [30]. In Kalman filtering, the predictive distributions are used when calculating innovation statistics.

Parameter estimation. The parameter estimation problem is often formulated using a cost function with parameters as arguments:

$$J(\theta) = \frac{1}{2}(y - \mathcal{H}(x,\theta))^{\mathrm{T}} R^{-1} (y - \mathcal{H}(x,\theta)) \tag{3.6}$$

In the Gaussian case, minimizing this cost function yields the same results as maximizing the likelihood function (3.3), since the cost function is the same as the negative log-likelihood function.

If both the prior and the likelihood are considered and are assumed to be Gaussian, the cost function is written as

$$J(\theta) = \frac{1}{2}(y - \mathcal{H}(x,\theta))^{\mathrm{T}} R^{-1}(y - \mathcal{H}(x,\theta)) + \frac{1}{2}(\theta^{\mathrm{p}} - \theta)(C^{\mathrm{p}})^{-1}(\theta^{\mathrm{p}} - \theta), \tag{3.7}$$

and again, minimizing this cost function yields the same result as maximizing the posterior function (3.2). In the cost function (3.7), θ^{p} and C^{p} indicate prior mean and the covariance matrix, respectively.

If we assume that the model $\mathcal{H}$ is a linear mapping $H \in \mathbb{R}^{p \times np}$, the argument θ^{est} that minimizes the cost function (3.7) and its error covariance matrix C^{est} can be calculated using basic matrix calculus [43]

$$\theta^{\mathrm{est}} = \theta^{\mathrm{p}} + C^{\mathrm{est}} H^{\mathrm{T}} R^{-1}(y - H\theta^{\mathrm{p}}) \tag{3.8}$$

$$C^{\mathrm{est}} = ((C^{\mathrm{p}})^{-1} + H^{\mathrm{T}} R^{-1} H)^{-1}. \tag{3.9}$$

This formulation is part of the classical Kalman filtering solution.

If the model is non-linear, different optimizing techniques can be used for minimizing the cost function (3.7) so that point estimates, such as the maximum *a posteriori* (MAP) or maximum likelihood (ML) estimates, can be obtained. Optimizing algorithms often provide also the Hessian matrix of the cost function at the optimum whose inverse can be used to approximate the covariance matrix of the solution. It is important to note that even when the prior and the likelihood are Gaussians, the posterior may not be, because of the non-linear observation model.

3.1 MCMC methods

Markov chain Monte Carlo (MCMC) techniques are statistical tools for sampling from probability distributions. They are often used to characterize the parameter errors and correlations. In addition, MCMC techniques can also be used to analyze the numerical validity of non-linear inverse algorithms [54].

In the Bayesian analysis, it is often our interest to study the full posterior distribution (3.2) and not consider just the point estimates. However, the problem of calculating the integral of the normalizing constant

$$p(y) = \int p(\theta) p(y \mid \theta)\, \mathrm{d}\theta \tag{3.10}$$

rules out direct use of the posterior distribution in all but trivial cases. Additionally, if the dimension of the parameter space is high, the use of traditional Monte Carlo methods to approximate the integral (3.10) fails.

In MCMC methodology (e.g, [42]), the aim is to generate random samples $\theta_0, \theta_1, \ldots, \theta_M$ from the posterior distribution by sampling from the artificial proposal distribution that

is typically selected to be Gaussian and centered at the previous sample θ_{m-1}. In the Metropolis algorithm, the candidate point $\widehat{\theta}$ from the proposal distribution is then accepted with the probability

$$\alpha(\theta_{m-1}, \widehat{\theta}) = \min\left(1, \frac{p(\widehat{\theta} \mid y)}{p(\theta_{m-1} \mid y)}\right). \tag{3.11}$$

If the candidate point $\widehat{\theta}$ is rejected, the previous point θ_{m-1} is repeated.

Two important features can be noted from formula (3.11). First, the normalizing constant (3.10) cancels out in the acceptance probability and hence it does not need to be calculated at any stage of the MCMC procedure. Second, the acceptance probability is 1 and the next point is always accepted, if the posterior density at the candidate point is higher than at the previous point. However, the candidate point can also be accepted, if the posterior density is lower, and in this way the algorithm explores the posterior distribution without converging to a single point.

The simplest MCMC method is the Metropolis algorithm [37]:

Step 0: Select an initial point θ_0, chain length M and set $m = 1$:

Step 1: Sample a candidate point $\widehat{\theta}$ from the proposal distribution;

Step 2: Accept $\widehat{\theta}$ with probability $\alpha(\theta_{m-1}, \widehat{\theta})$;

Step 3: If $\widehat{\theta}$ is accepted set $\theta_m = \widehat{\theta}$, else set $\theta_m = \theta_{m-1}$;

Step 4: Set $m = m + 1$ and go to step 1 until m is equal to the chain length.

As a result of a MCMC simulation a chain of samples is obtained. It can be proved that the Metropolis algorithm is ergodic [38], which states that the empirical distribution calculated from the chain is a consistent estimator of the posterior distribution [30].

In this dissertation, MCMC was used for statistical analysis in Papers **III** and **IV**. The adaptive MCMC approach [53, 19, 18], where the covariance matrix of the proposal distribution is automatically tuned using the previous samples, was adopted. All MCMC studies in this thesis have been conducted using Matlab or Fortran 90 versions of the MCMC toolbox developed in [30].

Chapter IV

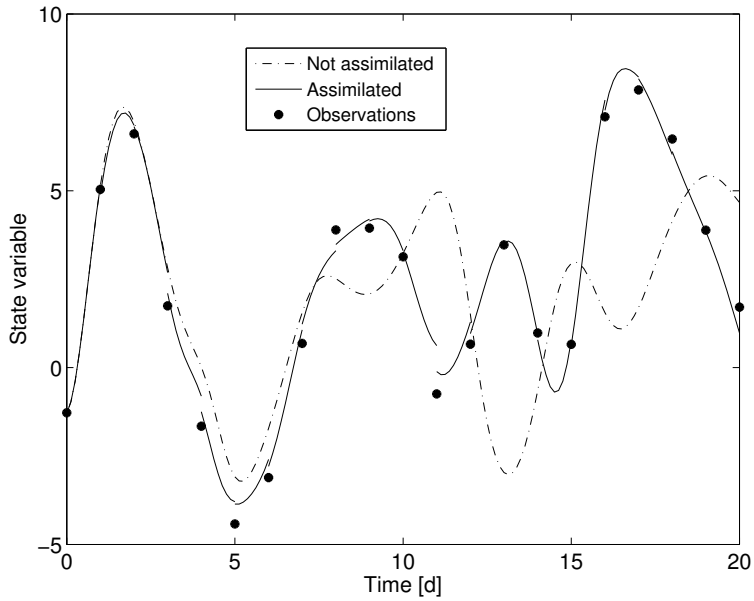
On state estimation methods

In many chaotic systems, the central problem is estimation of unknown state x_k at given time instant k. In NWP systems, for instance, state estimation (data assimilation) is used to obtain the optimal initial conditions for the next weather prediction. In addition to NWP, state estimation is also used in re-analysis studies, chemical data assimilation and in various engineering applications, such as navigation systems [2].

State estimation techniques can be divided into (i) variational methods (such as the 3D- and 4D-VAR used in many NWP applications [40]) and (ii) Kalman filtering and smoothing techniques. The main emphasis in this dissertation is on sequential filtering techniques. In this context, sequential means that the state of the system is updated sequentially as the new observations comes along. This procedure is illustrated in Fig. 4.1. It can be seen that the state estimation keeps the model on track, and problems related to chaos are alleviated. After one week, the non-assimilated control run starts to diverge from the trajectory defined by the observations.

The challenges of Kalman filtering methods in many atmospheric systems are related to the enormous size of the considered systems. In these applications, Kalman filtering quickly becomes unfeasible in its traditional form. First, the limitations in the computer memory do not allow full storage of the covariance matrices. Second, the linearization of the prediction model to be able to compute the prediction covariance matrices is, in general, not possible. In recent years, several new approaches have been developed to circumvent this problem. Many of the proposed methods are based on ensemble Kalman filtering.

In this chapter, first the general filtering equations are presented. Next, the standard Kalman filtering formulas and the ensemble Kalman filtering approach are reviewed. In Section 4.4, a novel technique from Paper **I** that is based on variational and ensemble Kalman filtering is presented. The method is designed for large-scale Kalman filtering applications.



An

Figure 4.1: An illustration of sequential state estimation in a chaotic system. Due to chaos, the non-assimilated control run diverges from the trajectory defined by the observations. See also Fig. 2.1.

4.1 General filtering formulas

Generally, the problem statement in filtering is given as a pair of state and observaraions

$$\begin{cases} x_k &= \mathcal{M}_k(x_{k-1}) + \xi_k, \\ y_k &= \mathcal{H}_k(x_k) + \varepsilon_k, \end{cases} \tag{4.1}$$

where $\mathcal{M}\colon \mathbb{R}^n \to \mathbb{R}^n$ is the dynamical state space model and $\mathcal{H}\colon \mathbb{R}^n \to \mathbb{R}^p$ is the observation operator that maps from the state space to the observation space. In addition, error terms ξ_k and ε_k are considered. With probabilistic notation, pair (4.1) can be written as

$$\begin{cases} x_k &\sim p(x_k \mid x_{k-1}), \\ y_k &\sim p(y_k \mid x_k). \end{cases} \tag{4.2}$$

In filtering, our aim is to find the posterior distribution $p(x_k \mid y_{1:K})$ of the state, given all the previous and current observations $y_{1:K}$ using Bayesian inference, as in Chapter 3. Often, a Gaussian approach is taken.

In general, filtering methods work in two stages: prediction and update. These are iterated, as illustrated in Fig. 4.1. In the prediction stage, the state estimate and its uncertainty are transported to the next time-step's prior using a dynamical model. The prior of the current state can be seen as a predictive distribution:

$$p(x_k \mid y_{1:k-1}) = \int p(x_k \mid x_{k-1}) p(x_{k-1} \mid y_{1:k-1}) \, \mathrm{d}x_{k-1}. \tag{4.3}$$

The predictive distribution of the observations y_k can be calculated as

$$p(y_k \mid y_{1:k-1}) = \int p(y_k \mid x_k) p(x_k \mid y_{1:k-1}) \, \mathrm{d}x_k. \tag{4.4}$$

In the update stage, the Bayes formula is used to obtain the posterior with the predictive distribution $p(x_k \mid y_{1:k-1})$ as prior

$$p(x_k \mid y_{1:k}) \propto p(y_{1:k} \mid x_k) p(x_k \mid y_{1:k-1}). \tag{4.5}$$

The next sections show how these general formulas work in practice.

4.2 Kalman filtering

Probably the most known and used state estimation technique is the Kalman filter and its non-linear counterpart, the extended Kalman filter (EKF) [24, 2].

In linear Kalman filtering, the state estimate of the previous time-step x_{k-1}^{est} and its error covariance matrix C_{k-1}^{est} are transported to the next time-step's prior x_k^{p} using linear model M_k

$$x_k^{\mathrm{p}} = M_k x_{k-1}^{\mathrm{est}}, \tag{4.6}$$

$$C_k^{\mathrm{p}} = M_k C_{k-1}^{\mathrm{est}} M_k^{\mathrm{T}} + Q_k, \tag{4.7}$$

where Q_k is the model error covariance matrix. The innovation (prediction residual) and its error covariance matrix can be calculated as

$$r_k = y_k - H x_k^{\mathrm{p}}, \tag{4.8}$$

$$C_k^{\mathrm{r}} = H_k C_k^{\mathrm{p}} H_k^{\mathrm{T}} + R_k, \tag{4.9}$$

where R_k is the observation error covariance matrix.

Finally, in the update phase, the following formulas are used

$$x_k^{\mathrm{est}} = x_k^{\mathrm{p}} + G_k r_k, \tag{4.10}$$

$$C_k^{\mathrm{est}} = (I - G_k H_k) C_k^{\mathrm{p}}, \tag{4.11}$$

where the Kalman gain matrix G can be calculated as

$$G_k = C_k^{\mathrm{p}} H_k^{\mathrm{T}} (C_k^{\mathrm{r}})^{-1}. \tag{4.12}$$

The updating formulas (4.10)–(4.11) are similar to the formulas (3.8)–(3.9) presented in Chapter 3.

If the dynamical model $\mathcal{M}$ or the observation operator $\mathcal{H}$ are not linear, the above formulas cannot be directly used. In this case, Kalman filtering formulas can be extended and used with the following tangent linear models

$$\begin{cases} M_k &= \dfrac{\partial \mathcal{M}_k(x_{k-1}^{\mathrm{est}})}{\partial x}, \\ H_k &= \dfrac{\partial \mathcal{H}_k(x_k^{\mathrm{p}})}{\partial x}. \end{cases} \tag{4.13}$$

These matrices are used in the Kalman filtering formulas, with the exception that the prediction (4.6) and the innovation (4.8) are computed using the actual non-linear models.

The extended Kalman filter, unlike the basic Kalman filter, is no-longer optimal, and the estimated error covariance matrix is likely to be underestimated [2]. This is because in the non-linear case the formulas (4.7) and (4.9) are not strictly valid. However, in practice, the EKF performs well in many application areas and in this dissertation, it is considered as the benchmark for other state estimation techniques.

As already mentioned, the full Kalman filtering becomes quickly unfeasible when the size of the state grows. The next sections show some alternative Kalman filtering techniques based on ensemble approaches.

4.3 Ensemble Kalman filtering

In non-linear systems, the main problem with extended Kalman filtering is the linearization (4.13) of the prediction model $\mathcal{M}$ for calculating the prior covariance matrix C_k^{p}. In ensemble Kalman filtering [15], the uncertainties are represented as a set of samples $x_k = (x_{1,k}, \ldots, x_{N,k})$ and the sample mean and covariance are considered. In this way, the linearization is avoided. The sample covariance matrix can be calculated as $\mathrm{Cov}(x_k) = X_k X_k^{\mathrm{T}}$, where

$$X_k = ((x_{1,k} - \overline{x_k}), \ldots, (x_{N,k} - \overline{x_k}))/\sqrt{N-1}, \tag{4.14}$$

and $\overline{x_k}$ denotes the sample mean. In many applications, the number of ensemble members is chosen to be smaller than the size of state n. Although, when $N < n$, the sample covariance matrix is not positive-definite.

In ensemble filtering, the uncertainty is transported to the next filtering time-step by transporting an ensemble of states using the forecast model $\mathcal{M}$. The difficulty (and variety) of the ensemble methods lies in how to update the prior ensemble in the analysis step.

In the basic ensemble Kalman filter (EnKF) algorithm [11], the update is done using Kalman filtering formulas, presented in Section 4.2, for individual ensemble members separately. In the prediction stage of the EnKF algorithm, the ensemble is pushed forward in time and perturbed according to model error covariance matrix Q_k

$$x_{i,k}^{\mathrm{p}} = \mathcal{M}_k(x_{i,k-1}^{\mathrm{est}}) + \xi_{i,k}, \tag{4.15}$$

where $\xi_{i,k} \sim N(0, Q_k)$. The matrix X_k is then formed from this perturbed prior ensemble.

The innovations are perturbed using realization of the observation error covariance matrix R_k

$$r_{i,k} = y_k - H x_{i,k}^{\mathrm{p}} + \varepsilon_{i,k}. \tag{4.16}$$

Finally, the posterior ensemble is acquired using Kalman gain (4.12)

$$x_{i,k}^{\mathrm{est}} = x_{i,k}^{\mathrm{p}} + G_k r_{i,k}. \tag{4.17}$$

In ensemble filtering, the sample mean is often considered as a state estimate. Sometimes this ensemble mean can be "non-representative" and produce "non-physical" estimates as

the ensemble mean does not provide a realistic scenario. If this is the case, it can be advantageous to consider a MAP point instead. This is done, e.g., in the VEnKF algorithm presented in the next section.

Other ensemble filters. As can be noted, the basic EnKF includes random perturbations that are essential to allow this method to work in practice and to produce correct ensemble statistics [11]. These perturbations make the filter stochastic and state estimates vary from one realization to another. An interesting subclass of ensemble filters are the so-called ensemble square root filters [56], where no random perturbation takes place and the filters are deterministic. The non-stochasticity is essential in some application like filter likelihood computations. In Papers **II** and **III** the local ensemble transform Kalman filter (LETKF) and the ensemble adjustment Kalman filter (EAKF) were considered in filter likelihood computations, respectively.

Besides EnKF and ensemble square root filters, in recent years several ensemble based Kalman filters have been proposed. In addition to ensemble filters, particle filters, where no Gaussian approximation takes place, exist. See, e.g., [46].

4.4 Variational ensemble Kalman filtering

In recent work [8, 7], a new approach for Kalman filtering, called variational Kalman filtering (VKF), was proposed. The approach is based on using the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) optimizing algorithm to obtain low memory approximations of the Hessian and inverse Hessian matrices. These low-storage matrices can be used to approximate the covariance matrices in the Kalman filtering formulas. In Paper **I**, it is shown how this variational approach can be used together with an ensemble approach. The joint method is called the variational ensemble Kalman filter (VEnKF).

In VKF, the following cost function

$$J(x) = \frac{1}{2}(y - \mathcal{H}_k(x))^{\mathrm{T}} R^{-1}(y - \mathcal{H}_k(x)) + \frac{1}{2}(x_k^{\mathrm{p}} - x)^{\mathrm{T}}(C_k^{\mathrm{p}})^{-1}(x_k^{\mathrm{p}} - x) \tag{4.18}$$

is minimized using the LBFGS method to obtain x_k^{est} and the low-storage approximation of C_k^{est}. The same cost function is also used in the VEnKF approach.

In VEnKF, the prior ensemble covariance matrix is approximated as

$$C_k^{\mathrm{p}} = \mathrm{Cov}(\mathcal{M}_k(x_{k-1}^{\mathrm{est}}) + \xi_k) = \mathrm{Cov}(\mathcal{M}_k(x_{k-1}^{\mathrm{est}})) + \mathrm{Cov}(\xi_k) = X_k X_k^{\mathrm{T}} + Q_k. \tag{4.19}$$

In formula 4.19, the standard assumption is made that the state and the observations are uncorrelated. In VEnKF, the MAP point is considered instead of the ensemble mean and the X_k matrix is computed as

$$X_k = ((x_{1,k} - x_k^{\mathrm{p}}), \ldots, (x_{N,k} - x_k^{\mathrm{p}}))/\sqrt{N}. \tag{4.20}$$

In the cost function (4.18), instead of directly forming and inverting the prior covariance matrix (4.19), the Sherman-Morrison-Woodbury (SMW) matrix inversion lemma [20] can

be used. Using the SMW formula, the inverse prior covariance matrix can be written as

$$(C_k^{\mathrm{p}})^{-1} = (X_k X_k^{\mathrm{T}} + Q_k)^{-1} \tag{4.21}$$

$$= Q_k^{-1} - Q_k^{-1} X (I + X^{\mathrm{T}} Q_k^{-1} X)^{-1} X^{\mathrm{T}} Q_k^{-1}. \tag{4.22}$$

This formulation is useful for two reasons. First, it can be used in the cost function without actually forming full $n \times n$ matrices. Second, the matrix to be inverted in the middle, has only a dimension of $N \times N$ instead of $n \times n$. The model error covariance matrix $Q_k \in \mathbb{R}^{n \times n}$ is typically chosen in such a way that the inverse Q_k^{-1} can be easily computed.

An alternative approach is to apply LBFGS to an auxiliary optimizing problem

$$J(u) = u^{\mathrm{T}}(X_k X_k^{\mathrm{T}} + Q_k)u \tag{4.23}$$

with a trivial solution $u = 0$, as in [8], to obtain the low-memory approximation of $(C_k^{\mathrm{p}})^{-1}$ to be used in the cost function (4.18)

In the update procedure of VEnKF, the cost function (4.18) is minimized using the LBFGS method. The MAP estimate x_k^{est} and the covariance matrix C_k^{est} in low memory form are acquired.

Obtaining the analysis ensemble is the most challenging part in ensemble filtering algorithms and it is the part where most of the ensemble based approaches differ. In VEnKF, the new ensemble is created by drawing samples $x_{i,k}^{\mathrm{est}} \sim N(x_k^{\mathrm{est}}, C_k^{\mathrm{est}})$. The low-memory approximation of C_k^{est} can be written in the form

$$C_k^{\mathrm{est}} = B_0 B_0^{\mathrm{T}} + \sum_{i=1}^{d} b_i b_i^{\mathrm{T}}, \tag{4.24}$$

where B_0 is a $n \times n$ matrix that does not need to be stored explicitly and b_i is a $n \times 1$ vector. The summation is done over the number of vectors stored in the LBFGS algorithm. The Appendix of Paper **I** gives details. From this representation of the covariance matrix, it is simple to draw random samples by calculating

$$r = B_0 z + \sum_{i=1}^{d} z_0 b_i^{\mathrm{T}}, \tag{4.25}$$

where $z \sim N(0, I)$ and $z_0 \sim N(0, 1)$. It is easy to confirm that

$$\mathrm{Cov}(r) = B_0 \mathrm{Cov}(z) B_0^{\mathrm{T}} + \sum_{i=1}^{d} b_i \mathrm{Cov}(z0) b_i^{\mathrm{T}} = C_k^{\mathrm{est}}. \tag{4.26}$$

All the building blocks needed for VEnKF method have now been presented. VEnKF can be written as the following pseudo algorithm:

Step 0: Select the initial MAP point x_0^{est} and ensemble $x_{i,0}^{\mathrm{est}}$, $i = 1, \ldots, N$, set $k = 1$;

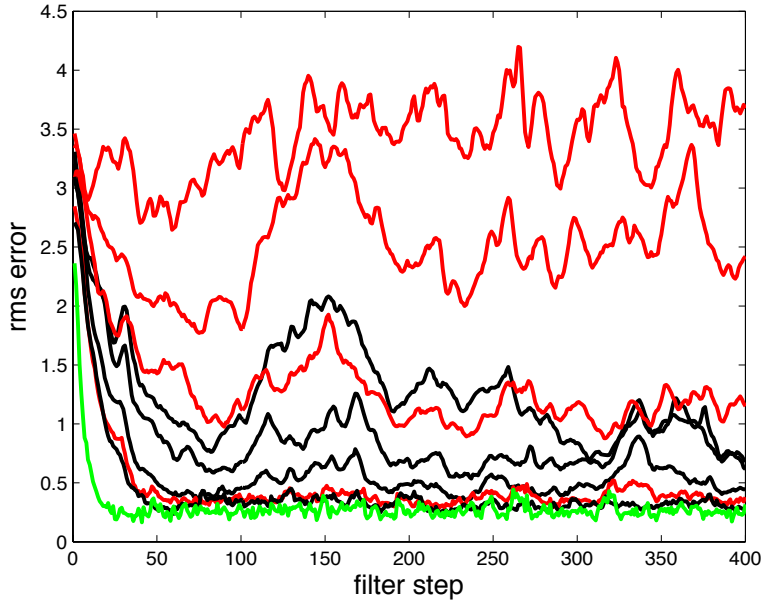


Figure 4.2: Comparison of EnKF (red), VEnKF (black) and EKF (green) with ensemble sizes $N = (10, 15, 20, 40)$ in the Lorenz 95 example. Increasing ensemble size leads to monotonically decreasing error levels for both EnKF and VEnKF. Taken from Paper **I**.

Step 1: Using the forecast model $\mathcal{M}$ push the ensemble and the MAP point forward in time;

Step 2: Define $(C_k^{\mathrm{p}})^{-1}$ using the SMW formula;

Step 3: Minimize the cost function (5.4) using LBFGS to obtain x_k^{est} and the low storage approximation of C_k^{est};

Step 4: Sample new ensemble $x_{i,k}^{\mathrm{est}} \sim N(x_k^{\mathrm{est}}, C_k^{\mathrm{est}})$, $i = 1, \ldots, N$;

Step 5: Set $k = k + 1$ and go to step 1 until the end of observations.

In step 2, one can alternatively apply LBFGS to the auxiliary optimizing problem (4.23) to obtain the low storage approximation of $(C_k^{\mathrm{p}})^{-1}$.

The presented algorithm has several attractive properties. First, the model error covariance matrix Q_k can be directly used in the computations. This is not the case in many other ensemble filtering techniques, where covariance inflation techniques are used [4]. Second, the method works in full space, rather than in a low-dimensional subspace, since the matrices produced by the LBFGS method are full-rank, although low-storage. Third, as the resampling is done in the filtering, the method avoids the inbreeding problems related to some ensemble filters.

In particular, it should be noted that the presented algorithm is, at least in principle, applicable in large-scale Kalman filtering systems and all the above computations can be done without computing full matrices. In addition, unlike the variational Kalman filter (VKF) method, VEnKF does not need adjoint and tangent-linear codes, which may be laborious to implement.

Two downsides of the method should be mentioned. First, the optimizing procedure is somewhat sensitive to the "parameters" of the LBFGS method, e.g., selection of the initial value for the optimizer can be reflected to the quality of the inverse Hessian matrix provided by the method. Second, the method does not allow the direct use of traditional localization techniques, which are often used in large-scale applications to account for sampling errors related to the limited ensemble size (undersampling).

In Paper **I**, numerical experiments are conducted and the VEnKF method is compared against EKF and EnKF using the Lorenz 95 test model [32] and a large-dimension heat equation example. It can be noted that the method performed well against EKF and better than the traditional ensemble Kalman filter method, especially when the ensemble size was low. This result can be seen in a Lorenz 95 comparison, Fig. 4.2, taken from Paper **I**.

Chapter V

On parameter estimation in chaotic systems

In recent years, numerical parameter estimation has become increasingly important in the atmospheric modeling community. Atmospheric general circulation models, like climate and weather models, contain physical parameterizations of sub-grid-scale processes, such as cloud formation and boundary layer turbulence [23], that are not affordable for explicit modeling. Traditionally, the optimal values of these parameterizations are defined using best expertise knowledge [23]. The validation procedure of these parameters is often based on a relatively small number of model simulations and is often done for different parameterizations separately. This procedure is somewhat subjective and therefore open to criticism.

In many engineering problems, parameter estimation is carried out by comparing model simulations directly to observations, e.g., in the least-squares sense. In chaotic systems, however, direct model–observation comparisons become meaningless when the prediction goes in a different trajectory than the observations, due to chaos. This phenomenon can be seen in Figures 2.1 and 4.1.

In this chapter, three different strategies for parameter estimations in chaotic systems, are discussed. The first approach is based on long model simulation runs, where the likelihood is computed using summary statistics. In principle, the modeled and the observed climatologies are compared against each other. In addition to this rather intuitive technique, two techniques based on Kalman filtering are discussed. The second approach is based on the state augmentation approach, where the model state is augmented with the model parameters. These parameters are then updated together with the actual model state. The third approach is based on filter likelihood calculations, where the likelihood is evaluated using the prediction residuals. The central idea is to "integrate out" the state using filtering techniques.

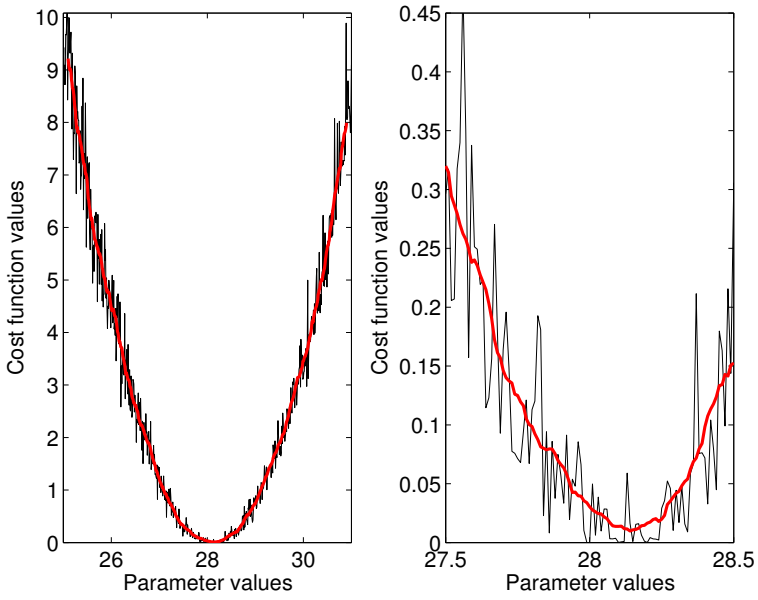


Figure 5.1: An illustration of a summary statistic cost function (black lines). Note the "deterministic noise". Red lines indicate the running mean. Right panel is zoomed near the optimum.

5.1 Likelihood based on summary statistics

Let us consider the following system

$$\begin{cases} x_k &= \mathcal{M}(x_{k-1}, \theta), \\ z_k &= \mathcal{H}(x_k), \end{cases} \tag{5.1}$$

where $\mathcal{M}$ is the state space model that depends on parameters θ and, as in filtering, $\mathcal{H}$ is the observation operator. Vector z_k denotes simulated observations that could, in principle, be compared with the actual observations y_k.

Let us consider the summary statistics computed—using the function f—from the observations $s = f(y_{1:K})$ and from the model simulations $s_\theta = f(z_{\theta,1:K})$. Our aim is to evaluate the posterior distribution

$$p(\theta \mid s) \propto p(\theta) p(s \mid \theta), \tag{5.2}$$

using these summary statistics. If a Gaussian assumption is made, the likelihood can be computed as

$$p(s \mid \theta) \propto \exp\left(-\frac{1}{2}(s - s_\theta)^{\mathrm{T}} C^{-1} (s - s_\theta) \right), \tag{5.3}$$

where C is an error covariance matrix of the residual $r = s - s_\theta$. For example, in a recent study [23] several summary statistics cost functions, like

$$J(\theta) = \frac{(s^{\mathrm{g}} - s_\theta^{\mathrm{g}})}{\sigma_{\mathrm{g}}^2} + \sum_{m=1}^{12} \sum_{\mathrm{z}} (s^{m,\mathrm{z}} - s_\theta^{m,\mathrm{z}})^{\mathrm{T}} (C^{mm,\mathrm{z}})^{-1} (s^{m,\mathrm{z}} - s_\theta^{m,\mathrm{z}}), \tag{5.4}$$

were considered in the ECHAM5 climate model application. In the cost function (5.4), global annual and zonal monthly top of the atmosphere net radiation fluxes were considered. The parameters θ were related to clouds and precipitation.

One problem related to this kind of likelihood formulation is that the cost function may not be a smooth function of the parameters. This non-smoothness causes the effect of "deterministic noise" that is illustrated in Fig. 5.1, where the summary statistics are calculated from the third trajectory of the Lorenz attractor [31]. The example is taken from the paper [5]. The optimization of this kind of spiky cost function is impossible with standard optimizing tools and typically only a local optimum is obtained. If the "noise" level is too high, even MCMC can be a challenging. One option to overcome this problem is to use so-called emulators as, e.g., in [48, 22].

Another challenge to this kind of likelihood formulation is related to the selection of which summary statistic should be used in the cost function. In principle, summary statistics cost functions are intuitive and easy to implement also in large scale application, as demonstrated in [23].

5.2 State augmentation

In state augmentation (SA) methodology (e.g. [5, 26, 21]), a Kalman filtering system where the parameters θ are estimated along the standard state estimation process is considered. In this process, the parameters are no longer static quantities but are updated as new data becomes available.

Let us consider the following system

$$\begin{cases} x_k &= \mathcal{M}_k(x_{k-1}, \theta_{k-1}) + \xi_k^x, \\ \theta_k &= \theta_{k-1} + \xi_k^\theta, \\ y_k &= \mathcal{H}_k(x_k) + \varepsilon_k, \end{cases} \tag{5.5}$$

where, in addition to the basic Kalman filtering pair (4.1), we have an identical forecast model for the model parameters θ. In SA, an augmented state vector $\widetilde{x}_k = (x_k, \theta_k)^{\mathrm{T}}$ is considered. The augmented model can be written as $\widetilde{\mathcal{M}}(\widetilde{x}_k) = (\mathcal{M}(x_k, \theta_k), \theta_k)^{\mathrm{T}}$ and the augmented model error as $\widetilde{\xi} = (\xi_k^x, \xi_k^\theta)^{\mathrm{T}}$. The augmented model can be linearized as

$$\widetilde{M}_k = \frac{\partial \widetilde{\mathcal{M}}(\widetilde{x}_k)}{\partial \widetilde{x}} = \begin{pmatrix} \partial \mathcal{M}(x_k, \theta_k)/\partial x & \partial \mathcal{M}(x_k, \theta_k)/\partial \theta \\ \partial \theta_k / \partial x & \partial \theta_k / \partial \theta \end{pmatrix} \tag{5.6}$$

$$= \begin{pmatrix} \partial \mathcal{M}(x_k, \theta_k)/\partial x & \partial \mathcal{M}(x_k, \theta_k)/\partial \theta \\ 0 & I \end{pmatrix}. \tag{5.7}$$

Matrix (5.7) can be used in standard Kalman filtering equations, presented in Chapter 4.

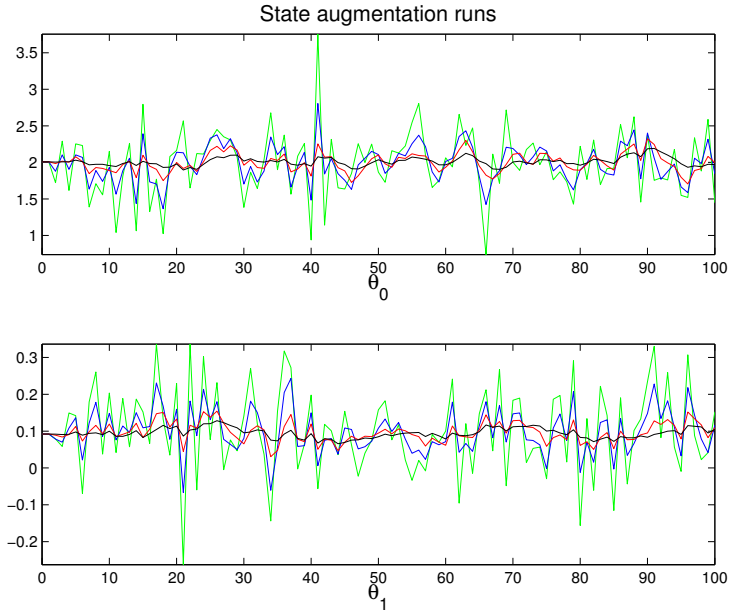


Figure 5.2: Runs using the EKF version of the SA method when the diagonal elements of Q^θ are taken to be 0.1% (black), 1% (red),10% (blue) and 100% (green) of the optimal initial values. The effect of the size of Q^θ was as expected. When Q^θ is set to be small, the method reacts slowly to new data and the parameter values take small steps. Taken from Paper **II**.

In Kalman filter formulation, a model error covariance matrix Q_k is also needed. In SA, the augmented model error covariance matrix can be modeled as a block diagonal matrix

$$\widetilde{Q}_k = \begin{pmatrix} Q_k^x & 0 \\ 0 & Q_k^\theta \end{pmatrix} \tag{5.8}$$

where Q_k^x is related to the actual model state and Q_k^θ is related to the model parameters. The latter term can be seen as an additional tuning handle that controls the rate of change of the parameters in the SA system. In addition, increasing Q_k^x will cause a smaller rate of change to the parameters. Further discussion of the effect of the matrix $\widetilde{Q}_k$ in an SA system is given in Appendix B of Paper **II**.

SA runs using a parametrized Lorenz 95 model [58] are illustrated in Fig. 5.2. The runs are made so that the diagonal elements of the static model error covariance matrix Q^θ are taken to be 0.1 %, 1 %, 10 % and 100 % of the optimal parameter values, see Paper **II** for details. In all cases, the SA method converges quickly to values near the optimum. It can be seen that when the diagonal elements are set to be small, the method reacts slowly to new data and the parameter values take small steps. On the other hand, if the diagonal elements are set to be large, the method allows larger deviations but can yield unrealistic values for the parameters. In this example, systematic temporal variations in

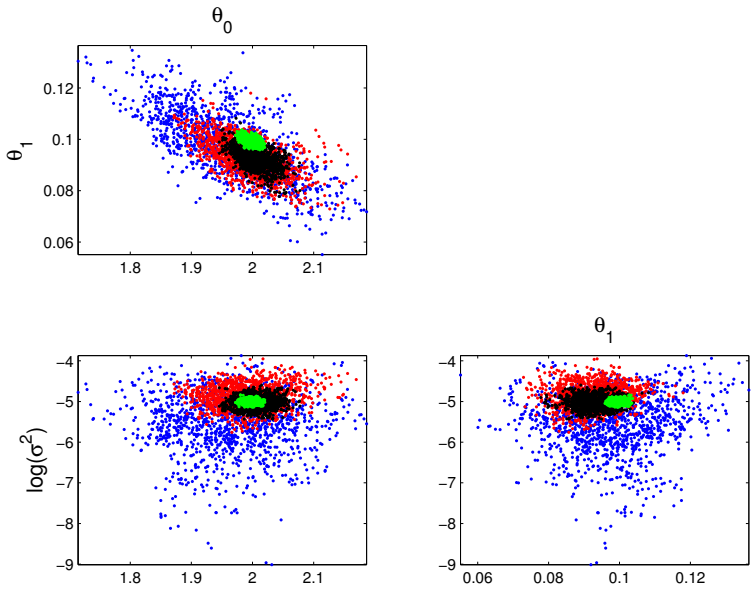


Figure 5.3: Scattering plot of parameter pairs from MCMC runs using 10 (blue), 20 (red), 50 (black) and 500 (green) day simulations. Clear tightening of the posterior can be observed as the simulation length increases. The third parameter is related to the model error covariance. Taken from Paper **II**.

the parameter value cannot be observed. However, if such variations existed, SA could be a useful method for tracking them.

In principle, the SA approach can be implemented in large-scale systems where a Kalman filter is available (see, e.g., [47]). It should be noted that in addition to EKF, SA can also be used together with ensemble and other filters. As an online method, SA has no problem with the stochasticity of the filtering algorithms. In SA, the parameters are modeled to be dynamic quantities and parameter trajectories like those illustrated in Fig. 5.2 are created. However, these trajectories cannot be interpreted in a statistical sense because they depend entirely on the additional tuning handle Q_k^θ.

5.3 Filter likelihood

Likelihood computations can be done via filtering methods. In this filter likelihood concept, the basic idea is to "integrate out" the state using filtering techniques and calculate the likelihood of the parameters using the prediction steps of the filter.

Let us write the filtering pair (4.2) in the case where the prediction model depends on

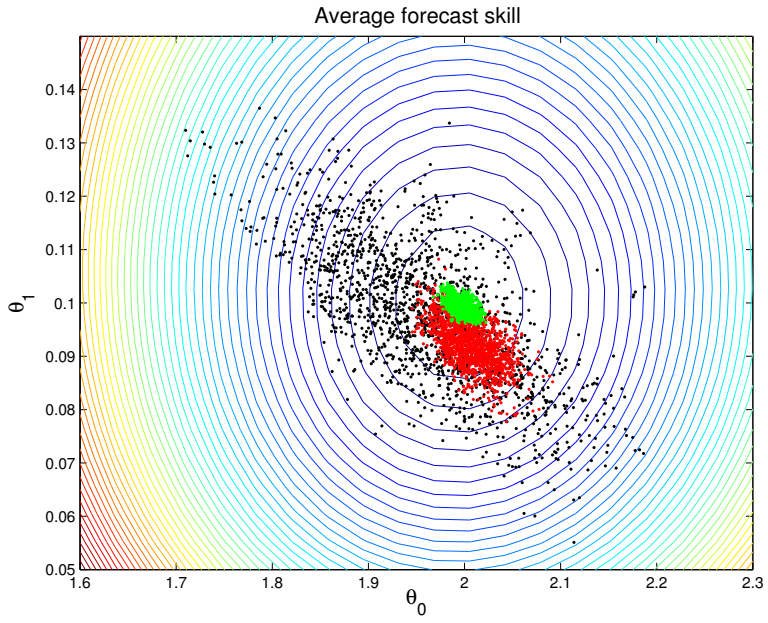


Figure 5.4: An illustration of the average forecast skill. Black, red and green dots indicate the results from 10, 50 and 500 day MCMC simulations, respectively. Blue contour colors indicate high forecast skill. Taken from Paper **II**.

the parameters θ and we have the prior information $p(\theta)$

$$\begin{cases} x_k & \sim p(x_k \mid x_{k-1}), \\ y_k & \sim p(y_k \mid x_k), \\ \theta & \sim p(\theta). \end{cases} \tag{5.9}$$

Our target is to find the posterior $p(\theta|y_{1:K})$ using Bayes' formula (3.2). Using the general assumption of the filtering methods that the observations are conditionally independent, we can use the chain rule for joint probability and write the posterior distribution as

$$p(\theta \mid y_{1:K}) \propto p(\theta)p(y_{1:K} \mid \theta) \tag{5.10}$$

$$= p(\theta)p(y_K \mid y_{1:K-1}, \theta)p(y_{K-1} \mid y_{1:K-2}, \theta)\dots \times p(y_2 \mid y_1, \theta)p(y_1 \mid \theta). \tag{5.11}$$

Hence,

$$p(\theta \mid y_{1:K}) \propto p(\theta)p(y_1 \mid \theta) \prod_{k=2}^{K} p(y_k \mid y_{1:k-1}, \theta). \tag{5.12}$$

The predictive distribution of the observations $p(y_k \mid y_{1:k-1}, \theta)$ can be computed using (4.4). Now the predictive distribution is written as

$$p(y_k \mid y_{1:k-1}, \theta) = \int p(y_k \mid x_k, \theta)p(x_k \mid y_{1:k-1}, \theta)\,\mathrm{d}x_k, \tag{5.13}$$

since the model depends on parameters θ.

In the context of EKF, the likelihood can be written using the innovation (4.8) and its error covariance matrix (4.9)

$$\begin{aligned} p(y_{1:K} \mid \theta) &= p(y_1 \mid \theta) \prod_{k=2}^{K} p(y_k \mid y_{1:k-1}, \theta) \\ &= \prod_{k=1}^{K} \exp\left(-\frac{1}{2} r_k^T (C_k^{\mathrm{r}})^{-1} r_k \right) (2\pi)^{-n/2} |C_k^{\mathrm{r}}|^{-1/2} \\ &\propto \exp\left(-\frac{1}{2} \sum_{k=1}^{K} \left[r_k^T (C_k^{\mathrm{r}})^{-1} r_k + \log |C_k^{\mathrm{r}}| \right] \right), \end{aligned} \tag{5.14}$$

where $|\cdot|$ denotes the matrix determinant. It should be noted that the normalizing "constant" $\log |C_k^{\mathrm{r}}|$ has to be included, since it depends on the parameters via the prediction model. The likelihood can be written similarly for other filters.

The filter likelihood function (5.14) can be written as a cost function, as done in Chapter 3. This cost function can be evaluated as the filter evolves and only the latest filtering outputs need to be stored. The cost function can be minimized using standard optimizing tools or the samples can be drawn using the MCMC methodology.

In Papers **II** and **III**, this filter likelihood (FL) method was studied using the parameterized Lorenz 95 model [58]. Example runs using different amounts of data in the simulations are illustrated in Fig. 5.3. It can be noted that the method "converges" when the amount of data increases. In Fig. 5.3, parameters θ_1 and θ_2 are related to the actual model parameters and the third parameter is related to the model error covariance matrix Q, which is also estimated.

In Paper **II**, the method is validated using metrics that take into account the true state of the system. In Fig. 5.4, 6-day average forecast skills are illustrated. It can be noted that MCMC samples converge quite accurately where the best forecast skills are located.

Filter likelihood is computationally a rather expensive technique, since in addition to a normal state estimation run, it needs several iterations, depending on the number of parameters and other factors, which can be a bottle-neck in large-scale applications. As mentioned in Chapter 4, although the Kalman filter in its traditional form is infeasible in large-scale applications, several alternatives exist. However, it should be noted that the use of stochastic state estimation techniques, like EnKF and VEnKF presented in Chapter 4, is challenging, since the use of a stochastic state estimation method yields a stochastic cost function, which is rather difficult to analyze using standard optimizing and sampling techniques.

5.4 Estimation of the filter related parameters

As noted in Section 5.3 and in Papers **II** and **III**, the filter likelihood approach can be used to estimate the filter related parameters, in addition to the standard model parameters. This is particularly useful with the model error covariance matrix Q, since

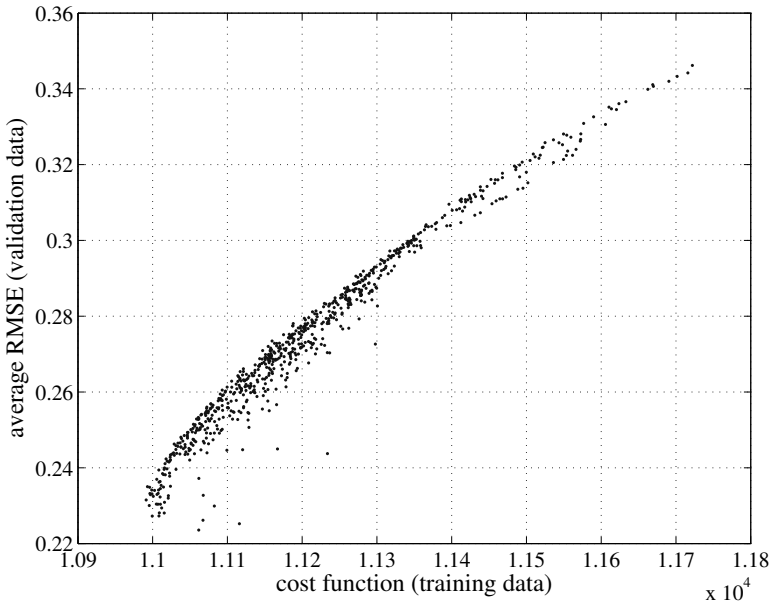


Figure 5.5: A scatter plot of cost function and RMSE values using different model error covariance matrix parameters in a 1600-dimensional two-layer quasi-geostrophic benchmark model. Taken from [51].

it is often considered as a tuning handle in the state estimation systems. In Paper **III**, the FL technique was also used to estimate the covariance inflation and localization parameters of the EAKF based filtering system. In [51], this idea was studied further in a model error context using a 1600-dimensional two-layer quasi-geostrophic benchmark model [39].

From [51], it can be concluded that the model error covariance matrix is an important factor of the prediction systems, in particular, when the available observations are noisy. In addition, it can be concluded that the model error covariance matrix parameters acquired using the FL technique agreed well against the metrics that use the information of the true state of system. An example is illustrated in Fig. 5.5, where the cost function values of different parameter combinations of the model error covariance matrix are plotted against RMSE values. A clear correlation between the two metrics can be seen.

5.5 Dilemma of the uniqueness of climate and weather model parameters

In this section, a dilemma of the uniqueness of climate and weather model parameters is discussed.

As noted earlier, atmospheric prediction systems contain several parameterization schemes and numerical approximations. These approximations render all model components

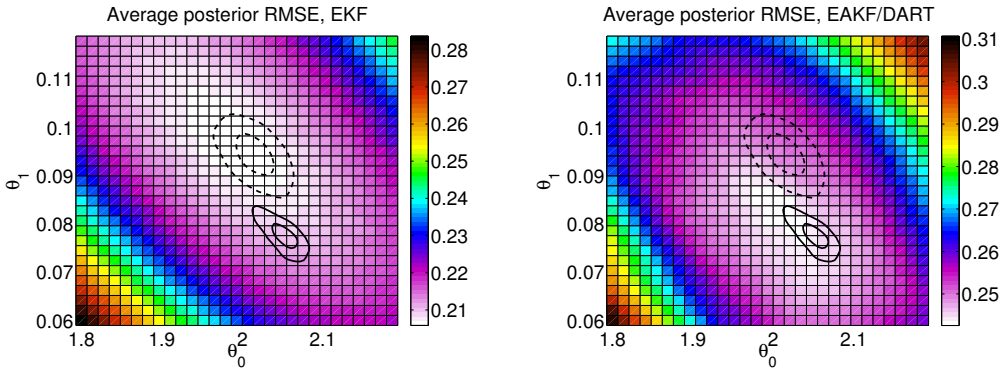


Figure 5.6: Average posterior RMSE values when using EKF (left panel) and EAKF (right panel). Lower values indicate better performance. Dashed and solid lines indicate the 50 % and 95 % probability regions as obtained with MCMC in the case of EKF and EAKF, respectively. RMSE values are depicted by the background color scheme. Note the difference in color scales. Taken from Paper **III**.

somewhat imperfect. One way to obtain the optimal parameter values of different components is to use dedicated laboratory or focused field campaign measurements separately for each model component. Another way is to tune all parameters at once, e.g., with the methods presented earlier in Sections 5.1, 5.2 and 5.3. It is qualitatively obvious that the two approaches do not necessarily coincide and a practical optimum is a compromise between the two. From a philosophical point of view, this yields a dilemma since the optimal model parameter values, which depend also on the approximations of the other model components, can correspond only to one observational truth.

In Paper **III**, this dilemma is demonstrated in practice using the parameterized Lorenz 95 system. The outline of the demonstration is that two different state estimation component, based on EKF and EAKF filtering methods, are used in the prediction system. The systems are then validated against the true simulation. It can be observed that the optimal model parameters, in an RMSE sense, correspond to a unique but different parameter setting depending on the choice of the state estimation component. These results are illustrated in Fig. 5.6. In addition, it can be noted that the FL based MCMC chains correspond to the correct parameter values of the systems.

The demonstration of this dilemma has a clear interpretation in many applications, since it has typically been considered that the model parameters and the assimilation component are somewhat independent of each other. The results of Paper **III** indicate that the model parameters are no longer only properties of the model alone, but the whole prediction system should be considered at once. From the practical point of view, it can be noted that if the data assimilation component of the prediction system is updated, the model parameters should be re-estimated too.

Chapter VI

GOMOS data and chemical state estimation application

Satellite instruments have provided information of Earth's atmosphere since the 1950s, when the firsts weather satellite instruments were launched. Satellite instruments to study the chemical composition of the atmosphere have been available since the 1970s. These instruments have been particularly important in global ozone and air-quality monitoring. In recent years, the need for global observations in climate studies has made the retrieval of greenhouse gases from space increasingly important.

Satellite-borne measurements of the atmospheric trace gases are always indirect and—in order to retrieve atmospheric parameters form measurement data—inverse algorithms are needed. The retrieval problems have created a rich field of science and many techniques to retrieve atmospheric trace gases have been proposed. In addition to satellite observations themselves, the role of data characterization, error analysis, validation and satellite-to-satellite intercomparisons have been recognized [43, 55].

State-of-the-art studies of the atmospheric composition rely on both observations and model studies [12]. The chemistry models vary from simple box models to complex thee-dimensional chemistry transport models. Typically, meteorological parameters like temperatures and wind fields are taken from external sources. Since the 1990s, chemical state estimation has been studied and, e.g., modern ozone forecasts are based on data assimilation. Successful forecast of ozone need satellite observations, chemistry transport model and high-quality forecast of dynamical fields from NWP models [29].

In this chapter, satellite-borne data from the GOMOS instrument are discussed and considered in a chemical state estimation application. In Section 6.1, an introduction to the operative GOMOS dataset is given, and an alternative retrieval algorithm based on the so-called one-step approach is presented. In section 6.2, the basics of the nitrogen trioxide (NO_3) chemistry is recalled and satellite-borne NO_3 vertical profiles—retrieved from the GOMOS stellar occultations and the SAGE III lunar occultations—are compared against each other. Finally, in Section 6.3, the GOMOS NO_3 observations are considered in a chemical data assimilation system in order to retrieve stratospheric temperature profiles.

6.1 GOMOS data and the one-step retrieval algorithm

The Global Ozone Monitoring by Occultation of Stars (GOMOS) satellite instrument was launched in March 2002 by the European Space Agency on board the Envisat platform

[10] and it operated until the end of the ENVISAT mission in 2012. During the mission, GOMOS provided 866 000 individual vertical profiles of ozone, NO_2, NO_3 and other species [27]. About half of the occultations were made during nighttime.

The GOMOS measurement principle is relatively simple and it is based on the stellar-occultation technique. The stellar spectra at different tangent altitudes are obtained with a sampling resolution of $0.3 - 1.7\,\mathrm{km}$. One occultation contains roughly 100 measurement spectra at 1 416 different wavelengths in the UV–visible wavelength region. Hence, roughly 150 000 individual measurements per occultation are obtained.

The transmission can be modeled using the Beer-Lambert law

$$T_{\mathrm{ext}}^{\mathrm{mod}}(\lambda, l) = \exp(-\tau(\lambda, l)) \tag{6.1}$$

where the optical depth $\tau(\lambda, l)$ is given as

$$\tau(\lambda, l) = \sum_{j=1}^{\mathrm{nconst}} \int_{l} \alpha_j(\lambda, \mathrm{T}(s))\rho_j(s)\,\mathrm{d}s, \tag{6.2}$$

where ρ_j is a local density profile and $\alpha_j(\lambda, \mathrm{T}(s))$ is the constituent, wavelength and temperature dependent cross-section. The integration is done along the line-of-sight.

When the problem has been discretized—using the vectorized notation—the full transmission model can be written as

$$T_{\mathrm{ext}}^{\mathrm{mod}}(\rho) = \exp(-(K \otimes A)\rho), \tag{6.3}$$

where $K \otimes A$ is the Kronecker product between the kernel operator K and the cross-sections A. Typically, six constituents (O_3, NO_2, NO_3 and three aerosol parameters), 1 500 wavelengths and 100 model layers are considered. Thus, the length of the input vector ρ is $6 \times 100 = 600$ and the length of the output vector $T_{\mathrm{ext}}^{\mathrm{mod}}$ is $1\,500 \times 100 = 150\,000$. Hence, the linear operator $K \otimes A$ has to be a $150\,000 \times 600$ matrix. In addition to O_3, NO_2, NO_3 and aerosol parameters, neutral density and some minor trace gases like OClO and BrO can be included in the model.

In the operative GOMOS algorithm, the retrieval problem has been split in two parts: the spectral inversion and the vertical inversion [28]. In the first part, horizontally integrated line densities of O_3, NO_2, NO_3 and aerosols are retrieved simultaneously using a combination of absolute and differential cross-sections. In the second part, local density profiles are retrieved from these horizontally integrated line densities at different tangent altitudes. In the latter part, a Tikhonov regularization is applied to compensate the low signal to noise ratio and the smoothness requirements are used as priors.

GOMOS one-step retrieval algorithm. The basic idea of the GOMOS one-step retrieval algorithm is to conduct the spectral and the vertical inversion of the operative GOMOS algorithm simultaneously using the full model (6.3). Historically, the use of a two-step strategy in the operative algorithm is mainly related to the enormous size of the problem. The possibility of one-step inversion was studied already in the early days of the GOMOS mission [57, 17]. Here the version of the GOMOS one-step retrieval algorithm from Paper **V** is presented.

The one-step retrieval of the GOMOS measurements can be seen as a minimization of the following cost function consisting of the likelihood and the prior:

$$J(\rho) = \frac{1}{2}(T_{\text{ext}}^{\text{mod}}(\rho) - T_{\text{ext}}^{\text{meas}})^{\text{T}} C_{\text{meas}}^{-1}(T_{\text{ext}}^{\text{mod}}(\rho) + T_{\text{ext}}^{\text{meas}}) + \frac{1}{2}(\rho_{\text{prior}} - L\rho)C_{\text{prior}}^{-1}(\rho_{\text{prior}} - L\rho). \tag{6.4}$$

where L is a linear operator based on how the prior is given. For example, when the smoothness requirements are considered as priors L is selected to be the second derivative and ρ_{prior} is set to zero. Cost function (6.4) could, in principle, be minimized using non-linear optimizing algorithms like the Levenberg-Marquardt algorithm used in the spectral inversion part of the operative algorithm. However, in practice, this problem can be solved using a generalized linear modeling (GLIM) approach with the iterative re-weighted least squares algorithm (IRLS), where the special exponential structure of the problem can be exploited for numerical efficiency [35].

Let us next study how the GLIM approach works in practice. Let us consider the following model

$$y = \exp(Ax) + \varepsilon, \tag{6.5}$$

where y are the measurements, A is the linear operator, $\varepsilon \sim N(0, \sigma I)$ is the measurement noise and x is the unknown state vector. Readers familiar with the GLIM vocabulary can note that the *link function* between the measurements and the linear operator is logarithmic and—as the measurement error is assumed Gaussian—the *variance function* is simply identity.

Using the IRLS algorithm, starting from the initial guess of x, the unknown x can be solved iteratively. In the iteration steps, the auxiliary problem based on the linear equation

$$z = Ax, \tag{6.6}$$

is solved using the standard-deviation weights w, where

$$\eta = Ax, \tag{6.7}$$
$$\mu = \exp(\eta), \tag{6.8}$$
$$z = \eta + (y - \mu)/\mu, \tag{6.9}$$
$$w = \sigma/\mu. \tag{6.10}$$

When prior information x_{prior} is considered, the iterative solution x can be seen as the argument that minimizes the following quadratic cost function

$$J(x) = \frac{1}{2}(Ax - z)^{\text{T}} C_w^{-1}(Ax - z) + \frac{1}{2}(x_{\text{prior}} - Lx)^{\text{T}} C_{\text{prior}}^{-1}(x_{\text{prior}} - Lx), \tag{6.11}$$

where C_w is the diagonal covariance matrix induced by the weights w. As the operators A and L in the cost function (6.11) are linear, we can obtain the solution x using basic linear algebra

$$x = (A^{\text{T}} C_w^{-1} A + L^{\text{T}} C_{\text{prior}}^{-1} L)^{-1}(A^{\text{T}} C_w^{-1} z + L^{\text{T}} C_{\text{prior}}^{-1} x_{\text{prior}}). \tag{6.12}$$

The non-trivial part in the one-step algorithm is the selection of the (diagonal) prior covariance matrix C_{prior} for the derivatives. From the statistical point of view, the

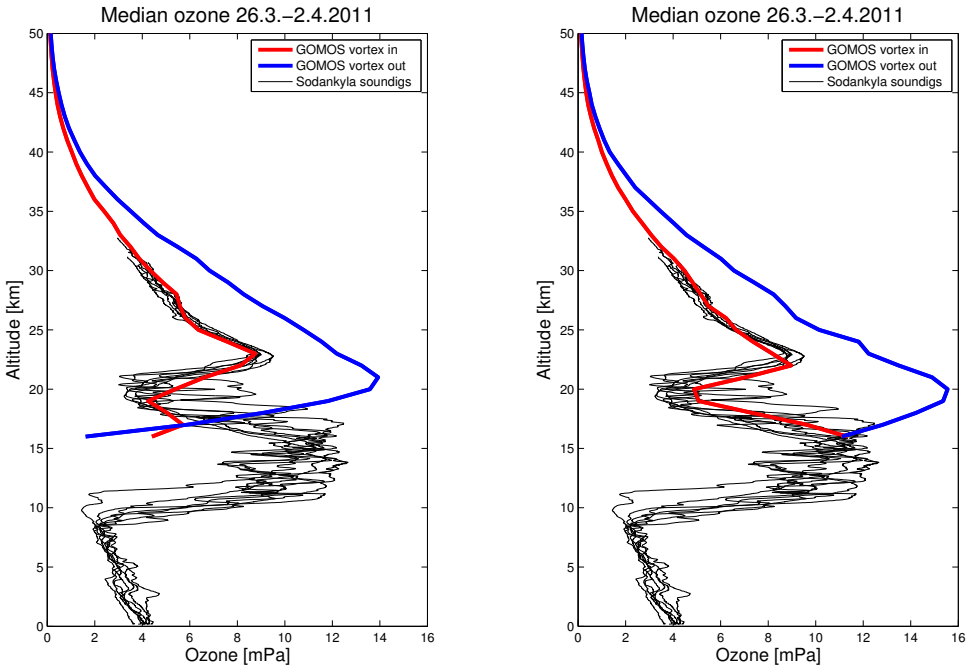


Figure 6.1: The median GOMOS ozone profiles retrieved using the operative (left panel) and the one-step (right panel) retrieval algorithms inside and outside the Arctic polar vortex. Sodankylä sounding profiles are illustrated too. Clear differences between the retrieval algorithms can be observed. Taken from Paper **V**.

covariance matrix should reflect our prior knowledge, e.g., from theory, about how smooth the profiles are or how big steps the profiles can make between the altitudes. Paper [17] provides details. From the practical point of view, the prior covariance matrix can be seen as a regularization tuning handle of the inversion similar to the Tikhonov regularization parameter in the operative algorithm.

In the case study part of Paper **V**, the GOMOS ozone retrievals—made with the one-step and operative retrieval algorithms in the Arctic region during the exceptional ozone-depletion conditions in spring 2011—were compared. The median ozone profiles inside and outside the polar vortex using the operative (left panel) and the one-step (right panel) retrieval algorithms are illustrated in Fig. 6.1. It can be noted that the one-step algorithm produces more realistic results than the operative algorithm. The median of the operative profiles starts to produce too low values at 21 km outside and at 17 km inside the polar vortex. The medians of the one-step profiles are as expected.

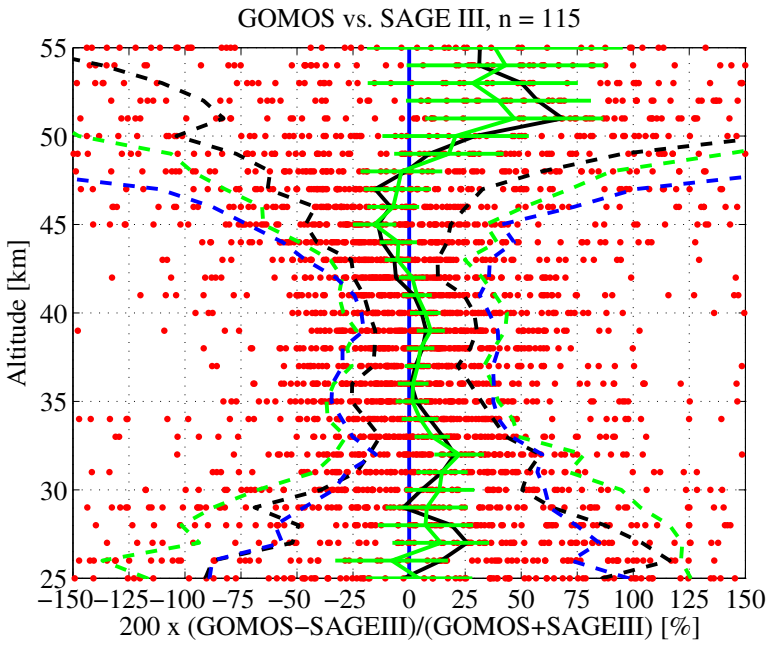


Figure 6.2: The statistics of the collocated GOMOS-SAGE III pairs from the years 2002–2004. The black solid line is the median of the individual differences and the black dashed lines correspond to median ± interquartile deviation. The green solid line is the median filtered mean and the green horizontal lines represents its 95 % confidence limits. Taken from Paper **IV**.

6.2 NO_3 chemistry and comparison results

NO_3 chemistry. The radical nitrate NO_3 plays an important role in stratospheric nighttime chemistry. It is chemically coupled to nitrogen oxides (NO_x = NO and NO_2), whose reactions in the middle atmosphere form the primary catalytic ozone destruction cycle. In the stratosphere, NO_3 has a strong diurnal variation and in the presence of sunlight it is extremely quickly destroyed by photolysis.

During nighttime, in the absence of heterogeneous processes, the NO_3 chemistry scheme is believed to be relatively simple with three reactions [45]:

$$NO_2 + O_3 \xrightarrow{k_1(T)} NO_3 + O_2, \tag{R1}$$

$$NO_3 + NO_2 + M \xrightarrow{k_2(T)} N_2O_5 + M, \tag{R2}$$

$$N_2O_5 + M \xrightarrow{k_3(T)} NO_3 + NO_2 + M, \tag{R3}$$

where k_1, k_2 and k_3 are (strongly) temperature dependent reaction constants and M is any molecule. JPL report [44] provides details. NO_3 is mainly produced by Reaction (R1) of NO_2 and O_3. The sink of NO_3 is Reaction (R2) with NO_2, which produces N_2O_5. The thermal decomposition of N_2O_5 is an additional source of NO_3.

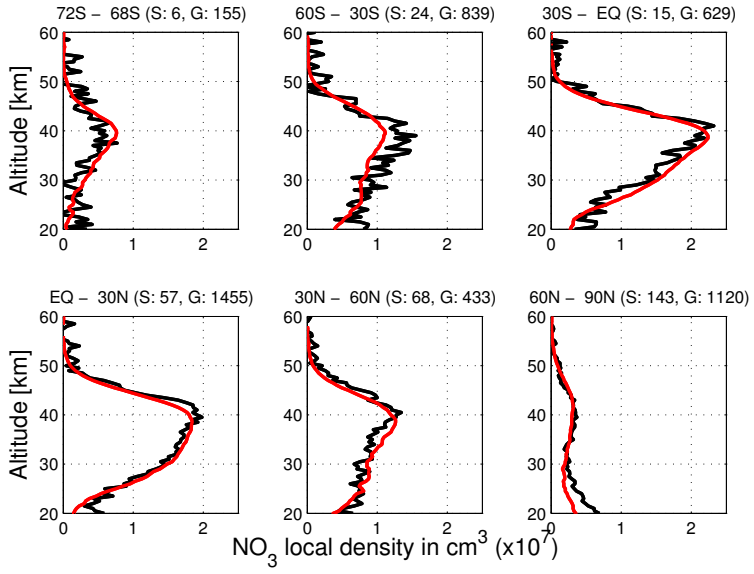


Figure 6.3: GOMOS (red) and SAGE III (black) zonal medians in six different latitude bands in 2004. Taken from Paper **IV**.

The steady state expression for NO_3, based on (R1)–(R3), can be written as [45]

$$NO_3 = \frac{k_1(T)O_3}{k_2(T)M} + \frac{k_3(T)N_2O_5}{k_2(T)NO_2}. \tag{S1}$$

If the thermal decomposition of N_2O_5 (R3) can be neglected (see, e.g., [34] for discussion), the expression (S1) is reduced to

$$NO_3 = \frac{k_1(T)O_3}{k_2(T)M}. \tag{S2}$$

In this expression, NO_3 simply depends on ozone, air and temperature and, in particular, is independent of NO_2. This steady state expression is sometimes used in global chemistry transport models [33]. It also used in the GOMOS NO_3 study [34].

NO_3 intercomparisons. Due to the strong diurnal variation of NO_3—it is undetectable during daytime—NO_3 datasets are scarce, since most of the satellite instruments use the Sun as the light source. In practice, only nighttime observations exist, although, a recent study [36] demonstrates that NO_3 slant-column densities can be observed through sunrise and sunset. The lack of validation datasets and the strong diurnal variation make the validation of satellite-borne NO_3 datasets challenging and previous validation studies of GOMOS NO_3 profiles have been very limited [33, 41].

As GOMOS uses stellar light as a light source, the quality of the measurements and the observations varies from star to star. NO_3 can be observed in the 25–45 km altitude range with a precision of 20–40 % with bright and medium bright stars. Moreover, cool stars are slightly more favorable for the NO_3 retrieval [55].

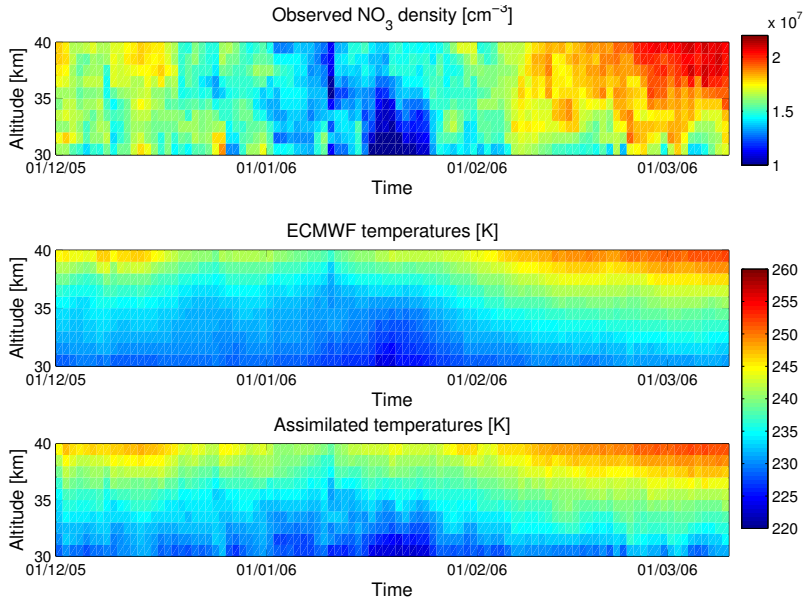


Figure 6.4: An illustration of assimilation results. In addition, GOMOS NO_3 observations and ECMWF temperature profiles are presented.

In Paper **IV**, first global comparisons between the two unique satellite-borne datasets of NO_3 are presented. These GOMOS and SAGE III intercomparison results are illustrated in Figures 6.2 and 6.3. From Fig. 6.2, it can be noted that between the altitudes 25 km and 45 km the median difference between these two datasets is within ±25 %. A small positive bias of some 10 % can also be observed below 40 km. The spread between the observations starts growing below 30 km and above 45 km. From the zonal median profiles, illustrated in Fig. 6.3, we can see visual agreement, showing that the climatological median profiles are comparable.

6.3 Retrieval of stratospheric temperature profiles using a Kalman filtering approach

Chemical data assimilation, pioneered in the 1990s [29], is a relatively recent application area of state estimation. Here, the assimilation of GOMOS NO_3 profiles in a simple system based on nighttime NO_3 chemistry is considered, in order to retrieve unobserved stratospheric temperature profiles.

State and evolution model. The state variable x of the system consists of temperature profiles and is written as

$$x = \mathrm{T}. \tag{6.13}$$

The evolution model $\mathcal{M}$ is simply an identical mapping

$$x_{k+1} = x_k. \tag{6.14}$$

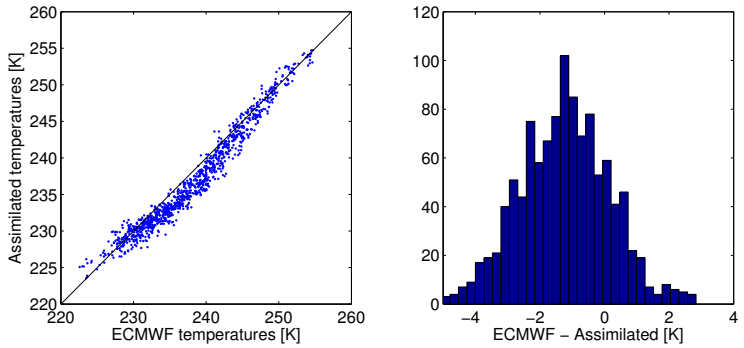


Figure 6.5: Comparisons between assimilated and ECMWF temperatures. Left panel: scatter plot and 1 : 1 line. Right panel: histogram of differences.

The error associated with the evolution model is set to 10 % of the initial values and the model error covariance matrix is simply diagonal

$$Q = \mathrm{diag}((0.10 \times x_0)^2). \tag{6.15}$$

These settings reflect the modeling assumptions of the situation. There is no transport model, instead it is assumed that night-to-night temperature variation can not be zonally more than 30 % of the initial values.

Observations. GOMOS nightly averaged NO_3 profiles are used as observations of the system. The observations are averaged zonally in the latitude band 30° S–30° N at 30–40 km. The observations are taken from the GOMOS IPF version 5.0 dataset and only the brightest stars are considered. The observation error covariance matrix is set diagonal, defined by the error values reported in the GOMOS products. The observations operator that maps from the state space to the observation space is based on the steady state expression (S2). It can be written as

$$\mathcal{H}(x_k) = \frac{k_1(x_k)\mathrm{O}_3}{k_2(x_k)\mathrm{M}}. \tag{6.16}$$

In the observation operator, O_3 and M are fixed to values reported by the GOMOS products. Factors for reaction rate constants k_1 and k_2 are taken from the JPL recommendations [44].

Analysis. The update step of the assimilation system is based on the basic EKF formulas (4.10)–(4.11). The innovation is calculated as a discrepancy between observed and modeled NO_3:

$$r_k = (\mathrm{NO}_3)_k^{\mathrm{obs}} - \mathcal{H}(x_k). \tag{6.17}$$

In order to compute the covariance matrix C_k^{r}, the non-linear observation model $\mathcal{H}(x_k)$ is linearized numerically.

Results. The assimilation system discussed above was considered in the tropical region 30° S–30° N for 100 days starting from 1st of December 2005. The initial values for NO_3

and temperature profiles were set to the values reported in the GOMOS products. The initial error was set to 50 % of the initial values.

The assimilated temperature profiles are illustrated in Fig. 6.4. In addition, the GOMOS NO_3 observations and ECMWF temperatures, also given in the GOMOS products, are illustrated. First, clear visual correlation between NO_3 and temperature profiles is observed, as expected. Second, it can be observed that the assimilated and ECMWF temperature profiles show similar pattern and time evolution. In general, the assimilated values are slightly higher.

In order to analyze the assimilated temperatures in more detail, comparisons against ECMWF temperatures are conducted. These results are illustrated in Fig. 6.5. In the left panel, a scatter plot between the values and $1:1$ line is illustrated and in the right panel, the histogram of differences is illustrated. Good overall agreement can be seen, although the scatter plot is slightly banana-shaped. The mean difference between the products is $-1.24\,\mathrm{K}$. It can be noted that these results are consistent with an earlier study [34], where no state estimation process was considered and only steady state expression (S2) was used. However, the retrievals of that study were slightly more biased against ECMWF temperatures, the bias being on average around $+3\,\mathrm{K}$. To the best of the author's knowledge, this is the first time that temperatures have been retrieved using a sequential chemical state estimation scheme.

Finally, it should be noted that because of the simplicity of the evolution model, it being only an identical operator, it was not possible to use FL computations to estimate parameters of the system related, e.g., to the model error covariance matrix and reactions (R1)–(R3).

Chapter VII

Conclusions

In this dissertation, numerical methods for state and parameter estimation in chaotic systems were studied. From the state estimation studies, it can be concluded that the presented variational ensemble Kalman filter (VEnKF) performed well against the state-of-the-art ensemble Kalman filter (EnKF) and extended Kalman filter (EKF). In the case studies, VEnKF worked with far fewer ensemble members than EnKF and it is, in principle, applicable in large-scale Kalman filtering systems.

The presented filter likelihood (FL) technique seems a promising method for parameter estimation in chaotic systems. In contrast to the well-known state augmentation approach, in FL, the parameters are modeled to be static and FL can be used for statistical analysis. In the case studies, FL was used together with Markov chain Monte Carlo (MCMC) methods to analyze the probability regions of the parameters. Good comparison results against standard validation metrics that take into account the true state of the system, were obtained. Previously, FL has been used in time series analysis and together with stochastic differential equations, but it has not been considered in large-scale climate and weather model applications. In addition to model parameters, the FL approach can be used to estimate filter related parameters related, e.g., to the model error covariance matrix. The dilemma related to the parameters of the climate and weather models was also discussed.

In the data-oriented part of this dissertation, GOMOS data were considered and the one-step retrieval algorithm presented. The one-step algorithm proved it usefulness in the Arctic region and produced more realistic ozone-profile results than the operative algorithm at the lower stratosphere, at 15–20 km altitude. In the validation study, GOMOS and SAGE III NO_3 vertical profiles were compared against each other. Statistical analysis of a limited amount of collocated pairs showed that between the altitudes 25 km and 45 km the median difference between these two datasets is within $\pm 25\,\%$. From the zonal median profiles, good overall agreement can be seen, showing that the climatological median profiles are comparable. In addition, the GOMOS data proved to be useful in a chemical state estimation study, where stratospheric temperature profile profiles were retrieved using a Kalman filtering approach.

The main result of this dissertation is consideration of the filter likelihood concept together with adaptive MCMC in atmospheric modeling. Previously, this kind of hybrid concept has not been considered in the atmospheric modeling community. In principle, the application area of FL can be extended as far as the Kalman filtering. This hybrid approach could become a standard tool also for analyzing non-chaotic models.

In non-chaotic systems, in contrast with the classical approach, the FL concept has several attractive features. For example, in many standard non-chaotic parameter estimation problems, the initial values of the system are poorly known and have to be estimated together with the parameters. This can make the parameter estimation problem substantially larger than originally assumed. In addition, in the classical approach, the models are assumed to be perfect and no model errors are considered. Using the FL concept, both these problems can be avoided, since the initial and model errors can be explicitly modeled and the estimation of the initial values can be neglected. It should be noted that if the initial and the model errors of the systems are set to zero, the filter likelihood reduces to the classical Gaussian likelihood approach.

In this dissertation, the FL concept was studied extensively using the standard low-order Lorenz 95 testing environment together with adaptive MCMC approach. Performance testing in real-life applications is left for future research.

In Paper **III**, the FL method was implemented to the data assimilation research testbed (DART), which is available for large-scale state estimation applications [3]. For example, in a recent study it was used to estimate ECHAM6 climate model closure parameters together with the SA approach [47]. Research of the authors of Paper **III** aims to estimate these parameters using the filter likelihood concept in the DART environment. The preliminary test hints that such parameter estimation is possible. In large-scale applications, the hybrid FL-MCMC approach can be used with the MCMC tricks presented in [52, 50].

Future work could also include considering the filter likelihood concept together with the GOMOS NO_3 data, in order to estimate the uncertainty of the reaction parameters of (R1)–(R3). In particular, it would be interesting to determine the factors of $k_1(\mathrm{T})$ that control the nighttime formation of NO_3 and N_2O_5.

Finally, motivated by the positive results of the case study of Paper **V**, future research should also study in more detail the capabilities of the one-step algorithm in the upper troposphere and lower stratosphere (UTLS) region, where satellite retrievals are challenging.

[1] Chaos theory. `http://en.wikipedia.org/wiki/Chaos_theory`. Accessed: 2013-07-05.

[2] Extended Kalman filter. `http://en.wikipedia.org/wiki/Extended_Kalman_filter`. Accessed: 2013-07-05.

[3] Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., and Avellano, A. The Data Assimilation Research Testbed: A Community Facility. *B. Am. Meteorol. Soc. 90*, 9 (2009), 1283–1296.

[4] Anderson, J. L., and Anderson, S. L. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev. 127*, 12 (2012/09/28 1999), 2741–2758.

[5] Annan, J., and Hargreaves, J. Efficient parameter estimation for a highly chaotic system. *Tellus A 56*, 5 (2004), 520–526.

[6] Annan, J. D., and Hargreaves, J. C. Efficient estimation and ensemble generation in climate modelling. *Phil. Trans. R. Soc. A 365*, 1857 (2007), 2077–2088.

[7] Auvinen, H. *Inversion and assimilation methods with applications in geophysical remote sensing*. PhD thesis, Lappeenranta University of Technology, 2009.

[8] Auvinen, H., Bardsley, J. M., Haario, H., and Kauranne, T. Large-scale Kalman filtering using the limited memory BFGS method. *Electron. Trans. Numer. Anal. 35* (2009), 217–233.

[9] Banks, J., Brooks, J., Cairns, G., Davis, G., and Stacey, P. On Devaney's definition of chaos. *Amer. Math. Mon. 99*, 4 (1992), 332–334.

[10] Bertaux, J. L., Kyrölä, E., Fussen, D., Hauchecorne, A., Dalaudier, F., Sofieva, V., Tamminen, J., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Mangin, A., Blanot, L., Lebrun, J. C., Pérot, K., Fehr, T., Saavedra, L., Leppelmeier, G. W., and Fraisse, R. Global ozone monitoring by occultation of stars: an overview of GOMOS measurements on ENVISAT. *Atmos. Chem. Phys. 10*, 24 (2010), 12091–12148.

[11] Burgers, G., Van Leeuwen, P., and Evensen, G. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev. 126*, 6 (1998), 1719–1724.

[12] Damski, J., Thölix, L., Backman, L., and Taalas, P. FinROSE — middle atmospheric chemistry transport model. *Boreal Env. Res. 12* (2007), 535–550.

[13] Devaney, R. L. *An Introduction to Chaotic Dynamical Systems.* Addison-Wesley, 1989.

[14] Durbin, J., and Koopman, S. *Time series analysis by state space methods.* Oxford university press, New York, 2001.

[15] Evensen, G. *Data assimilation: The ensemble Kalman filter.* Springer, 2007.

[16] Gelman, A., Carlin, J., Stern, H., and Rubin, D. *Bayesian Data Analysis*, second ed. Chapman & Hall, 2003.

[17] Haario, H., Laine, M., Lehtinen, M., Saksman, E., and Tamminen, J. Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. *J. R. Stat. Soc. (B) 66*, 3 (2004), 591–607.

[18] Haario, H., Laine, M., Mira, A., and Saksman, E. DRAM: Efficient adaptive MCMC. *Stat. Comput. 16*, 4 (2006), 339–354.

[19] Haario, H., Saksman, E., and Tamminen, J. An adaptive Metropolis algorithm. *Bernoulli 7*, 2 (2001), 223–242.

[20] Higham, N. *Accuracy and Stability of Numerical Algorithms*, second ed. SIAM, 2002.

[21] Ionides, E. L., Breto, C., and King, A. A. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences 103*, 1 (2006), 18438–18443.

[22] Jackson, C. S., Sen, M. K., Huerta, G., Deng, Y., and Bowman, K. P. Error Reduction and Convergence in Climate Prediction. *J. Climate 21*, 24 (2008), 6698–6709.

[23] Järvinen, H., Räisänen, P., Laine, M., Tamminen, J., Ilin, A., Oja, E., Solonen, A., and Haario, H. Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmos. Chem. Phys. 10*, 20 (2010), 9993–10002.

[24] Kalman, R. E. A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng., Ser. D 82* (1960), 35–42.

[25] Kapitaniak, T. *Chaos for Engineers: Theory, Applications, and Control*, second ed. Springer-Verlag Berlin Heidelberg New York, 2000.

[26] Kitagawa, G. A self-organizing state space model. *J. Amer. Statist. Assoc. 93*, 443 (1998), 1203–1215.

[27] Kyrölä, E., Tamminen, J., Sofieva, V., Bertaux, J. L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton D'Andon, O., Barrot, G., Guirlet, M., Fehr, T., and Saavedra de Miguel, L. GOMOS O_3, NO_2, and NO_3 observations in 2002-2008. *Atmos. Chem. Phys. 10*, 16 (2010), 7723–7738.

[28] Kyrölä, E., Tamminen, J., Sofieva, V., Bertaux, J. L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Guirlet, M., Mangin, A., Blanot, L., Fehr, T., Saavedra de Miguel, L., and Fraisse, R. Retrieval of atmospheric parameters from GOMOS data. *Atmos. Chem. Phys. 10*, 23 (2010), 11881–11903.

[29] Lahoz, W. A., Errera, Q., Swinbank, R., and Fonteyn, D. Data assimilation of stratospheric constituents: a review. *Atmos. Chem. Phys. 7*, 22 (2007), 5745–5773.

[30] Laine, M. *Adaptive MCMC methods with applications in environmental and geophysical models.* PhD thesis, Lappeenranta University of Technology, 2008.

[31] Lorenz, E. N. Deterministic nonperiodic flow. *J. Atmos. Sci. 20*, 2 (2013/07/05 1963), 130–141.

[32] Lorenz, E. N. Predictability: A problem partly solved. *Proceedings of the Seminar on Predicability, European Center on Medium Range Weather Forecasting 1* (1995), 1–18.

[33] Marchand, M., Bekki, S., Hauchecorne, A., and Bertaux, J.-L. Validation of the self-consistency of GOMOS NO_3, NO_2 and O_3 data using chemical data assimilation. *Geophys. Res. Lett. 31*, 10 (2004).

[34] Marchand, M., Bekki, S., Lefèvre, F., and Hauchecorne, A. Temperature retrieval from stratospheric O_3 and NO_3 GOMOS data. *Geophys. Res. Lett. 34*, 24 (2007).

[35] McCullagh, P., and Nelder, J. A. *Generalized Linear Models*, second ed. Chapman & Hall, London, 1989.

[36] McLinden, C. A., and Haley, C. S. Odin/OSIRIS observations of stratospheric NO_3 through sunrise and sunset. *Atmos. Chem. Phys. 8*, 18 (2008), 5529–5534.

[37] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. Equations of state calculations by fast computing machines. *J. Chem. Phys. 21*, 6 (1953), 1087–1092.

[38] Nummelin, E. MC's for MCMC'ists. *Int. Stat. Rev. 70*, 2 (2002), 215–240.

[39] Pedlosky, J. *Geophysical Fluid Dynamics*, second ed. Springer-Verlag, New York, 1987.

[40] Rabier, F., J ärvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A. The ECMWF operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc. 126*, 564 (2000), 1143–1170.

[41] Renard, J., Berthet, G., Brogniez, C., Catoire, V., Fussen, D., Goutail, F., Oelhaf, H., Pommereau, J., Roscoe, H. K., Wetzel, G., Chartier, M., Robert, C., Balois, J., Verwaerde, C., Auriol, F., François, P., Gaubicher, B., and Wursteisen, P. Validation of GOMOS-Envisat vertical profiles of O_3, NO_2, NO_3, and aerosol extinction using balloon-borne instruments and analysis of the retrievals. *J. Geophys. Res. 113*, A12 (2008), A02302.

[42] Robert, C. P., and Casella, G. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[43] Rodgers, C. D. *Inverse methods for atmospheric sounding: theory and practice*. World Scientific, 2000.

[44] Sander, S. P., Friedl, R. R., Golden, D. M., Kurylo, M. J., Moortgat, G. K., Keller-Rudek, H., Wine, P. H., Ravishankara, A. R., Kolb, C. E., Molina, M. J., Finlayson-Pitts, B. J., Huie, R. E., and Orkin, V. L. Chemical kinetics and photochemical data for use in atmospheric studies, evaluation number 15. `http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/41648/1/JPLPub06-2.pdf`. JPL Publication 06-2, Jet Propulsion Laboratory, Pasadena.

[45] Sanders, R. W., Solomon, S., Mount, G. H., Bates, M. W., and Schmeltekopf, A. L. Visible spectroscopy at McMurdo station, Antarctica: 3. observations of NO_3. *J. Geophys. Res. 92*, D7 (1987), 8339–8342.

[46] Särkkä, S. *Recursive Bayesian Inference on Stochastic Differential Equations*. PhD thesis, Helsinki University of Technology, 2006.

[47] Schirber, S., Klocke, D., Pincus, R., Quaas, J., and Anderson, J. L. Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world. *J. Adv. Model. Earth Syst. 5*, 1 (2013), 58–70.

[48] Sexton, D., Murphy, J., Collins, M., and Webb, M. Multivariate prediction using imperfect climate models part I: Outline of methodology. *Clim. Dynam.* (2011).

[49] Singer, H. Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood versus extended kalman filter and Itô–Taylor expansion. *J. Comp. Graph. Stat. 11* (2002), 972–995.

[50] Solonen, A. *Bayesian methods for estimation, optimization and experimental design*. PhD thesis, Lappeenranta University of Technology, 2011.

[51] Solonen, A., Hakkarainen, J., Ilin, A., Abbas, M., and Bibov, A. Tuning the model error covariance matrix in extended Kalman filtering. Submitted to Nonlin. Processes Geophys., 2013.

[52] 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*, 3 (2012), 715–736.

[53] Tamminen, J. *Adaptive Markov Chain Monte Carlo Algorithms with Geophysical Applications*. PhD thesis, University of Helsinki, 2004.

[54] Tamminen, J. Validation of nonlinear inverse algorithms with Markov chain Monte Carlo method. *J. Geophys. Res. 109*, D19 (2004).

[55] Tamminen, J., Kyrölä, E., Sofieva, V. F., Laine, M., Bertaux, J.-L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton-d'Andon, O., Barrot, G., Mangin, A., Guirlet, M., Blanot, L., Fehr, T., Saavedra de Miguel, L., and Fraisse, R. GOMOS data characterisation and error estimation. *Atmos. Chem. Phys. 10*, 19 (2010), 9505–9519.

[56] Tippett, M. K., Anderson, J. L., Bishop, G. H., Hamill, T. M., and Whitaker, J. S. Ensemble Square Root Filters. *Mon. Wea. Rev. 131*, 7 (2003), 1485–1490.

[57] Vanhellemont, F., Fussen, D., and Bingen, C. Global one-step inversion of satellite occultation measurements: A practical method. *J. Geophys. Res. 109*, D9 (2004).

[58] Wilks, D. Effects of stochastic parametrizations in the Lorenz '96 system. *Quart. J. Roy. Meteor. Soc. 131*, 606 (2005), 389–407.

Publ. I

Variational ensemble Kalman filtering using limited memory BFGS

Solonen, A., Haario, H., Hakkarainen, J., Auvinen, H., Amour, I., and Kauranne, T., *Electron. Trans. Numer. Anal. 39* (2012), 271–285.

Electronic Transactions on Numerical Analysis.
Volume 39, pp. 271-285, 2012.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

VARIATIONAL ENSEMBLE KALMAN FILTERING USING LIMITED MEMORY BFGS*

ANTTI SOLONEN†‡, HEIKKI HAARIO‡, JANNE HAKKARAINEN†, HARRI AUVINEN‡, IDRISSA AMOUR‡, AND TUOMO KAURANNE‡

Abstract. The extended Kalman filter (EKF) is one of the most used nonlinear state estimation methods. However, in large-scale problems, the CPU and memory requirements of EKF are prohibitively large. Recently, Auvinen et al. proposed a promising approximation to EKF called the variational Kalman filter (VKF). The implementation of VKF requires the tangent linear and adjoint codes for propagating error covariances in time. However, the trouble of building the codes can be circumvented by using ensemble filtering techniques, where an ensemble of states is propagated in time using the full nonlinear model, and the statistical information needed in EKF formulas is estimated from the ensemble. In this paper, we show how the VKF ideas can be used in the ensemble filtering context. Following VKF, we obtain the state estimate and its covariance by solving a minimization problem using the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) method, which provides low-storage approximations to the state covariances. The resulting hybrid method, the variational ensemble Kalman filter (VEnKF), has several attractive features compared to existing ensemble methods. The model error and observation error covariances can be inserted directly into the minimization problem instead of randomly perturbing model states and observations as in the standard ensemble Kalman filter. New ensembles can be directly generated from the LBFGS covariance approximation without the need of a square root (Cholesky) matrix decomposition. The frequent resampling from the full state space circumvents the problem of ensemble in-breeding frequently associated with ensemble filters. Numerical examples are used to show that the proposed approach performs better than the standard ensemble Kalman filter, especially when the ensemble size is small.

Key words. Data assimilation, state estimation, Kalman filtering, ensemble filtering, LBFGS

AMS subject classifications. 65K10, 15A29

1. Introduction.

Since the introduction of the Kalman filter (KF) in [21] and its nonlinear extension, the extended Kalman filter (EKF), many approaches to overcome the problems with computational complexity present in KF and EKF have been proposed; see, e.g., [6, 9, 13]. In these approaches, the state vector is often projected onto a fixed, low-dimensional subspace. It is known that a fixed projection operator might not correctly capture the dynamics of a nonlinear system; see [14]. In particular, such reduced rank Kalman filters tend to suffer from covariance leakage; see [20].

In [2, 3], a low-storage variational approach to approximate KF and EKF was proposed called the variational Kalman filter (VKF). In VKF, the large matrices in KF formulas are replaced with a low-storage approximation provided by the quasi-Newton optimization method called limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS). The applicability of VKF is hindered by the requirement of tangent linear and adjoint codes for the evolution model, which require a considerable development effort separately for every model.

In ensemble filtering methods, the problems related to standard EKF (large matrices, need for tangent linear and adjoint codes) are circumvented by representing uncertainty in the model state as a number of samples instead of covariance matrices. Instead of moving the covariance in time using the linearized model, uncertainty is propagated simply by moving the ensemble members in time with the full nonlinear evolution model. The simplest version of this idea is the ensemble Kalman filter (EnKF), first proposed in [11], where the covariance matrices in the KF formulas are essentially replaced with sample statistics calculated from

* Received January 24, 2011. Accepted January 22, 2012. Published online on July 9, 2012. Recommended by P. Van Dooren.

† Lappeenranta University of Technology, P.O. Box 20, FI-53851 Lappeenranta, Finland (antti.solonen@gmail.com).

‡ Finnish Meteorological Institute, P.O. Box 503, FI-00101 Helsinki, Finland.

ETNA
Kent State University
http://etna.math.kent.edu

the ensemble. However, EnKF suffers from some problems, e.g., sampling errors due to random perturbation of model state and observations and from ensemble in-breeding that results in a similar covariance leakage as that associated with reduced rank Kalman filters; see [23, 26, 28].

In this paper, we show how VKF ideas can be used in the ensemble filtering context to overcome some problems related to existing ensemble methods. In our approach, the model error and observation error covariances are inserted directly into the minimization problem instead of randomly perturbing model states and observations as in EnKF. New ensembles can be efficiently generated directly from the LBFGS covariance approximation without explicitly constructing the large covariance matrix. We call our hybrid approach the *variational ensemble Kalman filter (VEnKF)* and show by numerical examples that the hybrid method performs well compared to the standard EnKF, especially when the ensemble size is small.

The paper is organized as follows. In Section 2, we recall the basics of Kalman filtering and ensemble methods. We introduce the VEnKF algorithm in Section 3 and demonstrate its performance with numerical examples in Section 4. In Section 5, we discuss some specific topics related to our approach and the differences to existing ensemble filters. Section 6 concludes the paper.

2. Filtering methods.

In this section, we provide an overview of some existing Kalman filtering methods that are related to our approach. We start by recalling how the basic Kalman filter and some of its variants work and continue with an introduction to ensemble filtering methods.

2.1. Kalman filtering and variants.

The Kalman filter can be used to estimate the state $\mathbf{x}_k$ at discrete times k from observations $\mathbf{y}_k$, when the model and observation equations are linear:

$$\begin{aligned}\mathbf{x}_k &= \mathbf{M}_k\mathbf{x}_{k-1} + \varepsilon_k^p\\ \mathbf{y}_k &= \mathbf{K}_k\mathbf{x}_k + \varepsilon_k^o.\end{aligned}$$

In the above system, $\mathbf{M}_k$ is the $d \times d$ evolution model and $\mathbf{K}_k$ is the $m \times d$ observation operator. The $d \times 1$ vector $\mathbf{x}_k$ represents the model state, and the observed data are denoted by the $m \times 1$ vector $\mathbf{y}_k$. The model error ε_k^p and the observation error ε_k^o are assumed to be normally distributed zero mean random variables with covariance matrices $\mathbf{C}_{\varepsilon_k^p}$ and $\mathbf{C}_{\varepsilon_k^o}$, respectively. The Kalman filter algorithm for estimating states and their error covariances can be written as follows.

The Kalman filter algorithm

1. Move the state estimate and covariance in time:
   (a) Compute $\mathbf{x}_k^p = \mathbf{M}_k\mathbf{x}_{k-1}^{\text{est}}$.
   (b) Compute $\mathbf{C}_k^p = \mathbf{M}_k\mathbf{C}_{k-1}^{\text{est}}\mathbf{M}_k^{\mathrm{T}} + \mathbf{C}_{\varepsilon_k^p}$.
2. Combine the prior with observations:
   (a) Compute the Kalman gain $\mathbf{G}_k = \mathbf{C}_k^p\mathbf{K}_k^{\mathrm{T}}(\mathbf{K}_k\mathbf{C}_k^p\mathbf{K}_k^{\mathrm{T}} + \mathbf{C}_{\varepsilon_k^o})^{-1}$.
   (b) Compute the state estimate $\mathbf{x}_k^{\text{est}} = \mathbf{x}_k^p + \mathbf{G}_k(\mathbf{y}_k - \mathbf{K}_k\mathbf{x}_k^p)$.
   (c) Compute the covariance estimate $\mathbf{C}_k^{\text{est}} = \mathbf{C}_k^p - \mathbf{G}_k\mathbf{K}_k\mathbf{C}_k^p$.
3. Set $k \to k+1$ and go to step 1.

The extended Kalman filter directly uses the Kalman filter formulas in the nonlinear case by replacing the nonlinear model and observation operators with appropriate linearizations: $\mathbf{M}_k = \frac{\partial}{\partial \mathbf{x}}\mathcal{M}(\mathbf{x}_{k-1}^{\text{est}})$ and $\mathbf{K}_k = \frac{\partial}{\partial \mathbf{x}}\mathcal{K}(\mathbf{x}_k^p)$.

ETNA
Kent State University
http://etna.math.kent.edu

In the variational formulation of the Kalman filter, the state estimation at step k is viewed as an optimization problem, where a quadratic function

$$l(\mathbf{x}|\mathbf{y}_k) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_k^p\right)^{\mathrm{T}}\left(\mathbf{C}_k^p\right)^{-1}\left(\mathbf{x}-\mathbf{x}_k^p\right) + \frac{1}{2}\left(\mathbf{y}_k-\mathcal{K}(\mathbf{x})\right)^{\mathrm{T}}\left(\mathbf{C}_{\varepsilon_k^o}\right)^{-1}\left(\mathbf{y}_k-\mathcal{K}(\mathbf{x})\right) \tag{2.1}$$

is minimized with respect to $\mathbf{x}$. In the VKF method, introduced in [2], the minimization is done with the LBFGS optimization method, that produces both the state estimate and a low-storage approximation of the covariance (inverse hessian at the minimizer). In the VKF algorithm, the inverse of the prior covariance $\mathbf{C}_k^p$ is also approximated using LBFGS by setting up an auxiliary optimization problem

$$\operatorname*{argmin}_{\mathbf{u}} \frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{C}_k^p\mathbf{u}. \tag{2.2}$$

Thus, the LBFGS optimization routine provides low-storage approximation for both $(\mathbf{C}_k^p)^{-1}$ and $\mathbf{C}_k^{\mathrm{est}}$. All computations with the covariances can be carried out efficiently using the implicit low-storage representation without forming the full matrices. The VKF method is given as an algorithm below.

The variational Kalman filter algorithm

1. Move the state estimate and covariance in time:
   (a) Compute $\mathbf{x}_k^p = \mathbf{M}_k\mathbf{x}_{k-1}^{\mathrm{est}}$.
   (b) Define $\mathbf{C}_k^p = \mathbf{M}_k\mathbf{B}_{k-1}^{\#}\mathbf{M}_k^{\mathrm{T}} + \mathbf{C}_{\varepsilon_k^p}$.
   (c) Apply LBFGS to (2.2) to get an approximation $\mathbf{B}_k^*$ of $(\mathbf{C}_k^p)^{-1}$.
2. Combine the prior with observations:
   (a) Minimize expression (2.1) using LBFGS to get the state estimate $\mathbf{x}_k^{\mathrm{est}}$ and covariance estimate $\mathbf{B}_k^{\#}$.
3. Set $k \to k+1$ and go to step 1.

Note that while VKF can solve the storage problem related to EKF, it requires a way to evolve the covariance in time (step 1(b) in the algorithm above). Propagating the covariance using a direct linearization, as in EKF, is infeasible in high dimensions. In VKF, covariance propagation is done using tangent linear and adjoint codes that implement differentiation at the "code level". This is a standard technique in variational data assimilation; see, e.g., [15, 22]. These codes must be prepared separately for every model and their construction is laborious, although automatic code generators have been recently developed; see, e.g., [8]. In the ensemble filters that we discuss next, tangent linear and adjoint codes are not needed.

2.2. Ensemble filtering. In ensemble filtering, the uncertainty in the state estimate $\mathbf{x}_k$ is represented as N samples, here denoted as $\mathbf{s}_k = (\mathbf{s}_{k,1}, \mathbf{s}_{k,2}, \ldots, \mathbf{s}_{k,N})$, instead of a covariance matrix. The first ensemble filtering method was the ensemble Kalman filter (EnKF) introduced in [11] and implemented in operational numerical weather prediction, e.g., in [19]. The ensemble Kalman filter essentially replaces the state covariance matrices in EKF with the sample covariance calculated from the ensemble. The sample covariance can be written as $\mathrm{Cov}(\mathbf{s}_k) = \mathbf{X}_k\mathbf{X}_k^{\mathrm{T}}$, where

$$\mathbf{X}_k = \frac{1}{\sqrt{N-1}}\left((\mathbf{s}_{k,1}-\overline{\mathbf{s}_k}), (\mathbf{s}_{k,2}-\overline{\mathbf{s}_k}), \ldots, (\mathbf{s}_{k,N}-\overline{\mathbf{s}_k})\right).$$

The sample mean is denoted by $\overline{\mathbf{s}_k}$. Using our notation, the EnKF algorithm can be formulated as follows.

ETNA
Kent State University
http://etna.math.kent.edu

The ensemble Kalman filter algorithm

1. Move the state estimate and covariance in time:
   (a) Move ensemble forward and perturb members with model error: $\mathbf{s}^p_{k,i} = \mathcal{M}(\mathbf{s}^{\text{est}}_{(k-1),i}) + \mathbf{e}^p_{k,i}$, $i = 1, \dots, N$.
   (b) Calculate sample mean $\overline{\mathbf{s}_k}$ and covariance $\mathbf{C}^p_k = \mathbf{X}_k \mathbf{X}_k^{\mathrm{T}}$.
2. Combine the prior with observations:
   (a) Compute the Kalman gain $\mathbf{G}_k$.
   (b) Update ensemble members $\mathbf{s}^{\text{est}}_{k,i} = \mathbf{s}^p_{k,i} + \mathbf{G}_k(\mathbf{y}_k - \mathbf{K}_k \mathbf{s}^p_{k,i} + \mathbf{e}^o_{k,i})$.
   (c) Calculate state estimate as the sample mean $\overline{\mathbf{s}^{\text{est}}_{k,i}}$.

In the above algorithm, the vectors $\mathbf{e}^p_{k,i}$ and $\mathbf{e}^o_{k,i}$ are realizations of the model error and observation error distributions (Gaussians with covariances $\mathbf{C}_{\varepsilon^p_k}$ and $\mathbf{C}_{\varepsilon^o_k}$, respectively).

The ensemble Kalman filter is very simple to implement and it does not require tangent linear and adjoint codes. However, EnKF has various problems and numerous variants have been developed to overcome these issues; see, e.g., [1, 10, 19, 32]. In Section 5, we discuss these variants in light of the proposed VEnKF algorithm, which we introduce in the next section.

3. Variational ensemble Kalman filtering. Here we follow the VKF ideas and show how they can be implemented in the ensemble filtering context, resulting in a filter that we call the variational ensemble Kalman filter (VEnKF). As in VKF, the state estimation in VEnKF is based on minimizing the cost function in Equation (2.1). The prior covariance needed in the cost function is defined here as

$$\mathbf{C}^p_k = \operatorname{Cov}\left(\mathcal{M}(\mathbf{x}^{\text{est}}_{k-1}) + \varepsilon^p_k\right) = \operatorname{Cov}\left(\mathcal{M}(\mathbf{x}^{\text{est}}_{k-1})\right) + \operatorname{Cov}\left(\varepsilon^p_k\right) \approx \mathbf{X}_k \mathbf{X}_k^{\mathrm{T}} + \mathbf{C}_{\varepsilon^p_k}. \tag{3.1}$$

Note that the above formula contains the common assumption that the model error and model response are uncorrelated. The same assumption is made in KF and EKF. In VEnKF, we calculate the sample covariance using the state estimate evolved from the previous time as the expectation instead of the sample mean used in EnKF. Thus, we define

$$\mathbf{X}_k = \frac{1}{\sqrt{N}} \left((\mathbf{s}_{k,1} - \mathbf{x}^p_k), (\mathbf{s}_{k,2} - \mathbf{x}^p_k), \dots, (\mathbf{s}_{k,N} - \mathbf{x}^p_k) \right),$$

where $\mathbf{x}^p_k = \mathcal{M}(\mathbf{x}^{\text{est}}_{k-1})$ and $\mathbf{s}_{k,i} = \mathcal{M}(\mathbf{s}^{\text{est}}_{(k-1),i})$. Note that the ensemble members now do not contain random perturbations; the model error is included directly in Equation (3.1).

The inverse of the prior covariance $\mathbf{C}^p_k = \mathbf{X}_k \mathbf{X}_k^{\mathrm{T}} + \mathbf{C}_{\varepsilon^p_k}$, needed when evaluating the cost function (2.1), can be obtained in two ways. Following the VKF derivation, we can approximate the inverse by applying LBFGS to the artificial optimization problem

$$\operatorname*{argmin}_{\mathbf{u}} \mathbf{u}^{\mathrm{T}} (\mathbf{X}_k \mathbf{X}_k^{\mathrm{T}} + \mathbf{C}_{\varepsilon^p_k}) \mathbf{u}. \tag{3.2}$$

The sample covariance matrix in the above expression naturally does not have to be handled as a full matrix – in order to evaluate the cost function, we can just keep the covariance in "ensemble form" $\mathbf{X}_k \mathbf{X}_k^{\mathrm{T}}$ and evaluate the cost function in the form $\mathbf{u}^{\mathrm{T}} \mathbf{X}_k \mathbf{X}_k^{\mathrm{T}} \mathbf{u} + \mathbf{u}^{\mathrm{T}} \mathbf{C}_{\varepsilon^p_k} \mathbf{u}$. For the computation to remain efficient, we assume (as in VKF) that the model error covariance $\mathbf{C}_{\varepsilon^p_k}$ can be efficiently multiplied with a vector, which is the case, for instance, if the covariance is assumed to be diagonal. As a result of the above optimization, we obtain an LBFGS representation of the inverse of the prior covariance $(\mathbf{C}^p_k)^{-1}$. We can use the LBFGS representation to evaluate the first term when optimizing the cost function (2.1). For computing the matrix-vector product when the matrix is in the LBFGS form, there exists an efficient recursive algorithm; see Appendix A and [27] for details.

ETNA
Kent State University
http://etna.math.kent.edu

An attractive alternative way to proceed is to calculate the inverse of the prior covariance using the Sherman-Morrison-Woodbury (SMW) matrix inversion formula; see, e.g., [18]. The inverse of the prior covariance can be written as

$$\begin{aligned}(\mathbf{C}_k^p)^{-1} &= (\mathbf{X}_k\mathbf{X}_k^{\mathrm{T}} + \mathbf{C}_{\varepsilon_k^p})^{-1} \\ &= \mathbf{C}_{\varepsilon_k^p}^{-1} - \mathbf{C}_{\varepsilon_k^p}^{-1}\mathbf{X}_k(\mathbf{I} + \mathbf{X}_k^{\mathrm{T}}\mathbf{C}_{\varepsilon_k^p}^{-1}\mathbf{X}_k)^{-1}\mathbf{X}_k^{\mathrm{T}}\mathbf{C}_{\varepsilon_k^p}^{-1}.\end{aligned} \tag{3.3}$$

This representation of the inverse can be directly inserted into Equation (2.1) when it it minimized. The computation of the quadratic expression $(\mathbf{x} - \mathbf{x}_k^p)^{\mathrm{T}}(\mathbf{C}_k^p)^{-1}(\mathbf{x} - \mathbf{x}_k^p)$ can be organized so that we do not have to store full matrices of size $d \times d$. With this formulation, the first LBFGS approximation can be avoided and the prior can be included 'exactly' in the second optimization. The application of the formula requires that the inverses of the model error covariances $\mathbf{C}_{\varepsilon_k^p}$ are available. If the model error is assumed to be constant (same for all k), this matrix inversion needs to be computed only once. In addition, the inversion of $\mathbf{I} + \mathbf{X}_k^{\mathrm{T}}\mathbf{C}_{\varepsilon_k^p}^{-1}\mathbf{X}_k$ needs to be computed at every step. However, this matrix is only of size $N \times N$, where N is the ensemble size, which is always very small compared to the dimension of the state space in large-scale applications. In the examples of this paper, we use the SMW formula for inverting the prior covariance.

When the LBFGS optimization is applied to minimize the function (2.1), we get a low-storage approximation for the covariance $\mathbf{C}_k^{\mathrm{est}}$. After that, we sample a new ensemble of state vectors from $\mathrm{N}(\mathbf{x}_k^{\mathrm{est}}, \mathbf{C}_k^{\mathrm{est}})$. Samples can be drawn efficiently, since the LBFGS representation for $\mathbf{C}_k^{\mathrm{est}}$ can be written in the form

$$\mathbf{C}_k^{\mathrm{est}} = \mathbf{B}_0\mathbf{B}_0^{\mathrm{T}} + \sum_{i=1}^{n} \mathbf{b}_i\mathbf{b}_i^{\mathrm{T}},$$

where $\mathbf{B}_0$ is a $d \times d$ matrix and $\mathbf{b}_i$ are $d \times 1$ vectors. From this representation one can produce a zero mean random vector $\mathbf{r} \sim N(\mathbf{0}, \mathbf{C}_k^{\mathrm{est}})$ simply by calculating

$$\mathbf{r} = \mathbf{B}_0\mathbf{z} + \sum_{i=1}^{n} \omega_i\mathbf{b}_i,$$

where $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$ and $\omega_i \sim N(0, 1)$. The matrix $\mathbf{B}_0$ does not have to be stored explicitly, since the product $\mathbf{B}_0\mathbf{z}$ can be computed implicitly using the stored LBFGS vectors. See Appendix A for details about constructing $\mathbf{B}_0$ and $\mathbf{b}_i$.

Finally, we are ready to present our *variational ensemble Kalman filter* (VEnKF) as an algorithm. After setting the initial guesses for the state and its covariance to $\mathbf{x}_0^{\mathrm{est}}$ and $\mathbf{C}_0^{\mathrm{est}}$, respectively, and setting $k = 1$, we write our algorithm as follows:

The VEnKF algorithm

1. Move the ensemble forward and build the prior:
   (a) Compute prior center point $\mathbf{x}_k^p = \mathcal{M}(\mathbf{x}_{k-1}^{\mathrm{est}})$.
   (b) Compute prior ensemble $\mathbf{s}_{k,i}^p = \mathcal{M}(\mathbf{s}_{(k-1),i}^{\mathrm{est}})$, $i = 1, \ldots, N$.
   (c) Define $(\mathbf{C}_k^p)^{-1}$ using SMW formula (3.3)
   (**Alternatively:** apply LBFGS to (3.2) to get $(\mathbf{C}_k^p)^{-1}$).
2. Calculate the posterior estimate and generate the new ensemble:
   (a) Apply LBFGS to minimize (2.1) to get $\mathbf{x}_k^{\mathrm{est}}$ and $\mathbf{C}_k^{\mathrm{est}}$.
   (b) Sample new ensemble $\mathbf{s}_{k,i}^{\mathrm{est}} \sim N(\mathbf{x}_k^{\mathrm{est}}, \mathbf{C}_k^{\mathrm{est}})$; see Appendix A for details.
3. Set $k \to k + 1$ and go to step 1.

ETNA
Kent State University
http://etna.math.kent.edu

Since the optimization tasks are both quadratic, only a quadratic version of the LBFGS method is needed. The LBFGS algorithm is given in Appendix A, for a detailed analysis we refer to, e.g., [27].

The proposed VEnKF algorithm has several attractive features. First of all, it does not suffer from ensemble in-breeding that is often encountered with many ensemble filtering methods, since the new ensemble is re-generated at each step. The ensembles are generated from dynamically changing covariances that are not restricted to any fixed subspace as in reduced rank methods. The ensemble generation is performed directly by the low storage LBFGS construction without the need of any further matrix (Cholesky) decomposition. Moreover, the model error covariance term can be added explicitly to the optimization problem in step 1(c) of the algorithm, whereas in EnKF it is handled by randomly perturbing the prior ensemble with the model error. In addition, VEnKF uses the state estimate given by the optimizer as the expectation in the sample covariance calculations instead of the sample mean. Mainly due to these reasons, VEnKF has a better performance than EnKF in our examples, when the ensemble size is small (see Section 4). Note that VEnKF especially applies to high dimensional problems, since all calculations are performed using the LBFGS covariance representations without handling full covariance matrices.

It is worth noting that the covariance matrices produced by LBFGS (with a diagonal initial inverse Hessian) are full rank and VEnKF is therefore not a "reduced rank" method. Thus, the new ensemble members generated in VEnKF perturb also the directions of small eigenvalues. This is not the case for many other covariance approximation/optimization methods that could be used instead of LBFGS. For example, Lanczos and conjugate gradient methods operate on a low-dimensional subspace.

4. Numerical experiments. In this section, we demonstrate the performance of VEnKF with two synthetic examples, adopted from [2]. The first example is the well-known Lorenz 95 benchmark problem (a low-order nonlinear chaotic ODE system) that shares some characteristics with weather models. The second example is a linear system, where the dimension of the problem can be controlled by changing the discretization.

For comparing methods, we use the root mean square error (rms), written as

$$[\mathrm{rms}]_k = \sqrt{\frac{1}{d}||\mathbf{x}_k^{\mathrm{est}} - \mathbf{x}_k^{\mathrm{true}}||^2},$$

where $\mathbf{x}_k^{\mathrm{est}}$ is the filter estimate and $\mathbf{x}_k^{\mathrm{true}}$ is the truth used in the data generation at iteration k.

4.1. Lorenz 95. In this example, we consider the well-known nonlinear and chaotic Lorenz 95 model introduced in [24] and analyzed in [25]. The model shares many characteristics with realistic atmospheric models and it is often used as a low-order test case for weather forecasting schemes. We use a 40-dimensional version of the model, given as an ODE system

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + 8, \quad i = 1, 2, \ldots, 40.$$

The state variables are periodic: $x_{-1} = x_{39}$, $x_0 = x_{40}$, and $x_{41} = x_1$. Out of the 40 model states, measurements are obtained from 24 states. We define the observation operator (following [2]) as $\mathcal{K}(\mathbf{x}) = \mathbf{K}\mathbf{x}$, where

$$[\mathbf{K}]_{rp} = \begin{cases} 1, & (r,p) \in \{(3j+i, 5j+i+2)\} \\ 0, & \text{otherwise} \end{cases},$$

ETNA
Kent State University
http://etna.math.kent.edu

where $i = 1, 2, 3$ and $j = 0, 1, \ldots, 7$. Thus, we observe the last three states in every set of five. To generate data, we add Gaussian noise to the model solution with zero mean and covariance $(0.15\sigma_{\text{clim}})^2\mathbf{I}$, where $\sigma_{\text{clim}} = 3.641$ ('climatological' standard deviation computed from long model simulations). In the filtering methods, we use $\mathbf{C}_{\varepsilon_k^p} = (0.05\sigma_{\text{clim}})^2\mathbf{I}$ as the model error covariance and $\mathbf{C}_{\varepsilon_k^o} = (0.15\sigma_{\text{clim}})^2\mathbf{I}$ as the observation error covariance. As initial guesses in the filtering, we use $\mathbf{x}_0^{\text{est}} = \mathbf{1}$ and $\mathbf{C}_0^{\text{est}} = \mathbf{I}$. For more details about the example, we refer to [2].

We run experiments with varying ensemble size N and varying number of LBFGS iterations. In Figure 4.1, we compare the performance of EKF, EnKF, and VEnKF with $N = (10, 15, 20, 40)$ in terms of the rms error. Since EnKF and VEnKF are stochastic methods, we display rms errors averaged over 10 repetitions. In VEnKF, the number of LBFGS iterations and the number of LBFGS vectors stored was the same as the ensemble size. From the results it is clear that VEnKF works better when the ensemble size is small. When the ensemble size gets large, the performances of VEnKF and EnKF approach each other.

In Figure 4.2, we compare the forecast skills given by different methods using the same ensemble sizes as above. The forecast skill is here defined as the mean squared difference between the "truth" and the forecast made with the model scaled with σ_{clim}; see [2] for details. Again, VEnKF performs better, especially when N is small. For instance, VEnKF with $N = 10$ performs equally well as EnKF with $N = 20$. Even with larger N, VEnKF is better on average.

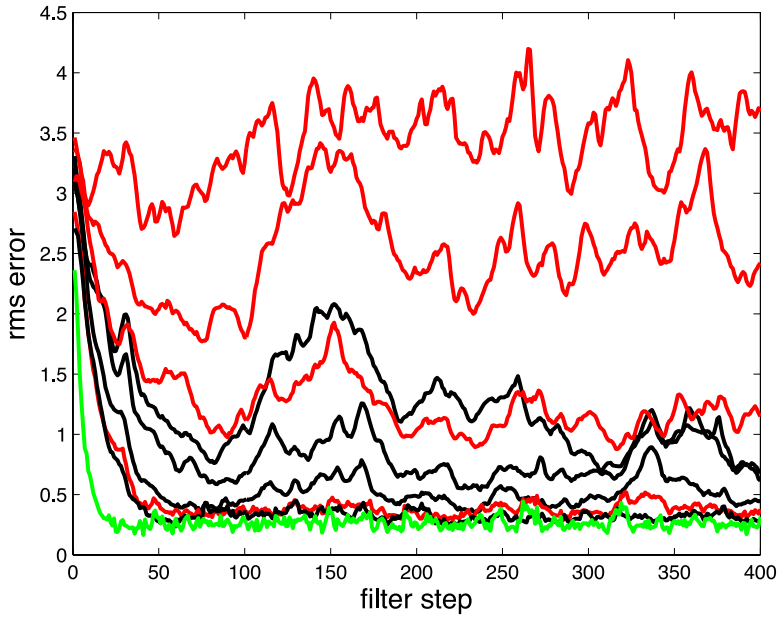


FIG. 4.1. *Comparison of EnKF (red), VEnKF (black), and EKF (green) with ensemble sizes* $N = (10, 15, 20, 40)$ *in the Lorenz 95 example. Increasing ensemble size leads to monotonically decreasing error levels for both EnKF and VEnKF.*

To further demonstrate the behavior of VEnKF, in Figure 4.3 we compare the rms errors (averaged over time) with varying ensemble sizes and varying number of the LBFGS iterations used. As a reference, we also plot the EnKF performance. One can see that 30 LBFGS iterations practically give an as good performance as 40 or 100 iterations, and EnKF starts to produce acceptable results when $N \geq 30$.

ETNA
Kent State University
http://etna.math.kent.edu

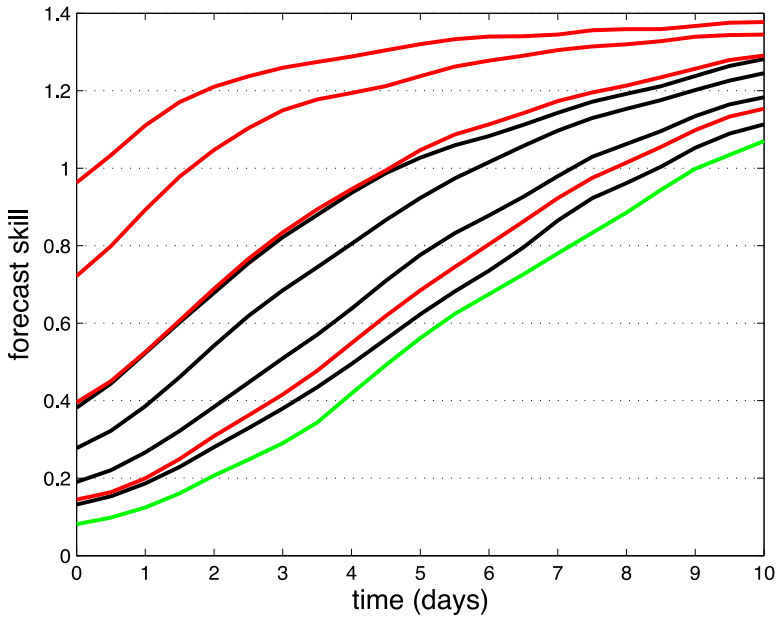


FIG. 4.2. *Forecast skill comparison of EnKF (red), VEnKF (black), and EKF (green) with ensemble sizes* $N = (10, 15, 20, 40)$ *in the Lorenz 95 example. Increasing ensemble size leads to monotonically increasing forecast skill for both EnKF and VEnKF.*

4.2. Heat equation. The purpose of this example, adopted from [2], is to demonstrate the behavior of VEnKF when the dimension is large. The example is linear, so we can directly compare the results with KF. However, as the dimension of the problem is increased, KF cannot be run anymore due to memory issues. Note that while this example does illustrate some computational aspects related to the methods, this system is well-behaved and we cannot conclude much about how the methods work in a high-dimensional chaotic case such as numerical weather prediction, for example.

The model describes heat propagation on a two-dimensional grid and is written as a PDE

$$\frac{\partial x}{\partial t} = -\frac{\partial^2 x}{\partial u^2} - \frac{\partial^2 x}{\partial v} + \alpha \exp\left(-\frac{(u-2/9)^2 + (v-2/9)^2}{\sigma^2}\right),$$

where x is the temperature at coordinates u and v over the domain $\Omega = \{(u, v)|u, v \in [0, 1]\}$. The last term in the equation is an external heat source, whose magnitude can be controlled with the parameter $\alpha \geq 0$.

We discretize the model using a uniform $S \times S$ grid. This leads to a linear forward model $\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k + \mathbf{f}$, where $\mathbf{M}$ corresponds to the heat diffusion and $\mathbf{f}$ to the external forcing; see [2] for details. The dimension of the problem can be controlled by changing S. The observation operator $\mathbf{K}$ is defined as in [2]: the measured temperature is a weighted average of temperatures at neighboring points at $S^2/64$ evenly spaced locations.

The data are generated by adding normally distributed random noise to the model state and the corresponding response:

$$\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k + \mathbf{f} + \mathrm{N}(\mathbf{0}, (0.5\sigma_{ev})^2\mathbf{I})$$
$$\mathbf{y}_{k+1} = \mathbf{K}\mathbf{x}_{k+1} + \mathrm{N}(\mathbf{0}, (0.8\sigma_{obs})^2\mathbf{I}).$$

In the data generation we use $\alpha = 0.75$ and choose σ_{ev} and σ_{obs} so that the signal to noise

ETNA
Kent State University
http://etna.math.kent.edu

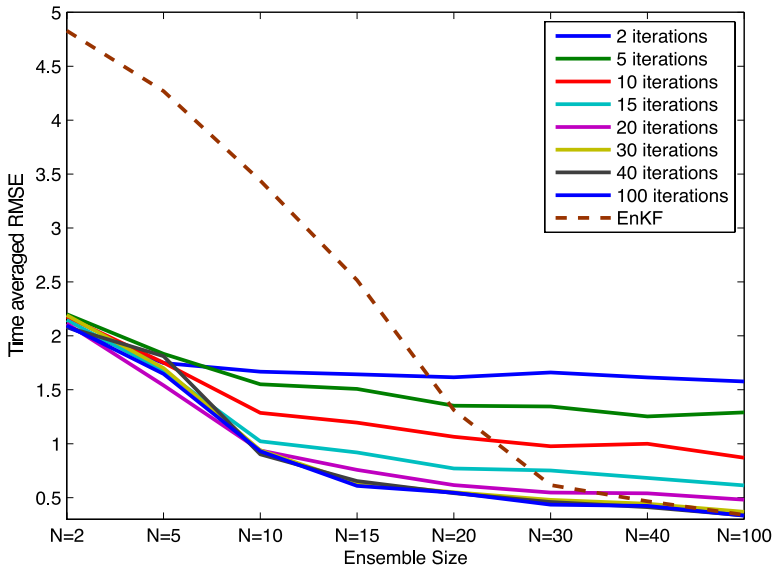


FIG. 4.3. *Rms error averaged over time for EnKF and for VEnKF with varying ensemble sizes and LBFGS iterations in the Lorenz 95 example.*

ratios at the initial condition defined by $\frac{||\mathbf{x}_0||^2}{S^2\sigma_{ev}^2}$ and $\frac{||\mathbf{K}\mathbf{x}_0||^2}{m^2\sigma_{obs}^2}$ are both 50. The initial condition for the data generation is

$$[\mathbf{x}_0]_{ij} = \exp\left(-(u_i - 1/2)^2 - (v_j - 1/2)^2\right).$$

For the filtering we use a biased model, where the forcing term is dropped by setting $\alpha = 0$. The error covariances used for model and observations are $\sigma_{ev}^2\mathbf{I}$ and $\sigma_{obs}^2\mathbf{I}$, respectively. We start all filters from the initial guess $\mathbf{x}_0 = \mathbf{0}$. For ensemble filters, all members are initialized to the same value and for KF we set the initial covariance estimate to $\mathbf{C}_0^{est} = 0$.

In our first test, we take $S = 2^j$ and choose $j = 5$, which is the largest integer so that KF can still be computed on a standard desktop computer. Thus, the dimension of the initial test was $d = S^2 = 1024$. In Figure 4.4, we compare KF, VEnKF, and EnKF using ensemble sizes $N = (5, 10, 20, 50, 100)$ for VEnKF and EnKF. In VEnKF, we always take 20 LBFGS iterations and store 20 LBFGS vectors. The performance of VEnKF approaches that of KF as the ensemble size increases, but EnKF performs poorly: only with larger ensemble sizes we get acceptable results. We think that such a dramatic difference between VEnKF and EnKF is related to the handling of the measurement error in the method. In this case, we have a simple linear dynamic and a rather good guess about the model error, and just a plain "3D-Var" method, where the prior covariance is kept constant, performs rather well. This is the lower limit of the performance of VEnKF upon which we can improve by adding ensemble members. In EnKF, the model error can be added only by perturbing the (few) samples randomly, which can lead to large errors. In addition, in EnKF the state estimate is calculated as the sample mean, whereas VEnKF uses the MAP estimate. In this case this might produce the large errors in EnKF.

Next, we compared VEnKF to EnKF in a case, where the dimension is much higher ($j = 7, d = S^2 = 16384$) using the same ensemble sizes and the same LBFGS settings. In this case, KF cannot be used anymore due to memory issues. As it can be seen in Figure 4.5, the difference between EnKF and VEnKF is even more dramatic in this case: the EnKF performance is poor, whereas VEnKF is able to improve the simple 3D-Var results.

ETNA
Kent State University
http://etna.math.kent.edu

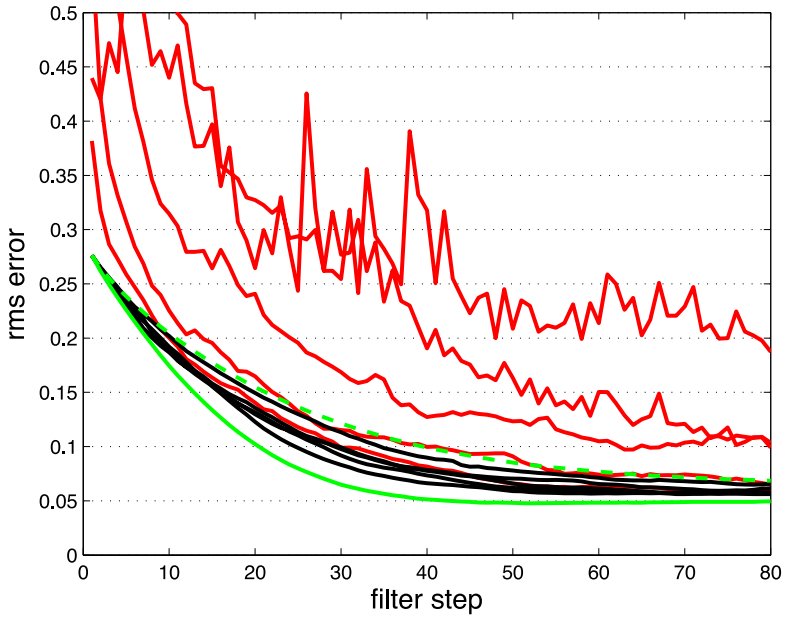


FIG. 4.4. *Performance comparison of KF (solid green), VEnKF (black), and EnKF (red) with different ensemble sizes in the case where* $d = 1024$. *The dashed green line is 3D-Var, where only the fixed model error covariance is used. Increasing ensemble size leads to monotonically decreasing error levels for both EnKF and VEnKF.*

5. Discussion. In the past decade, a wide literature about ensemble filtering has emerged. We start this section by reviewing some of it and discuss the existing approaches in light of our VEnKF method. Later, we discuss some specific topics related to VEnKF.

The standard EnKF is criticized in many papers because of the additional sampling errors brought in by randomly perturbing the observations. In so called square root ensemble filters (SRFs) this is not needed; see, e.g., [12] and the review of SRF methods given in [29]. Similar approaches include the ensemble adjustment Kalman filter (EAKF) and the ensemble transform Kalman filter (ETKF) given in [1] and [5], respectively. In SRF methods, the prior ensemble is deterministically transformed so that the posterior statistics match with the theory in the linear case (Kalman filter equations). This is done basically by writing the Kalman formulas for matrix "square roots" (symmetric decompositions) instead of covariance matrices in a manner that avoids forming full covariance matrices. This transformation is non-unique leading to various SRF formulations; see [29] for a comparison. One difficulty in SRF methods, as pointed out in [29], is the handling of the model error. A simple way is to include it by perturbing the ensemble (as in EnKF), but this potentially leads to sampling errors much in the same way as with perturbed observations in the standard EnKF. Dropping the model error altogether leads to underestimation of errors and possible divergence issues, and different "covariance inflation" mechanisms need to be developed for this purpose. In our VEnKF algorithm, both the model error and the observation error covariances are explicitly present in the minimized cost functions, and these problems do not appear here. Moreover, the square root filters operate, as for the standard EnKF, only in the subspace spanned by the ensemble vectors, whereas the VEnKF samples from the full state space.

Hybrid EnKF methods that incorporate features from both EnKF and variational methods (as VEnKF) have been found appealing in many studies. In [16], a hybrid approach is adopted, which combines elements of ensemble filtering and 3D-Var. However, the hybrid method uses perturbed observations and makes the perfect model assumption; model error

ETNA
Kent State University
http://etna.math.kent.edu

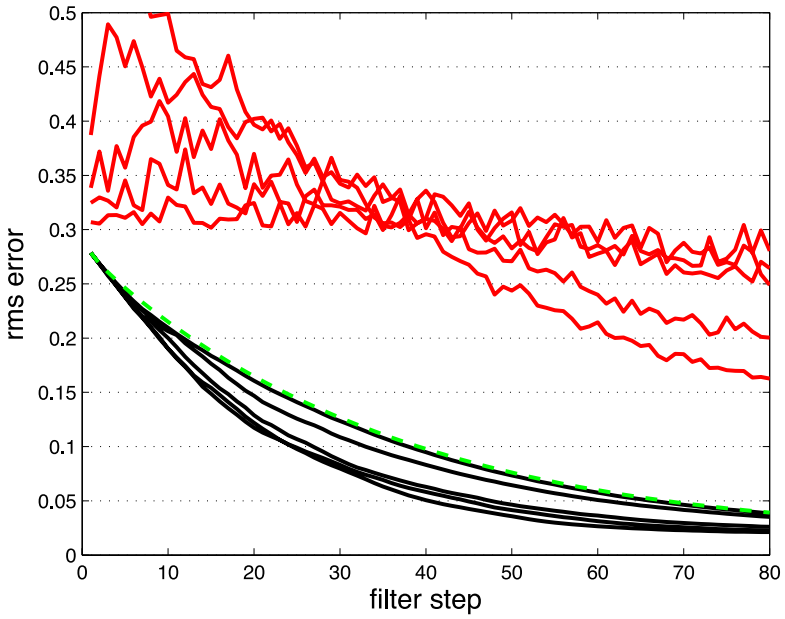


FIG. 4.5. *Performance comparison of VEnKF (black) and EnKF (red) with different ensemble sizes in the case where $d = 16384$. The dashed green line is 3D-Var, where only the fixed model error covariance is used. Increasing ensemble size leads to monotonically decreasing error levels for both EnKF and VEnKF.*

covariance cannot be easily incorporated. The prior covariance is defined as a linear combination of the sample covariance and the static model error covariance used in 3D-Var, which is rather ad-hoc and introduces a tuning parameter (the weight given for the sample covariance). Similar approaches are introduced in [10] and [30]. In [31], the hybrid approach is extended so that it can be used with the 4D-Var method, but the same problems remain.

Probably the closest method to our VEnKF approach is the maximum likelihood ensemble Filter (MLEF) proposed in [32]. In MLEF, an iterative optimization method is used to optimize a 3D-Var type of cost function, and the found optimum is used as the state estimate instead of the sample mean. However, the model error term is neglected and the method operates only in the ensemble subspace.

One criticism that is faced by all Kalman filter based methods are the Gaussian approximations: in all of the methods discussed so far in this paper, a Gaussian form is used for the prior. The ensemble methods are more nonlinear than EKF in the sense that the covariance information is propagated using the nonlinear model instead of the linearized model. Purely nonlinear filtering methods exist as well; see, e.g., [7] for a recent introduction to particle filters. Their benefit over the linear methods can be easily shown in small-dimensional cases, but they become infeasible in large systems. Some nonlinearity in the prior formulation can be induced, e.g., by representing the prior as a mixture of Gaussians or a kernel density estimate fitted to the ensemble; see, e.g., [1, 4] for some discussion. We note that different, nonlinear prior and likelihood formulations can be rather easily incorporated into the cost functions used in the proposed VEnKF method. The only complication is that the cost function is no longer quadratic and one has to use the full LBFGS algorithm instead of the quadratic version used in this paper and to worry about e.g., the Wolfe conditions in the optimization. Extending VEnKF to this type of nonlinear filtering is a topic for further research.

In the VEnKF algorithm, we sample new ensemble members at each iteration. Tradition-

ETNA
Kent State University
http://etna.math.kent.edu

ally, sampling from a multivariate Gaussian distribution is done by performing a symmetric decomposition on the covariance matrix, $\mathbf{C} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$, for example using the Cholesky decomposition, and then producing random vectors as $\mathbf{L}\mathbf{z}$, where $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$. The work in [17] comes close to our approach as it employs the BFGS covariance approximation for the proposal distribution in the Metropolis MCMC algorithm, but using again the Cholesky decomposition. In high dimensions, performing these decompositions is infeasible, since we cannot even store the full covariance matrix. However, as shown in this paper, random sampling can be done directly using the LBFGS vectors. This is potentially useful in other contexts as well, where high dimensional random sampling is needed, for example in the field of inverse problems.

Although we have shown that VEnKF performs well, the approach has its downsides. In particular, the LBFGS optimization is sensitive to certain "parameters" and therefore VEnKF requires some case specific tuning. One tuning parameter is the preconditioner for the inverse Hessian used in the LBFGS optimization. In this paper, we use a heuristic given in [27] that eliminates this tuning parameter (see Appendix A for details). Other tuning parameters remain, such as the choice of the initial guess, the number of LBFGS vectors that we store, and the number of LBFGS iterations that we take. At present, we have not found any general way to define these parameters other than "trial and error".

Naturally, the performance of VEnKF depends on the accuracy of the LBFGS covariance approximations. In [2], the quality of LBFGS covariances was found to be good in low-dimensional numerical examples. However, the LBFGS performance in approximating covariance in high dimensions and realistic data assimilation problems remains a topic of future research.

6. Conclusions. In this paper, we propose a hybrid method called the variational ensemble Kalman filter (VEnKF) for high-dimensional data assimilation that combines elements from ensemble filtering and variational methods. VEnKF is based on the variational Kalman filter (VKF) method [2], where the memory issues related to EKF are solved by low-storage approximations of the state covariances obtained using the LBFGS optimization method. The proposed approach can solve some problems often encountered with ensemble methods, such as sampling errors due to random perturbation of states and observations and ensemble in-breeding. All of the calculations in VEnKF, noting especially the generation of new ensembles, can be made using the LBFGS covariance representation without handling full covariance matrices. Since VEnKF is an ensemble method, it does not need tangent linear and adjoint codes. We show with synthetic examples that the method can perform better than the standard ensemble Kalman filter. Testing the applicability of the method on real data assimilation problems remains a topic of future research.

Acknowledgements. The work was supported by the Centre of Excellence in Inverse Problems of the Academy of Finland.

Appendix A. In this appendix, we give details of some computational issues in VEnKF. First, we recall how the LBFGS optimization algorithm for quadratic minimization problems works. Then, we shown how random samples can be produced from the LBFGS covariance representation.

The LBFGS algorithm for minimizing a quadratic function $q(\mathbf{u}) = \frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u}$, given an initial guess $\mathbf{u}_0$, reads as

LBFGS algorithm for quadratic problems

1. Choose an inverse Hessian approximation $\mathbf{H}_k^0$.
2. Compute the gradient $\mathbf{g}_k = \nabla q(\mathbf{u}_k) = \mathbf{A}\mathbf{u}_k$.

ETNA
Kent State University
http://etna.math.kent.edu

3. Compute a search direction $\mathbf{p}_k = \mathbf{H}_k \mathbf{g}_k$, where $\mathbf{H}_k$ is the LBFGS approximation of the inverse Hessian (details below).
4. Compute step size $\alpha_k = (\mathbf{g}_k^{\mathrm{T}} \mathbf{p}_k)/(\mathbf{p}_k^{\mathrm{T}} \mathbf{A} \mathbf{p}_k)$.
5. Set $\mathbf{u}_{k+1} = \mathbf{u}_k - \alpha_k \mathbf{p}_k$ and $k \to k+1$ and go to step 1.

The LBFGS algorithm uses the BFGS formula for approximating $\mathbf{H}_k$, which is recursively defined as

$$\mathbf{H}_{k+1} = \mathbf{V}_k^{\mathrm{T}} \mathbf{H}_k \mathbf{V}_k + \rho_k \mathbf{s}_k \mathbf{s}_k^{\mathrm{T}},$$

where

$$\begin{aligned} \rho_k &= 1/(\mathbf{y}_k^{\mathrm{T}} \mathbf{s}_k) \\ \mathbf{V}_k &= \mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^{\mathrm{T}} \\ \mathbf{s}_k &= \mathbf{u}_{k+1} - \mathbf{u}_k \\ \mathbf{y}_k &= \mathbf{g}_{k+1} - \mathbf{g}_k. \end{aligned}$$

In LBFGS only a certain number n of the vectors $\mathbf{s}_k$ and $\mathbf{y}_k$ are used in the above formula. Writing out the recursive formula and storing only n most recent vectors leads to the following formula for the LBFGS inverse Hessian approximation:

$$\begin{aligned} \mathbf{H}_k = (\mathbf{V}_{k-1}^{\mathrm{T}} \dots \mathbf{V}_{k-n}^{\mathrm{T}}) \mathbf{H}_k^0 (\mathbf{V}_{k-n} \dots \mathbf{V}_{k-1}) \\ + \rho_{k-n} (\mathbf{V}_{k-1}^{\mathrm{T}} \dots \mathbf{V}_{k-n+1}^{\mathrm{T}}) \mathbf{s}_{k-n} \mathbf{s}_{k-n}^{\mathrm{T}} (\mathbf{V}_{k-n+1} \dots \mathbf{V}_{k-1}) \\ + \rho_{k-n+1} (\mathbf{V}_{k-1}^{\mathrm{T}} \dots \mathbf{V}_{k-n+2}^{\mathrm{T}}) \mathbf{s}_{k-n+1} \mathbf{s}_{k-n+1}^{\mathrm{T}} (\mathbf{V}_{k-n+2} \dots \mathbf{V}_{k-1}) \\ + \dots \\ + \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}_{k-1}^{\mathrm{T}}. \end{aligned}$$

Note that in the LBFGS formulation, the initial inverse Hessian approximation $\mathbf{H}_k^0$ can vary from one iteration to another. One can choose, e.g., a fixed diagonal covariance $\mathbf{H}_k^0 = \gamma \mathbf{I}$, but then the issue of tuning γ emerges. We use a heuristic from [27] in which $\mathbf{H}_k^0 = \gamma_k \mathbf{I}$ and $\gamma_k = (\mathbf{s}_{k-1}^{\mathrm{T}} \mathbf{y}_{k-1})/(\mathbf{y}_{k-1}^{\mathrm{T}} \mathbf{y}_{k-1})$, which attempts to estimate the size of the covariance along the last search direction; see [27].

In our applications, we never want to calculate and store the full inverse Hessian, but to keep it in the above "vector form". There exists an efficient iterative algorithm for computing matrix-vector products with the inverse Hessian needed, e.g., when calculating the search direction in the LBFGS algorithm; see, e.g., [27] for details.

Assuming that the initial inverse Hessian can be decomposed into $\mathbf{H}_k^0 = \mathbf{L}_0 \mathbf{L}_0^{\mathrm{T}}$, the above LBFGS inverse Hessian formula can be written in the form

$$\mathbf{H}_k = \mathbf{B}_0 \mathbf{B}_0^{\mathrm{T}} + \sum_{i=1}^{n} \mathbf{b}_i \mathbf{b}_i^{\mathrm{T}},$$

where

$$\begin{aligned} \mathbf{B}_0 &= (\mathbf{V}_{k-1}^{\mathrm{T}} \dots \mathbf{V}_{k-n}^{\mathrm{T}}) \mathbf{L}_0 \\ \mathbf{b}_1 &= \sqrt{\rho_{k-1}} \mathbf{s}_{k-1} \\ \mathbf{b}_i &= \sqrt{\rho_{k-i}} (\mathbf{V}_{k-1}^{\mathrm{T}} \dots \mathbf{V}_{k-i+1}^{\mathrm{T}}) \mathbf{s}_{k-i}, \quad i = 2, \dots, n. \end{aligned}$$

Note that the square roots $\sqrt{\rho_i}$ can always be calculated, since in the LBFGS algorithm we choose the step length so that $\rho_i \geq 0$ for all i; see [27] for details. Thus, we can sample zero

ETNA
Kent State University
http://etna.math.kent.edu

mean random variables from the covariance $\mathbf{H}_k$ by calculating

$$\mathbf{r} = \mathbf{B}_0\mathbf{z} + \sum_{i=1}^{n} \omega_i \mathbf{b}_i,$$

where $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$ and $\omega_i \sim N(0, 1)$. It is simple to verify that $\mathrm{Cov}(\mathbf{r}) = \mathbf{H}_k$. The needed products can be calculated efficiently without storing full matrices of size $d \times d$. For the first product $\mathbf{q} = \mathbf{B}_0\mathbf{z}$, we can use the iteration

- Set $\mathbf{q} \to \mathbf{L}_0\mathbf{z}$.
- For $i = 1 \ldots n$
  1. $\rho_k = 1/(\mathbf{y}_k^{\mathrm{T}}\mathbf{s}_k)$.
  2. Set $\mathbf{q} \to \mathbf{q} - (\rho_k \mathbf{y}_k^{\mathrm{T}}\mathbf{q})\mathbf{s}_k$.
- Return $\mathbf{q}$.

For the other products $\omega_i\mathbf{b}_i$ we can first calculate the vectors $\mathbf{b}_i$ explicitly using a similar loop as above and then take sums of the vectors, weighted by random numbers ω_i.

REFERENCES

[1] J. L. ANDERSON, *An ensemble adjustment Kalman filter for data assimilation*, Mon. Wea. Rev., 129 (2001), pp. 2884–2903.

[2] H. AUVINEN, J. M. BARDSLEY, H. HAARIO, AND T. KAURANNE, *The variational Kalman filter and an efficient implementation using limited memory BFGS*, Internat. J. Numer. Methods Fluids, 64 (2009), pp. 314–335.

[3] ———, *Large-scale Kalman filtering using the limited memory BFGS method*, Electron. Trans. Numer. Anal., 35 (2009), pp. 217–233.
http://etna.math.kent.edu/vol.35.2009/pp217-233.dir

[4] T. BENGTSSON, C. SNYDER, AND D. NYCHKA, *Toward a nonlinear ensemble filter for high-dimensional systems*, J. Geophys. Res., 108 (2003), 8775 (10 pages).

[5] C. H. BISHOP, B. J. ETHERTON, AND S. J. MAJUMDAR, *Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects*, Mon. Wea. Rev, 129 (2001), pp. 420–436.

[6] M. A. CANE, R. N. MILLER, B. TANG, E. C. HACKERT, AND A. J. BUSALACCHI, *Mapping tropical Pacific sea level: data assimilation via reduced state Kalman filter*, J. Geophys. Res., 101 (1996), pp. 22599–22617.

[7] O. CAPPE, S. GODSILL, AND E. MOULINES, *An overview of existing methods and recent advances in sequential Monte Carlo*, Proc. IEEE, 95 (2007), pp. 899–924.

[8] Q. CHENG, J. CAO, B. WANG, AND H. ZHANG, *Adjoint code generator*, Sci. China Ser. F, 52 (2009), pp. 926–941.

[9] D. P. DEE, *Simplification of the Kalman filter for meteorological data assimilation*, Q. J. R. Meteorol. Soc., 117 (1990), pp. 365–384.

[10] B. J. ETHERTON AND C. H. BISHOP, *Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error*, Mon. Wea. Rev., 132 (2004), pp. 1065–1080.

[11] G. EVENSEN, *Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res., 99 (1994), pp. 10143–10162.

[12] ———, *Sampling strategies and square root analysis schemes for the EnKF*, Ocean Dynam., 54 (2004), pp. 539–560.

[13] M. FISHER, *Development of a simplified Kalman filter*, ECMWF Technical Memorandum n. 260 (1998), European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2-9AX, UK.

[14] M. FISHER AND E. ANDERSSON, *Developments in 4D-var and Kalman filtering*, ECMWF Technical Memorandum n. 347 (2001), European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2-9AX, UK.

[15] R. GIERING AND T. KAMINSKI, *Recipes for adjoint code construction*, ACM Trans. Math. Software, 24 (1998), pp. 437–474.

[16] T. M. HAMILL AND C. SNYDER, *A hybrid ensemble Kalman filter-3D variational analysis scheme*, Mon. Wea. Rev., 128 (2000), pp. 2905–2919.

[17] K. M. HANSON AND G. CUNNINGHAM, *Posterior sampling with improved efficiency*, in Medical Imaging 1998: Image Processing, K. M. Hanson, ed., Proceedings of SPIE, 3338, SPIE, Bellingham WA, 1998, pp. 371–382.

[18] N. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.

ETNA
Kent State University
http://etna.math.kent.edu

[19] P. HOUTEKAMER AND H. L. MITCHELL, *Data assimilation using an ensemble Kalman filter technique*, Mon. Wea. Rev., 126 (1998), pp. 796–811.
[20] L. ISAKSEN, M. FISHER, E. ANDERSSON, AND J. BARKMEIJER, *The structure and realism of sensitivity perturbations and their interpretation as 'Key Analysis Errors'*, Q. J. R. Meteorol. Soc., 131 (2005), pp. 3053–3078.
[21] R. E. KALMAN, *A new approach to linear filtering and prediction problems*, Trans. ASME Ser. D J. Bas. Engrg., 82 (1960), pp. 35–45.
[22] F. X. LE DIMET AND O. TALAGRAND, *Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects*, Tellus Ser. A., 38 (1986), pp. 97–110.
[23] H. LI, E. KALNAY, AND T. MIYOSHI, *Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter*, Q. J. R. Meteorol. Soc., 135 (2009), pp. 523–533.
[24] E. N. LORENZ, *Predictability: a problem partly solved*, in Proceedings of the Seminar on Predictability, Vol. 1, T. Palmer, ed., ECMWF, Reading, UK, (1996), pp. 1–18.
[25] E. N. LORENZ AND K. A. EMANUEL, *Optimal sites for supplementary weather observations: simulation with a small model*, J. Atmospheric Sci., 55 (1998), pp. 399–414.
[26] Z. MENG AND F. ZHANG, *Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: imperfect model experiments*, Mon. Wea. Rev., 135 (2007), pp. 1403–1423.
[27] J. NOCEDAL AND S. WRIGHT, *Numerical Optimization*, Springer, Berlin, 1999.
[28] W. SACHER AND P. BARTELLO, *Sampling errors in ensemble Kalman filtering. Part I: theory*, Mon. Wea. Rev., 136 (2008), pp. 3035–3049.
[29] M. K. TIPPETT, J. L. ANDERSON, C. H. BISHOP, T. M. HAMILL, AND J. S. WHITAKER, *Ensemble square root filters*, Mon. Wea. Rev., 131 (2003), pp. 1485–1490.
[30] X. WANG, D. M. BARKER, C. SNYDER, AND T. M. HAMILL, *A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment*, Mon. Wea. Rev., 136 (2008), pp. 5116–5131.
[31] F. ZHANG, M. ZHANG, AND J. A. HANSEN, *Coupling ensemble Kalman filter with four-dimensional data assimilation*, Adv. Atmospher. Sci., 26 (2009), pp. 1–8.
[32] M. ZUPANSKI, *Maximum likelihood ensemble filter: theoretical aspect*, Mon. Wea. Rev., 133 (2005), pp. 1710–1726.

Publ. II

On closure parameter estimation in chaotic systems

Hakkarainen, J., Ilin, A., Solonen, A., Laine, M., Haario, H., Tamminen, J., Oja, E., and Järvinen, H., *Nonlin. Processes Geophys. 19*, 1 (2012), 127–143, doi:10.5194/npg-19-127-2012.

Nonlin. Processes Geophys., 19, 127–143, 2012
www.nonlin-processes-geophys.net/19/127/2012/
doi:10.5194/npg-19-127-2012

Nonlinear Processes in Geophysics

On closure parameter estimation in chaotic systems

J. Hakkarainen[1,2,*], A. Ilin[3,*], A. Solonen[1,2,*], M. Laine[1], H. Haario[2], J. Tamminen[1], E. Oja[3], and H. Järvinen[1]

[1]Finnish Meteorological Institute, Helsinki, Finland
[2]Lappeenranta University of Technology, Lappeenranta, Finland
[3]Aalto University School of Science, Espoo, Finland
[*]These authors contributed equally to this work.

Correspondence to: J. Hakkarainen (janne.hakkarainen@fmi.fi)

Received: 21 November 2011 – Revised: 3 February 2012 – Accepted: 7 February 2012 – Published: 15 February 2012

Abstract. Many dynamical models, such as numerical weather prediction and climate models, contain so called closure parameters. These parameters usually appear in physical parameterizations of sub-grid scale processes, and they act as "tuning handles" of the models. Currently, the values of these parameters are specified mostly manually, but the increasing complexity of the models calls for more algorithmic ways to perform the tuning. Traditionally, parameters of dynamical systems are estimated by directly comparing the model simulations to observed data using, for instance, a least squares approach. However, if the models are chaotic, the classical approach can be ineffective, since small errors in the initial conditions can lead to large, unpredictable deviations from the observations. In this paper, we study numerical methods available for estimating closure parameters in chaotic models. We discuss three techniques: off-line likelihood calculations using filtering methods, the state augmentation method, and the approach that utilizes summary statistics from long model simulations. The properties of the methods are studied using a modified version of the Lorenz 95 system, where the effect of fast variables are described using a simple parameterization.

1 Introduction

Many dynamical models in atmospheric sciences contain so called closure parameters. These parameters are usually connected to processes that occur on smaller and faster scales than the model discretization allows. For instance, the computational grid used in modern climate and numerical weather prediction (NWP) models is too coarse to directly model cloud micro-physics and many cloud-related phenomena are therefore represented by parameterization schemes. For example, consider the cloud shortwave optical properties, which are related to the cloud liquid water amount, and affect the shortwave radiation fluxes in atmospheric models. These properties can be specified with parameters, such as the mean effective radius of cloud water droplets (Martin et al., 1994).

The closure parameters act as "tuning handles" of the models. Parameter tuning is particularly necessary whenever new and improved parameterized processes are implemented into the models. Currently, the parameters are usually predefined by experts using a relatively small number of model simulations. This tuning procedure is somewhat subjective and therefore open for criticism.

In this paper, we discuss different algorithmic ways to estimate the tuning parameters, that is, how to find the optimal closure parameters by fitting the model to available observations. While this problem has not been studied much, there are a few recent papers that address the problem in the context of climate modeling (Jackson et al., 2008; Villagran et al., 2008; Järvinen et al., 2010; Sexton et al., 2011). Numerical weather prediction (NWP) is considered to be more of an initial value problem than a parameter estimation problem (Annan and Hargreaves, 2007) and tuning of closure parameters is in general done manually by using samples of test forecasts. In a recently proposed approach, NWP parameter estimation is embedded into ensemble prediction systems (Järvinen et al., 2011; Laine et al., 2011).

While the motivation behind this work is the closure parameter estimation problem in atmospheric models, we note that similar parameterizations appear in many multi-scale models in computational fluid dynamics. The parameter estimation is complicated by the fact that these models are often chaotic, which means that a small change in the initial conditions can lead to a completely different simulated trajectory. Therefore, the classical parameter estimation approaches that are based on directly comparing model simulations and observations using, for instance, a least squares approach, may be inefficient.

To fix the notation, let us assume for simplicity that a dynamical model can be described by a discrete state space model

$$\boldsymbol{x}_k = \mathcal{M}(\boldsymbol{x}_{k-1}, \boldsymbol{\theta}) \quad (1)$$

$$\boldsymbol{z}_k = \mathcal{K}(\boldsymbol{x}_k), \quad (2)$$

where $\boldsymbol{x}$ denotes the state of the system, the model operator $\mathcal{M}$ solves the equations that describe the dynamics of the system, k is the index of the time, $\boldsymbol{z}$ are the variables that can be observed, $\mathcal{K}$ is the observation operator and $\boldsymbol{\theta}$ denotes the (closure) parameters. The model operator $\mathcal{M}$ is assumed to contain everything that is needed to simulate the system, including also as external forcing terms and boundary conditions. In the real-world setting, we would like to tune parameters $\boldsymbol{\theta}$ of the model in Eqs. (1)–(2) using a set of available observations $\boldsymbol{y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ taken at some time instances $\{t_1, \ldots, t_n\}$. Note that $\boldsymbol{y}$ are measured values, while $\boldsymbol{z}$ are simulated values of the same variables.

In parameter estimation, we follow the Bayesian methodology, in which the knowledge about the unknown parameters is inferred from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{\theta})\, p(\boldsymbol{y}|\boldsymbol{\theta}), \quad (3)$$

which is evaluated using the prior $p(\boldsymbol{\theta})$ and the likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$. The likelihood function specifies how plausible the observed data are given model parameter values. Therefore, defining a proper likelihood function is the central problem in parameter estimation. The prior contains the information that we have about the parameters based on the accumulated information from the past. For an introduction to Bayesian estimation, see, for example, the book by Gelman et al. (2003).

Traditionally, parameters of dynamical systems are estimated by comparing model simulations to observed data using a measure such as a sum of squared differences between $\boldsymbol{z}$ and $\boldsymbol{y}$. This corresponds to the assumption that the observations are noisy realizations of the model values. The problem in applying these techniques directly to chaotic systems is that the dynamically changing model state $\boldsymbol{x}$ is not known exactly, and small errors in the state estimates can grow in an unpredictable manner, making direct comparisons of model simulations and observations meaningless over long time periods.

In this paper, we consider three ways to estimate the closure parameters of chaotic models. In the first approach, observations and model simulations are summarized in the form of statistics, which are typically some temporal and spatial averages of the data. The likelihood model is constructed in terms of the summary statistics such that model parameters producing statistics that are closer to the observed statistics would have higher likelihood. This kind of an approach is employed in climate model parameter estimation in several recent studies (Jackson et al., 2008; Järvinen et al., 2010; Sexton et al., 2011). In the summary statistics approach, the problem of chaotic behavior can be alleviated, since the statistics computed from long simulations are less dependent on the initial conditions than the specific values of the state variables.

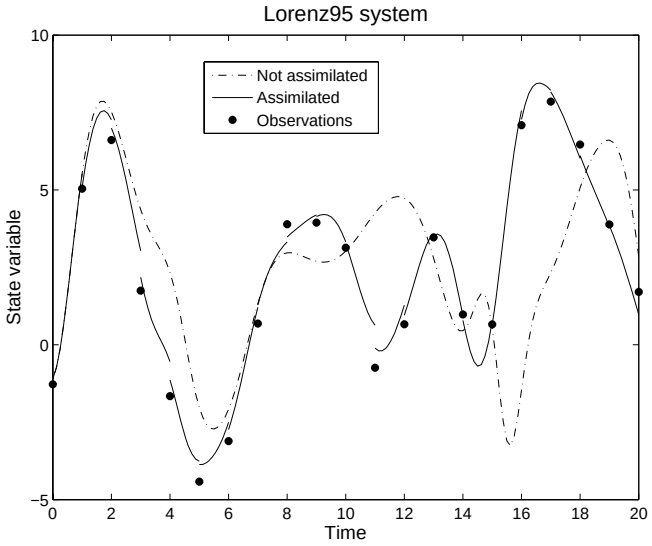


Fig. 1. An illustration of sequential data assimilation in a chaotic system. After some time the control run, even with optimal parameter values, gets "off track" due to chaos. Data assimilation keeps the model in the same trajectory with the data.

The other two approaches are based on embedding the parameter estimation techniques into dynamical state estimation (data assimilation) methods that constantly update the model state as new observations become available. Thus, the model is kept in the vicinity of the data, and the problems caused by chaotic behavior can be alleviated. This is illustrated in Fig. 1 by running the Lorenz system – that is used for experimentation in Sect. 5 – two times from the same initial values, with and without data assimilation. One can see that the model run without assimilation eventually deviates from the trajectory of the observations.

We consider two ways to implement parameter estimation within a data assimilation system. In the *state augmentation* approach (see Sect. 4), the model parameters are treated as artificial states and assimilated together with the actual model state (see, e.g. Kitagawa, 1998; Ionides et al., 2006; Dowd, 2011). In the *likelihood* approach, detailed in Sect. 3, the likelihood of a parameter value is evaluated by running a state estimation method over a chosen data set, keeping the parameter value fixed. The likelihood is constructed using the filter residuals (the squared differences between the observations and the short-range forecasts), see Fig. 1. This resembles classical parameter estimation, but the uncertainty in the model state is "integrated out" using a state estimation technique. The problem of chaoticity is circumvented by computing the likelihood components from short simulations, where chaotic behavior does not yet appear. The likelihood approach is a standard technique in

parameter estimation of stochastic models (see, e.g. Kivman, 2003; Singer, 2002; Dowd, 2011), but less studied in connection with deterministic, chaotic systems.

The paper is organized as follows. In Sect. 2, we present the methodology for the summary statistics approach. The likelihood approach is discussed in Sect. 3, and the sequential parameter estimation via state augmentation is presented in Sect. 4. In Sect. 5, the case study setup and numerical results are presented. In Sect. 6 we discuss some specific issues related to the properties of the methods. Section 7 concludes the paper.

2 Likelihood based on summary statistics

Several recent studies (Jackson et al., 2008; Järvinen et al., 2010; Sexton et al., 2011) on parameter estimation in climate models formulated the likelihood in terms of summary statistics. The advantage of this approach is that it is computationally feasible and rather straightforward to implement. It avoids the problem of chaotic behavior as in sufficiently long simulations the effect of the initial values diminishes.

In this approach, the observations are transformed to a set of summary statistics $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{y}_{1:n})$, where $\boldsymbol{y}_{1:n}$ denotes all observations for n steps of the simulation model Eqs. (1)–(2). The posterior distribution of model parameters $\boldsymbol{\theta}$ is evaluated as

$$p(\boldsymbol{\theta}|\boldsymbol{s}) \propto p(\boldsymbol{\theta})\, p(\boldsymbol{s}|\boldsymbol{\theta}). \tag{4}$$

The likelihood $p(\boldsymbol{s}|\boldsymbol{\theta})$ is constructed so that models $\boldsymbol{\theta}$ producing summary statistics $\boldsymbol{s}_{\boldsymbol{\theta}}$ which are close to the observed values $\boldsymbol{s}$ get higher probability. Here, $\boldsymbol{s}_{\boldsymbol{\theta}} = \boldsymbol{s}(\boldsymbol{z}_{\boldsymbol{\theta},1:n})$ denotes the summary statistics computed from data $\boldsymbol{z}_{\boldsymbol{\theta},1:n}$ simulated for n steps with model parameters $\boldsymbol{\theta}$. The approach is related to approximate Bayesian computation (ABC, see, e.g. Cornuet et al., 2008), where summary statistics are used to do statistical inference in situations where the exact likelihood is intractable.

2.1 Matching observed and simulated statistics

When the simulated and observed summary statistics are directly matched, the likelihood can be formulated, for instance, as

$$p(\boldsymbol{s}|\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2} C(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{\theta}})\right), \tag{5}$$

where $C(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{\theta}})$ is a cost function penalizing the misfit between $\boldsymbol{s}$ and $\boldsymbol{s}_{\boldsymbol{\theta}}$. For example, one could use the Gaussian assumption yielding the form

$$C(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{\theta}}) = (\boldsymbol{s} - \boldsymbol{s}_{\boldsymbol{\theta}})^T \Sigma^{-1} (\boldsymbol{s} - \boldsymbol{s}_{\boldsymbol{\theta}}), \tag{6}$$

where the covariance matrix Σ takes into account possible correlations between the summary statistics. When some of the correlations are ignored (Eq. 6) becomes a sum of multiple terms. For example, in Järvinen et al. (2010) the cost function was similar to

$$C(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{\theta}}) = (s^{\mathrm{g}} - s^{\mathrm{g}}_{\boldsymbol{\theta}})^2 / \sigma^2_{\mathrm{g}} + \sum_{t=1}^{12} \sum_{i,k} \left(\boldsymbol{s}^{ikt} - \boldsymbol{s}^{ikt}_{\boldsymbol{\theta}}\right)^T \Sigma^{-1}_{ikt} \left(\boldsymbol{s}^{ikt} - \boldsymbol{s}^{ikt}_{\boldsymbol{\theta}}\right), \tag{7}$$

where s^{g} is the annual global mean of the net radiative flux at the top of the atmosphere (TOA) and $\boldsymbol{s}^{ikt}$ are zonal and monthly averages of the k-th variable computed for latitude i and month t. The first term in Eq. (7) penalizes unstable models which have an unrealistic balance of the global-mean TOA radiation, whereas the second term ensures a realistic annual cycle for the radiation. The same statistics computed from simulated data are denoted by $s^{\mathrm{g}}_{\boldsymbol{\theta}}$ and $\boldsymbol{s}^{ikt}_{\boldsymbol{\theta}}$.

The goal of the studies by Järvinen et al. (2010) was to explore the uncertainty of the parameters which have a large effect on the radiative balance and therefore only the net radiative flux at TOA was used to compute the zonal and monthly averages $\mathbf{s}^{ikt}$ in Eq. (7). In Jackson et al. (2008), several variables were included in the cost function and the covariance matrix Σ^{ikt} was formulated in terms of a few leading empirical orthogonal functions (EOFs).

One problem of the direct matching of observed and simulated statistics is that the resulting likelihood Eq. (5) may not be a smooth function of the parameters, as will be seen in the experimental results (e.g. Fig. 2). A possible reason for this are the random effects caused by the finite length of the simulations. The noise in the objective function may complicate the parameter estimation procedure. Another problem is that the matching approach is not based on a well justified statistical model for the summary statistics: it is not easy to define what values for the summary statistics are "good" in the statistical sense. For example, it is not straightforward to select the scaling parameter Σ_{ikt} in Eq. (7).

2.2 Fitting a probabilistic model for summary statistics

The problems mentioned above can be partly overcome by building a probabilistic model for the summary statistics. The summary statistics are treated as random variables which are systematically affected by varying the model parameters $\boldsymbol{\theta}$:

$$\boldsymbol{s}_{\boldsymbol{\theta}} = \mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\epsilon}, \tag{8}$$

where function $\mathbf{f}$ is often called an emulator or a surrogate model (Rougier, 2008; Sexton et al., 2011) and the noise term $\boldsymbol{\epsilon}$ can be assumed Gaussian with zero mean and covariance matrix Σ. The emulator and the noise model can be estimated from training samples which are pairs $\{\boldsymbol{s}_{\boldsymbol{\theta}_i}, \boldsymbol{\theta}_i\}$ of simulated statistics $\boldsymbol{s}_{\boldsymbol{\theta}_i}$ and parameter values $\boldsymbol{\theta}_i$ used in the simulations. This is a nonlinear regression problem which have been studied intensively in statistics and machine learning (see, e.g. Bishop, 2006). Examples of parametric models

for **f** include polynomials, radial basis function networks and multi-layer perceptrons.

Thus, prior to constructing the likelihood, the model has to be simulated many times with different parameter values selected over a suitable range. This is computationally expensive, but a necessary step. The likelihood is then constructed as

$$\boldsymbol{s}|\boldsymbol{\theta} \sim N(\mathbf{f}(\boldsymbol{\theta}), \Sigma). \tag{9}$$

One difficulty of learning Eq. (8) is that the set of all possible statistics in $\boldsymbol{s}_{\boldsymbol{\theta}}$ is highly multidimensional while the number of examples to train the emulator Eq. (8) is potentially limited because of the high computational costs of model simulations. This problem can be solved by reducing the number of summary statistics somehow before training the model Eq. (8). The simplest way is to consider only a linear combination of the summary statistics, which means neglecting the variability of summary statistics outside a selected subspace. Thus, Eq. (8) is replaced by

$$\mathbf{W}\boldsymbol{s} = \mathbf{f}_*(\boldsymbol{\theta}) + \boldsymbol{\epsilon}_*, \tag{10}$$

where **W** is a properly chosen matrix and the likelihood is formulated in terms of the projected data:

$$\mathbf{W}\boldsymbol{s}|\boldsymbol{\theta} \sim N(\mathbf{f}_*(\boldsymbol{\theta}), \Sigma_*). \tag{11}$$

Sexton et al. (2011), computed **W** using principal component analysis (PCA) and the dimensionality of the summary statistics was reduced from 175 000 to only six. Thus, the criterion for dimensionality reduction used there was the maximum amount of variance retained in the projected statistics. Another possible approach is to find projections $\mathbf{W}\boldsymbol{s}$ which are most informative about closure parameters $\boldsymbol{\theta}$. For example, canonical correlation analysis is mentioned as a more appropriate dimensionality method by Rougier (2008). In the experiments presented in Sect. 5.2.1, we find **W** by fitting a linear regression model $\boldsymbol{\theta} \approx \mathbf{W}\boldsymbol{s}_{\boldsymbol{\theta}}$ to the training data $\{z_{\boldsymbol{\theta}_i}, \boldsymbol{\theta}_i\}$. The emulator (10) is then estimated for the projected statistics. Note that one can analyze the elements of **W** computed to maximize correlations between $\boldsymbol{\theta}$ and $\boldsymbol{s}_{\boldsymbol{\theta}}$ in order to have a clue on which summary statistics are affected by varying the closure parameters.

3 Likelihood with filtering methods

Traditionally, the likelihood for a parameter in non-chaotic dynamical models is calculated by comparing the model to data using a goodness-of-fit measure, such as the sum of squared differences between the model and the observations. In many cases, however, the model state is not known accurately and it has to be estimated together with the parameters. This is especially important with chaotic models, where small errors in the model state can grow quickly when the model is integrated in time.

State estimation in dynamical models can be carried out using filtering methods, where the distribution of the model state is evolved with the dynamical model and sequentially updated as new observations become available. When static parameters are estimated, filtering can be used to "keep the model on track" with the measurements. In this section, we present how the likelihood in chaotic models can be computed using filtering methods. First, we present the general filtering formulas and then consider the special case of Gaussian filters.

3.1 General formulas

Let us consider the following discrete state space model at time step k with unknown parameters $\boldsymbol{\theta}$:

$$\boldsymbol{x}_k \sim p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{\theta}) \tag{12}$$

$$\boldsymbol{y}_k \sim p(\boldsymbol{y}_k|\boldsymbol{x}_k) \tag{13}$$

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta}). \tag{14}$$

Thus, in addition to the unknown, dynamically changing model state $\boldsymbol{x}_k$, we have static parameters $\boldsymbol{\theta}$, from which we have some prior information $p(\boldsymbol{\theta})$. As mentioned in the introduction, the goal in parameter estimation, in Bayesian terms, is to find the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:n})$ of the parameters, given a fixed data set $\boldsymbol{y}_{1:n}$. Here, the notation $\boldsymbol{y}_{1:n} = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_n\}$ means all observations for n time steps.

Filtering methods (particle filter, Kalman filters, etc.) estimate the dynamically changing model state sequentially. They give the marginal distribution of the state given the measurements obtained until the current time k. Thus, for a given value for $\boldsymbol{\theta}$, filtering methods estimate $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}, \boldsymbol{\theta})$.

Filters work by iterating two steps: prediction and update. In the prediction step, the current distribution of the state is evolved with the dynamical model to the next time step:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{\theta}) \times p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}) d\boldsymbol{x}_{k-1}. \tag{15}$$

When the new observation $\boldsymbol{y}_k$ is obtained, the model state is updated using the Bayes' rule with the predictive distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta})$ as the prior:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}, \boldsymbol{\theta}) \propto p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}). \tag{16}$$

This posterior is used inside the integral (15) to obtain the prior for the next time step.

Using the state posteriors obtained in the filtering method, it is also possible to compute the predictive distribution of the next observation. For observation $\boldsymbol{y}_k$, the predictive distribution, given all previous observations, can be written as

$$p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}) = \int p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}) d\boldsymbol{x}_k. \tag{17}$$

The latter term $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta})$ in the integral is the predictive distribution given by Eq. (15).

Let us now proceed to the original task of estimating static parameters $\boldsymbol{\theta}$ from observations $\boldsymbol{y}_{1:n}$, i.e., computing the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:n})$. Applying the Bayes' formula and the chain rule for joint probability, we obtain

$$p(\boldsymbol{\theta}|\boldsymbol{y}_{1:n}) \propto p(\boldsymbol{y}_{1:n}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (18)$$

$$= p(\boldsymbol{y}_n|\boldsymbol{y}_{1:n-1},\boldsymbol{\theta})p(\boldsymbol{y}_{n-1}|\boldsymbol{y}_{1:n-2},\boldsymbol{\theta})\ldots \times p(\boldsymbol{y}_2|\boldsymbol{y}_1,\boldsymbol{\theta})p(\boldsymbol{y}_1|\boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (19)$$

In the filtering context, the predictive distributions $p(\boldsymbol{y}_i|\boldsymbol{y}_{1:i-1},\boldsymbol{\theta})$ are calculated based on the marginal posterior of the model states, see Eq. (17).

Thus, the likelihood of the whole data $\boldsymbol{y}_{1:n}$ can be calculated as the product of the predictive distributions of the individual observations. That is, to check how well the model with parameter vector $\boldsymbol{\theta}$ fits the observations, one can check how individual predictions made from the current posterior fit the next observations. The only difference to traditional model fitting is that the state distribution is updated after each measurement.

Note that the above analysis only tells how the parameter likelihood is related to filtering methods. We have not yet discussed how the parameter estimation can be implemented in practice. In order to obtain the parameter estimates, two steps are required: (a) a filtering method to compute the posterior density for a given parameter value and (b) a parameter estimation algorithm to obtain the estimates. In this paper, we use variants of the Kalman filter for task (a), but other filtering methods, such as particle filters (see, e.g. Cappe et al., 2007), could be applied as well. For task (b) there are several optimization and Monte Carlo approaches available, and the method of choice depends on the case. In our examples, we use three different methods to estimate the posterior distribution: a maximum a posteriori (MAP) optimization approach with a Gaussian approximation of the posterior, a Markov chain Monte Carlo (MCMC) algorithm, and an importance sampling approach. For the sake of completeness, these algorithms are briefly reviewed in Appendix A.

Next, we will present how the parameter estimation is performed in the more familiar case, where the distributions are assumed to be Gaussian, and the extended Kalman filter (EKF) is used as the filtering method.

3.2 EKF likelihood

The extended Kalman filter is one of the most popular methods for state estimation in nonlinear dynamical models. EKF is an extension to the Kalman filter (KF, Kalman, 1960), where the model is assumed to be linear.

Let us now write the state space model as follows:

$$\boldsymbol{x}_k = \mathcal{M}(\boldsymbol{x}_{k-1},\boldsymbol{\theta}) + \boldsymbol{E}_k \quad (20)$$

$$\boldsymbol{y}_k = \mathcal{K}(\boldsymbol{x}_k) + \boldsymbol{e}_k. \quad (21)$$

Unlike in the standard EKF, the model $\mathcal{M}$ now depends on parameters θ. The model and observation errors are assumed to be zero mean Gaussians: $\boldsymbol{E}_k \sim N(\mathbf{0}, \mathbf{C}_k^{\mathrm{E}})$ and $\boldsymbol{e}_k \sim N(\mathbf{0}, \mathbf{C}_k^{\mathrm{e}})$.

In KF, the prediction and update steps can be written down analytically, since everything is linear and Gaussian. EKF uses the KF formulas, but the model matrix in KF is replaced by a linearization of the nonlinear model. In EKF, the predictive distribution for the state at time k is

$$\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1},\boldsymbol{\theta} \sim N(\boldsymbol{x}_k^p, \mathbf{C}_k^p), \quad (22)$$

where $\boldsymbol{x}_k^p = \mathcal{M}(\boldsymbol{x}_{k-1}^{\mathrm{est}},\boldsymbol{\theta})$ is the posterior mean $\boldsymbol{x}_{k-1}^{\mathrm{est}}$ from the previous time evolved with the model. The prior covariance $\mathbf{C}_k^p = \mathbf{M}_k^{\boldsymbol{\theta}}\mathbf{C}_{k-1}^{\mathrm{est}}\mathbf{M}_k^{\boldsymbol{\theta}^T} + \mathbf{C}_k^{\mathrm{E}}$ is the covariance of the state estimate evolved with the linearized model $\mathbf{M}_k^{\boldsymbol{\theta}} = \partial\mathcal{M}(\boldsymbol{x}_{k-1}^{\mathrm{est}},\boldsymbol{\theta})/\partial\boldsymbol{x}_{k-1}^{\mathrm{est}}$.

In the update step, the prior in Eq. (22) is updated with the new observation $\boldsymbol{y}_k$. In the EKF formulation, the posterior is

$$\boldsymbol{x}_k|\boldsymbol{y}_{1:k},\boldsymbol{\theta} \sim N(\boldsymbol{x}_k^{\mathrm{est}}, \mathbf{C}_k^{\mathrm{est}}), \quad (23)$$

where $\boldsymbol{x}_k^{\mathrm{est}}$ and $\mathbf{C}_k^{\mathrm{est}}$ are given by the Kalman filter formulas:

$$\boldsymbol{x}_k^{\mathrm{est}} = \boldsymbol{x}_k^p + \mathbf{G}_k(\boldsymbol{y}_k - \mathcal{K}(\boldsymbol{x}_k^p)) \quad (24)$$

$$\mathbf{C}_k^{\mathrm{est}} = \mathbf{C}_k^p - \mathbf{G}_k\mathbf{K}_k\mathbf{C}_k^p. \quad (25)$$

Here $\mathbf{G}_k = \mathbf{C}_k^p\mathbf{K}_k^T(\mathbf{K}_k\mathbf{C}_k^p\mathbf{K}_k^T + \mathbf{C}_k^{\mathrm{e}})^{-1}$ is the Kalman gain matrix and $\mathbf{K}_k = \partial\mathcal{K}(\boldsymbol{x}_k^p)/\partial\boldsymbol{x}_k^p$ is the linearized observation operator. The predictive distribution of measurement $\boldsymbol{y}_k$, needed in the likelihood evaluation, is given by

$$\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1},\boldsymbol{\theta} \sim N(\mathcal{K}(\boldsymbol{x}_k^p), \mathbf{C}_k^y), \quad (26)$$

where $\mathbf{C}_k^y = \mathbf{K}\mathbf{C}_k^p\mathbf{K}^T + \mathbf{C}_k^{\mathrm{e}}$. Now, applying the general formula (19), the total likelihood of observing $\boldsymbol{y}_{1:n}$, given parameters $\boldsymbol{\theta}$, can be written as

$$p(\boldsymbol{y}_{1:n}|\boldsymbol{\theta}) = p(\boldsymbol{y}_1|\boldsymbol{\theta})\prod_{k=2}^{n} p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1},\boldsymbol{\theta})$$

$$= \prod_{k=1}^{n}\exp\left(-\frac{1}{2}\boldsymbol{r}_k^T(\mathbf{C}_k^y)^{-1}\boldsymbol{r}_k\right)(2\pi)^{-d/2}|\mathbf{C}_k^y|^{-1/2}$$

$$\propto \exp\left(-\frac{1}{2}\sum_{k=1}^{n}\left[\boldsymbol{r}_k^T(\mathbf{C}_k^y)^{-1}\boldsymbol{r}_k + \log|\mathbf{C}_k^y|\right]\right), \quad (27)$$

where $\boldsymbol{r}_k = \boldsymbol{y}_k - \mathcal{K}(\boldsymbol{x}_k^p)$ and $|\cdot|$ denotes the determinant. Note that the normalization constants of the likelihood terms depend on θ implicitly through the covariances $\mathbf{C}_k^p$, and the term $\log|\mathbf{C}_k^y| = \log|\mathbf{K}_k\mathbf{C}_k^p\mathbf{K}_k^T + \mathbf{C}_k^{\mathrm{e}}|$ therefore needs to be included.

The above likelihood resembles the traditional "least squares" type of Gaussian likelihoods. The difference is that the model state is allowed to change between time steps, and the residuals are weighted by the model prediction uncertainty term $\mathbf{K}_i\mathbf{C}_i^p\mathbf{K}_i^T$ in addition to the measurement error covariance $\mathbf{C}_i^{\mathrm{e}}$. In fact, removing the model uncertainty terms $\mathbf{K}_i\mathbf{C}_i^p\mathbf{K}_i^T$ reduces the likelihood to the classical Gaussian

likelihood often used in parameter estimation. Adding the prediction error covariances to the sum of squares terms essentially gives more weight to situations that are predictable, and down-weights the terms where the model prediction is uncertain, due to, e.g. chaotic behavior of the system.

If, in addition to the actual model parameters $\boldsymbol{\theta}$, the model error covariance $\mathbf{C}_{\mathrm{E}}$ is unknown, we can parameterize it and estimate its parameters from the measurements together with the model parameters. As discussed later in more detail, the ability to estimate the variance parameters is one of the advantages of the likelihood approach, compared to the state augmentation and summary statistics methods.

Unfortunately, as the dimension of the model increases, EKF soon becomes practically infeasible. An approximation to EKF that can be implemented in large-scale systems is the ensemble Kalman filter (EnKF). In EnKF and its numerous variants (see, e.g. Evensen, 2007; Ott et al., 2004; Whitaker and Hamill, 2002; Zupanski, 2005), the computational issues in EKF are circumvented by using sample statistics in EKF formulas, computed from a relatively small number of ensembles. Hence, when EnKF is used, the likelihood (27) can be computed simply by defining $\boldsymbol{x}_i^p$ and $\mathbf{C}_i^p$ as sample mean and covariance matrix estimated from the ensemble.

Most implementations of the EnKF involve random perturbations of the model states and observations. This introduces randomness in the likelihood function (27): two evaluations with the same parameter value give different likelihood values. As noted by Dowd (2011), this complicates the parameter inference, and one has to resort to stochastic optimization methods that can handle noise in the target function (see Shapiro et al., 2009, for an introduction). Note that some recent variants of EnKF, such as many of the so called ensemble square root filters (Tippett et al., 2003) do not involve random components, and they might be more suitable for parameter estimation purposes. We test a variant called Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007) in the experiments of Sect. 5.

4 Parameter estimation with state augmentation

In the previous section, the parameter estimation was carried out off-line by repeatedly sweeping through the data using a filter. In state augmentation (SA), the parameters are added to the state vector and estimated on-line in the filter. In practice this means that model parameters are updated together with the state, whenever new observations become available. Next, we will present how SA can be implemented with EKF.

Let us consider the following state space model, where the parameter vector is modeled as an additional dynamical variable:

$$\boldsymbol{x}_{k+1} = \mathcal{M}(\boldsymbol{x}_k, \boldsymbol{\theta}_k) + \boldsymbol{E}_{\boldsymbol{x}} \tag{28}$$

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \boldsymbol{E}_{\boldsymbol{\theta}} \tag{29}$$

$$\boldsymbol{y}_{k+1} = \mathcal{K}(\mathbf{x}_{k+1}) + \boldsymbol{e}. \tag{30}$$

For notational convenience, we have dropped the time index k from the error terms.

In SA, we treat the combined vector $\boldsymbol{s}_k = [\boldsymbol{x}_k, \boldsymbol{\theta}_k]^T$ as the state vector that is updated at each time step k. The model for the combined vector can be written as

$$\boldsymbol{s}_{k+1} = \tilde{\mathcal{M}}(\boldsymbol{s}_k) + \boldsymbol{E}_{\boldsymbol{x},\boldsymbol{\theta}}, \tag{31}$$

where $\tilde{\mathcal{M}}(\boldsymbol{s}_k) = [\mathcal{M}(\boldsymbol{x}_k, \boldsymbol{\theta}_k), \boldsymbol{\theta}_k]^T$ and $\boldsymbol{E}_{\boldsymbol{x},\boldsymbol{\theta}}$ is the error of the augmented model $\tilde{\mathcal{M}}$, here assumed to be zero mean Gaussian with covariance matrix $\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}}$.

In EKF, we now need to linearize $\tilde{\mathcal{M}}(\boldsymbol{s}_k)$ with respect to $\boldsymbol{s}_k$, which results in the following Jacobian matrix:

$$\mathbf{M}_k = \frac{\partial \tilde{\mathcal{M}}(\boldsymbol{s}_k)}{\partial \boldsymbol{s}_k} = \begin{bmatrix} \partial \mathcal{M}(\boldsymbol{s}_k)/\partial \boldsymbol{x}_k & \partial \mathcal{M}(\boldsymbol{s}_k)/\partial \boldsymbol{\theta}_k \\ \partial \boldsymbol{\theta}_k/\partial \boldsymbol{x}_k & \partial \boldsymbol{\theta}_k/\partial \boldsymbol{\theta}_k \end{bmatrix} \tag{32}$$

$$= \begin{bmatrix} \partial \mathcal{M}(\boldsymbol{s}_k)/\partial \boldsymbol{x}_k & \partial \mathcal{M}(\boldsymbol{s}_k)/\partial \boldsymbol{\theta}_k \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{33}$$

Now, this matrix $\mathbf{M}_k$ can be used in the EKF formulas. Note that the top left term in the matrix is the linearization with respect to the actual states, which is needed in the standard states-only EKF as well. In addition, the derivative with respect to the parameters is needed (the top right term).

In EKF, we also need to define the model error covariance matrix $\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}}$. In SA, this must be defined in the joint space of the state and the parameters. The errors in the state and parameters are hardly uncorrelated, but for simplicity we model them here as independent random variables, which yields a block diagonal error covariance matrix

$$\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{C}_{\boldsymbol{x}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\boldsymbol{\theta}} \end{bmatrix}. \tag{34}$$

The model error in the state, $\mathbf{C}_{\boldsymbol{x}}$, has a clear interpretation: it represents the statistical properties of the error that the model makes in a filter time step. However, the parameter error covariance matrix $\mathbf{C}_{\boldsymbol{\theta}}$ lacks such an interpretation. We consider $\mathbf{C}_{\boldsymbol{\theta}}$ as an additional tuning parameter of the SA approach. Roughly speaking, increasing $\mathbf{C}_{\boldsymbol{\theta}}$ allows more sudden changes from $\boldsymbol{\theta}_k$ to $\boldsymbol{\theta}_{k+1}$. Note that, unlike in the full likelihood approach, the model error covariance $\mathbf{C}_{\boldsymbol{x}}$ cannot be estimated from data using SA. A simple example illustrating this problem is shown by DelSole and Yang (2010). The effect of (and the sensitivity to) $\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}}$ is studied in more detail in the experimental section. In Appendix B, we give some theoretical discussion of the effect of the selected $\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}}$.

As in the likelihood approach, the SA parameter estimation method can be implemented using other filtering methods besides EKF. For instance, replacing EKF with EnKF is straightforward: the ensembles now contain perturbations of both the model states and the parameters. The EnKF SA approach has been implemented, for instance, for a marine biochemistry model by Dowd (2011) and for an atmospheric model by Annan et al. (2005).

Conceptually, SA is straightforward, although, like noted by Järvinen et al. (2010), it implicitly assumes the static parameters as dynamical quantities and the parameter estimates therefore change at every update step of the filter. In some applications, such as numerical weather prediction (NWP), this may be critical. Operational NWP systems perform under strict requirements of timely product delivery to end-users. The "drifting" model parameters have to be therefore carefully considered from the system reliability point-of-view.

5 Case study: parametrized Lorenz 95

In this section, we will demonstrate the discussed parameter estimation approaches using a modified Lorenz system. We start by describing the model and the experiments, and then present the results for the three different methods.

5.1 Description of the experiment

To demonstrate and compare the parameter estimation approaches, we use a modified version of the Lorenz 95 ODE system, detailed by Wilks (2005). The chaotic Lorenz model (Lorenz, 1995) is commonly used as a low order test model to study estimation algorithms. The system used here is similar to the original system, but the state variables x_i are affected by forcing due to fast variables y_j, too. The full system is written as

$$\frac{dx_k}{dt} = -x_{k-1}(x_{k-2} - x_{k+1}) - x_k + F - \frac{hc}{b}\sum_{j=J(k-1)+1}^{Jk} y_j \tag{35}$$

$$\frac{dy_j}{dt} = -cby_{j+1}\left(y_{j+2} - y_{j-1}\right) - cy_j + \frac{c}{b}F_y + \frac{hc}{b}x_{1+\lfloor\frac{j-1}{J}\rfloor} \tag{36}$$

where $k = 1,\ldots,K$ and $j = 1,\ldots,JK$. That is, each of the "slow" state variables x_i are forced by a sum of the additional fast variables y_j. The fast variables have dynamics similar to the slow variables, but they are also coupled with the slow variables. We use values $K = 40$, $J = 8$, $F = F_y = 10$, $h = 1$ and $c = b = 10$, adopted from (Leutbecher, 2010).

The system (35)–(36) is considered as the "truth" and used for generating synthetic data. As a forecast model, we use a version where the net effect of the fast variables is described using a deterministic parameterization. The forecast model reads as

$$\frac{dx_k}{dt} = -x_{k-1}(x_{k-2} - x_{k+1}) - x_k + F - g(x_k,\boldsymbol{\theta}), \tag{37}$$

where $g(x_k,\boldsymbol{\theta})$ is the parameterization in which the missing fast variables y_j are modeled using the "resolved" variables. Here, we use a polynomial parameterization,

$$g(x_k,\boldsymbol{\theta}) = \sum_{i=0}^{d}\theta_i x_k^{(i)}, \tag{38}$$

with $d = 1$. The goal is to "tune" the parameters $\boldsymbol{\theta}$ so that the model fits the observations as well as possible. The parameter estimation resembles the closure parameter estimation problem in atmospheric models: the forecast model is solved with a time step $\Delta t = 0.025$, which is too crude for modeling the fast variables that operate on a finer time scale.

The observations for parameter estimation are generated as follows. The model is solved with dense time stepping ($\Delta t = 0.0025$) for altogether 2500 days (in the Lorenz model, one day corresponds to 0.2 time units). Then Gaussian noise is added to the model output with zero mean and covariance $(0.1\sigma_{\text{clim}})^2\mathbf{I}$, where $\sigma_{\text{clim}} = 3.5$ (standard deviation from long simulations). When the parameters are estimated using the filtering approaches, only 24 out of the 40 slow variables are assumed to be observed each day. The observation operator, used also in previous data assimilation studies (Auvinen et al., 2009, 2010), picks the last three state variables from every set of five states and we thus observe states $3,4,5,8,9,10,\ldots,38,39,40$. Partial observations were assumed to emulate a realistic data assimilation setting. In the experiments with the summary statistics approach, all the 40 states are assumed to be observed because hiding some of the states would introduce problems in the computation of the statistics.

Note that with this set-up, it is possible to use the values of the fast variables y_j simulated in the full system (35)–(36) to estimate parameters θ of the forcing model

$$g(x_k,\boldsymbol{\theta}) \approx \frac{hc}{b}\sum_{j=J(k-1)+1}^{Jk} y_j\,.$$

We will use the term "reference parameter values" for $\boldsymbol{\theta}$ which minimize the errors of this forcing model in the least squares sense. Naturally, such fitting cannot be performed in real applications since the actual sub-grid scale forcing is not known.

5.2 Results

In this section, we will present the results using the summary statistics, likelihood and state augmentation approaches. Our emphasis is on comparing the accuracy and the properties of these different approaches.

5.2.1 Summary statistics

In a synthetic example like Lorenz 95, summarizing the data in the form of a few statistics is not a trivial task. For example, if one wants to repeat the parameter estimation procedure similarly to Järvinen et al. (2010); Sexton et al. (2011), it is not clear what would be a suitable counterpart for the zonal and monthly averages in Eq. (7).

Matching observed and simulated statistics

As mentioned in Sect. 2.1, the simulated statistics may not be a smooth function of the parameters. One method that is rather insensitive to that kind of behavior in the likelihood is importance sampling (see Appendix A3 for details). We perform importance sampling for the two parameters $[\theta_0, \theta_1]$ of the model Eqs. (37)–(38). First, we draw 1000 candidate values uniformly and independently for the two parameters from the intervals $\theta_0 \in [1.4, 2.2]$ and $\theta_1 \in [0, 0.12]$. Then, the system defined by Eqs. (37)–(38) is simulated for each candidate value, the summary statistics are computed and the likelihood is calculated. The parameter ranges were chosen so that the shape of the posterior distribution is clearly visible.

In the first experiment, the cost function was constructed around a set of summary statistics which were selected arbitrarily. We used six statistics: mean, variance, auto-covariance with time lag 1, covariance of a node with its neighbor and cross-covariance of a node with its two neighbors for time lag 1. Since the model is symmetric with respect to the nodes, we averaged these statistics across different nodes. The cost function was

$$C(\hat{\boldsymbol{s}}, \boldsymbol{s}_{\boldsymbol{\theta}}) = \sum_{i=1}^{6} (\hat{s}^i - s^i_{\boldsymbol{\theta}})^2 / \hat{\sigma}_i^2, \tag{39}$$

where $s^i_{\boldsymbol{\theta}}$ is one of the six statistics computed from data simulated with parameters $\boldsymbol{\theta}$, and $\hat{s}^i$, $\hat{\sigma}_i^2$ are the mean and variance of the same statistics computed from a relatively long simulation of the full-system (35)–(36) similarly to (Järvinen et al., 2010) . All the 40 variables x_k were assumed to be observed and the observation period was taken to be 200 days.

Figure 2 shows the importance sampling weights obtained for the 1000 candidate values. The results show that the cost function (39) does not restrict the model parameters much and the posterior distribution is rather broad. The parameter estimates are also clearly biased: the values obtained using the knowledge of the simulated fast variables are outside the obtained distribution. Note also the "spiky" behavior of the cost function: the weights do not vary smoothly as a function of the parameters.

Likelihood based on an emulator

In the next experiment, the likelihood was computed using an emulator trained on the same set of six summary statistics. We again performed importance sampling of parameters θ_0 and θ_1 of the model (37)–(38) using the same 1000 candidate values drawn from $\theta_0 \in [1.4, 2.2]$ and $\theta_1 \in [0, 0.12]$. The likelihood Eq. (11) was constructed using an emulator, as explained in Sect. 2.2.

Figure 3a presents the results for the likelihood Eq. (11) in which the dimensionality reduction was performed using PCA with only two principal components retained. Figure 3b

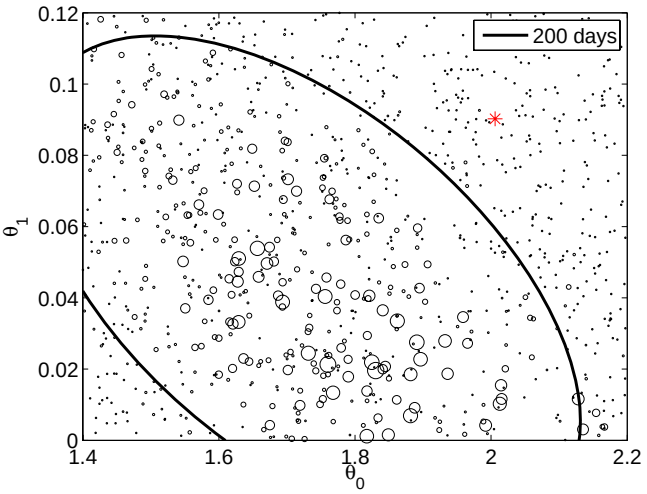


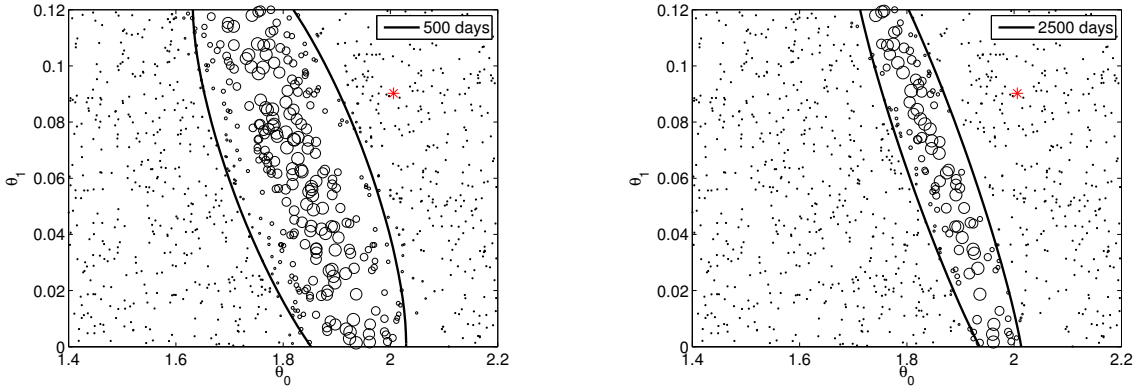
Fig. 2. The scatter plots of the parameter value candidates and their likelihood (represented by the size of the markers) obtained with cost function (39). The marker size reflects the value of the weights for importance sampling. The black ellipse represents the first two moments estimated with importance sampling. The red star represents the parameter values estimated by fitting the forcing model to the simulated fast variables (see Sect. 5 for details).

presents similar results for the case when the dimensionality of the features was reduced from six to two by fitting a linear regression model $\boldsymbol{\theta} \approx \mathbf{W}\boldsymbol{s}_{\boldsymbol{\theta}}$ and by using a feed-forward neural network (see, e.g. Bishop, 2006) to build an emulator.

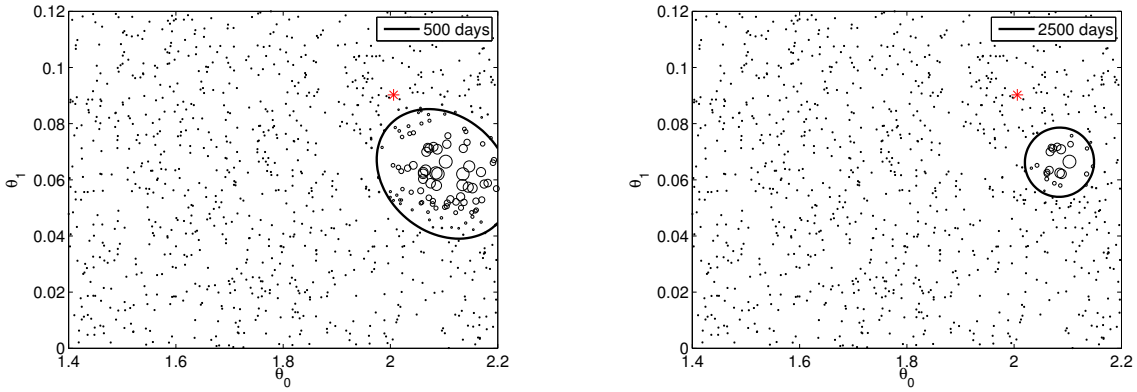
There are a few remarks that we can make based on the obtained results. Using the emulator results in a posterior density which is a smooth function of the parameters, thus the problem of the spiky behavior is solved. The parameters found with this approach are close to the reference values but there is a bias which is affected by the choice of the summary statistics. This effect is related to the known result from the field of Approximate Bayesian Computation that only using *sufficient* summary statistics yields the same posterior distribution as when the full data set is used (Marjoram and Tavaré, 2006). A longer observational period results in a more narrow posterior distribution but the bias problem remains.

The results are generally sensitive to the dimensionality reduction approach and to the number of components retained in the model. In this simple example, using more or less components leads to qualitatively similar results (biased estimates). In more complex cases, a cross-validation procedure (e.g. similar to Sexton et al., 2011) might be needed to estimate the right number of the required components.

Another problem is that the observed values of some of the summary statistics cannot be obtained by varying the parameters of the surrogate model. This situation can easily occur in real model tuning and it may result in over-fitting of the parameters to such problematic statistics. In the results shown in Fig. 3a, this problem is concealed by the fact that



(a) Dimensionality reduction by PCA



(b) Dimensionality reduction by fitting a linear regression model $\theta \approx \mathbf{W}\mathbf{s}_\theta$

Fig. 3. The scatter plots of the parameter value candidates and their likelihood (represented by the size of the markers) obtained with cost function (11). The simulation length is 500 days (left) and 2500 days (right). The marker size reflects the value of the weights for importance sampling. The black ellipses represent the first two moments estimated with importance sampling. The red start represents the parameter values estimated by fitting the forcing model to the simulated fast variables (see Sect. 5 for details).

only two principal components are retained in the analysis and those principal components can be simulated well by the surrogate model.

The summary statistics approach has a few other potential problems. The choice of the summary statistics has a critical impact on the estimated parameter values. Some arbitrarily selected statistics may not be affected by varying the model parameters and the idea behind the most informative projections is to diminish this problem. In the tuning of climate models, summary statistics often include only monthly and regional averages of some state variables, which means the focus is on how well climate models reproduce the seasonal cycle. It may be that some model parameters have little effect on the seasonal cycle but they can be important for the overall quality of a climate model.

Thus, the summary statistics approach has a few practical problems and can result in biased estimates. We think that the essential problem is the averaging procedure in which a lot of important information is lost. We argue that the sequential methods provide a more appropriate way to determine the likelihood for the parameters.

5.2.2 Likelihood calculations using filtering methods

In this section, we estimate the parameters of the forecast model (37) for the Lorenz 95 system using the filtering methodology presented in Sect. 3.

Filtering with EKF

Implementation of EKF requires linearization of the model (37), which is rather straightforward in this synthetic example. As mentioned in Sect. 3.2, the EKF filtering procedure also requires the model error covariance matrix $\mathbf{C}_\mathrm{E}$. We use a simple parameterization:

$$\mathbf{C}_\mathrm{E} = \sigma^2 \mathbf{I}, \tag{40}$$

where σ^2 is a parameter which needs to be estimated together with parameters $\boldsymbol{\theta}$ of the forecast model (37). In practice, we estimate the logarithm of σ^2, which guarantees the positivity of the variance and yields a posterior distribution whose shape is closer to Gaussian. We perform parameter estimation using delayed rejection adaptive Metropolis (DRAM) MCMC (Haario et al., 2006, Appendix A).

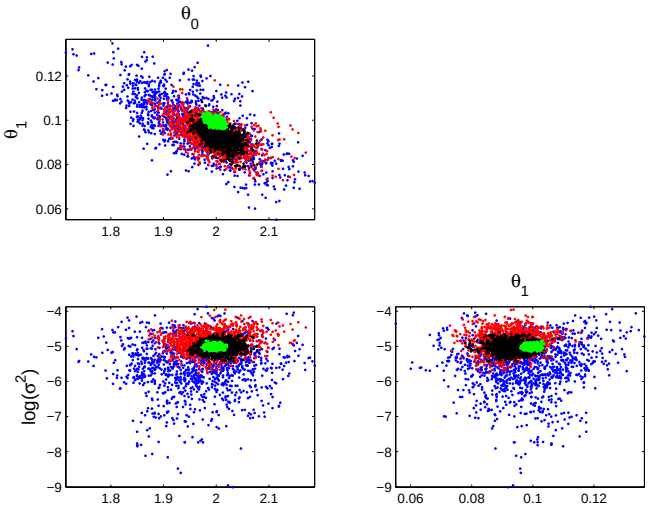


Fig. 4. Scattering plot of parameter pairs from MCMC runs using 10 (blue), 20 (red), 50 (black) and 500 (green) day simulations. Clear tightening of posterior can be observed as the simulation length increases. The third parameter is related to the model error covariance.

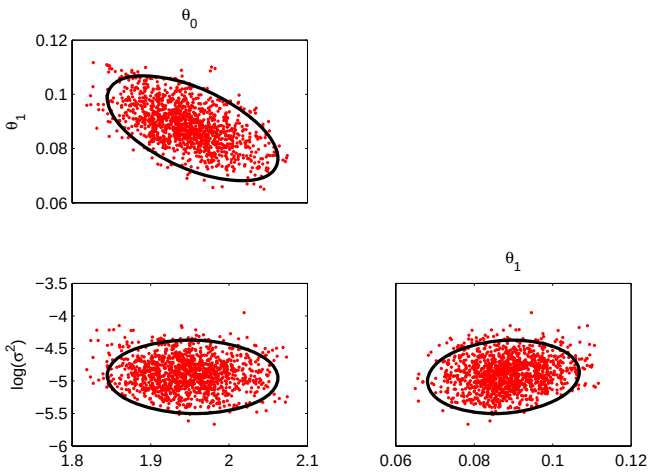


Fig. 5. Posterior distribution using 20 day simulations with MCMC (red dots) and Gaussian approximation (black ellipse).

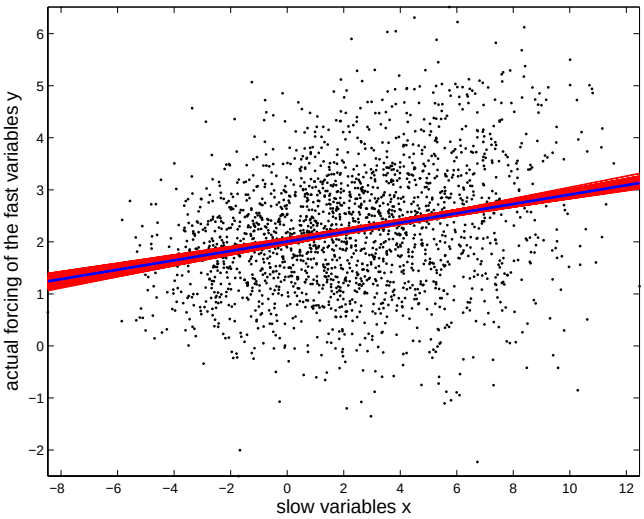


Fig. 6. Actual forcing of fast variables (black cloud). Red lines indicate forcing from MCMC runs. Red lines in the figure represent results from the 50 day MCMC simulations. Blue line is gotten by formally fitting the parameter values in the cloud.

In Fig. 4, the pairwise marginal distributions for the parameters are illustrated using 10, 20, 50 and 500 day simulations. As expected, the distribution becomes tighter as the simulation length increases. Note that the posterior distribution is much tighter compared to the summary statistics approach (see, e.g. Fig. 3) even though almost half of the states are not observed and the filtering procedure is applied to a relatively short observation sequence. The parameter estimates are closer to the reference values obtained using the knowledge of the simulated fast variables, compared to the estimates obtained via the summary statistics approach. We also observe that the parameter distribution is approximately Gaussian when sufficiently long simulations are used, as shown in Fig. 5.

In Fig. 6, we plot the true forcing in the simulated full system (35)–(36) against the slow variables. The red lines in the figure represent the parameter values from the 50 day MCMC simulation. The blue line represent the parameter values obtained by fitting a line to the true forcing in the least squares sense. We observe good agreement with our results and the fitted line.

The estimates obtained by the likelihood approach are close to the reference values obtained using the knowledge of the fast variables. However, there is no reason to think that the reference values are optimal, for instance in the sense of forecast accuracy. Therefore, to further demonstrate that likelihood approach produces good parameter estimates, we study the effect of the parameters on the forecast skill of the model. We grid the 2-dimensional parameter space and calculate the average 6 day forecast skill for different parameter values. The average forecast skill is computed by making a 6 day forecast starting every 24 h for 100 days. The averaged forecast skill can be written as

$$f(\boldsymbol{\theta}) = \frac{1}{NK\sigma_{\text{clim}}^2} \sum_{i=1}^{N} \|\mathcal{M}_6(\boldsymbol{x}_i^{\text{true}}, \boldsymbol{\theta}) - \boldsymbol{x}_{i+6}^{\text{true}}\|_2^2,$$

where $N = 100$, $K = 40$ and $\sigma_{\text{clim}} = 3.5$. The notation $\mathcal{M}_6(\boldsymbol{x}_i^{\text{true}}, \boldsymbol{\theta})$ means a 6 day prediction launched from the true state $\boldsymbol{x}_i^{\text{true}}$ with parameter values $\boldsymbol{\theta}$. The contour lines of the average forecast skill and the parameter values obtained by 10, 50 and 500 day MCMC runs are shown in Fig. 7. Again, we observe a good agreement: parameters tuned with the likelihood approach yield a good average forecast skill provided that the simulation length is sufficient.

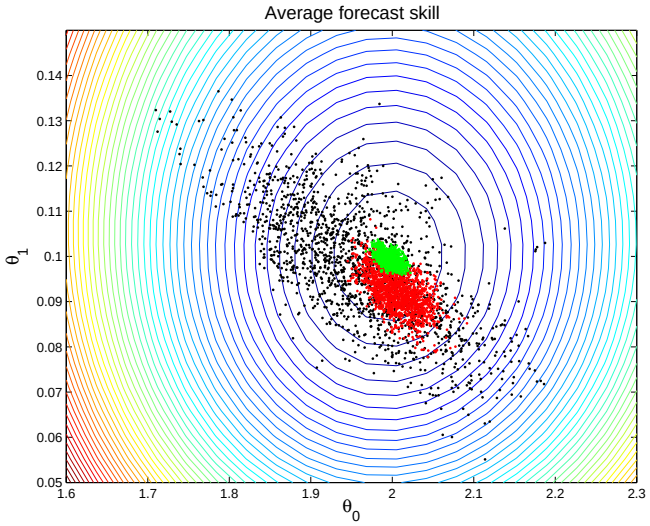


Fig. 7. An illustration of the average forecast skill. Black, red and green dots indicate the results from 10, 50 and 500 day MCMC simulations, respectively. Blue contour colors indicate high forecast skill.

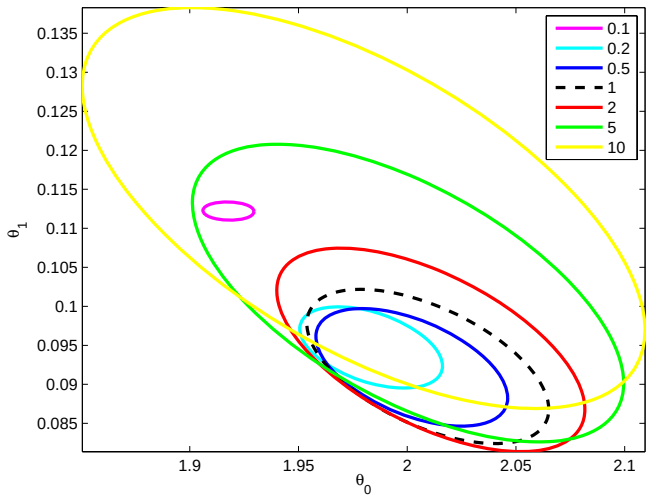


Fig. 8. Gaussian posterior approximations with 10, 5 and 2 times too small and too large σ^2.

Sensitivity to model error covariance

The possibility to estimate the model error covariance $\mathbf{C}_E$ from data is an advantage of the parameter estimation based on filtering. In data assimilation, $\mathbf{C}_E$ is often considered as a tuning parameter which has to be selected somehow. In large scale models like NWP, the model error covariance is usually estimated in a separate procedure, and finding an effective parametrization of $\mathbf{C}_E$ is an open question (see, e.g. Bonavita et al., 2008).

Therefore, in the following experiment, we test how specifying non-optimal values for the model error covariance affects the quality of parameter estimation. We use the same parameterization (40) and vary the variance parameter σ^2 so that it is two, five and 10 times smaller or greater than the optimal value obtained in the EKF exercise with the 500 day-long simulations (see Fig. 4). We run the likelihood approach with only 50 days of data. Since the posterior distribution is approximately Gaussian, we do not perform the computationally heavy MCMC runs, but compute the MAP estimate using an optimizer and calculate the Gaussian approximation of the posterior at the MAP estimate (see Appendix A).

The results are shown in Fig. 8. We observe that specifying the model error covariance wrongly can lead to biased parameter estimates: too small values of σ^2 lead to an underestimated posterior covariance of the parameters and vice versa. In this specific example, we change the correct $\mathbf{C}_E$ only by one order of magnitude and still obtain reasonable parameter estimates. For larger errors in $\mathbf{C}_E$, parameter estimates can be severely biased.

Filtering with EnKF

As discussed in Sect. 3.2, replacing the deterministic EKF with a stochastic filter, such as the EnKF, leads to a noisy likelihood function. We performed importance sampling for the two parameters similarly as in the summary statistics experiment in Section 5.2.1. The EnKF likelihood was evaluated by fixing the model error variance σ^2 to the optimum found in the EKF exercise, and setting the number of ensemble members to 100.

In our experiment, the noise in the likelihood function dominated the statistical analysis: most of the importance weights were assigned to only a few candidate values. That is, statistical inference could not be performed properly, and the noise in the likelihood seems to be a real issue in the likelihood approach computed with EnKF. Here, we settle for plotting the negative log-likelihood surface, as was done with EnKF parameter estimation by Dowd (2011). From Fig. 9 we observe that the general pattern is good: low values are found from the correct region, and a reasonable MAP estimate might be found using stochastic optimization techniques. However, statistical inference is complicated by the noisy likelihood. Smoothing methods could be used to alleviate the noise in the likelihood, but this question is not pursued further here.

In addition, we also tested the likelihood set-up with LETKF (Hunt et al., 2007) which falls into the category of ensemble square-root filters with no random perturbations. In this method, the model error term is neglected, but a covariance inflation term is used to inflate the posterior covariance and to account for the missing model error term. The covariance inflation parameter (see Hunt et al., 2007, for details) can be estimated together with the parameters like the model error term in EKF. In Fig. 10 we show MCMC runs

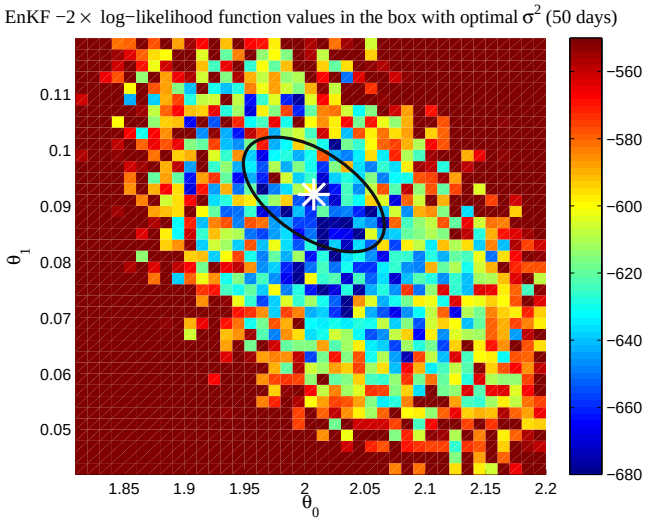


Fig. 9. The use of stochastic filtering method yields a noisy likelihood function. The results of 50 day run using EnKF based likelihood function is illustrated in the Figure. The values indicate negative log-likelihood times two values. White star and black ellipse is acquired from correspondent EKF likelihood calculations.

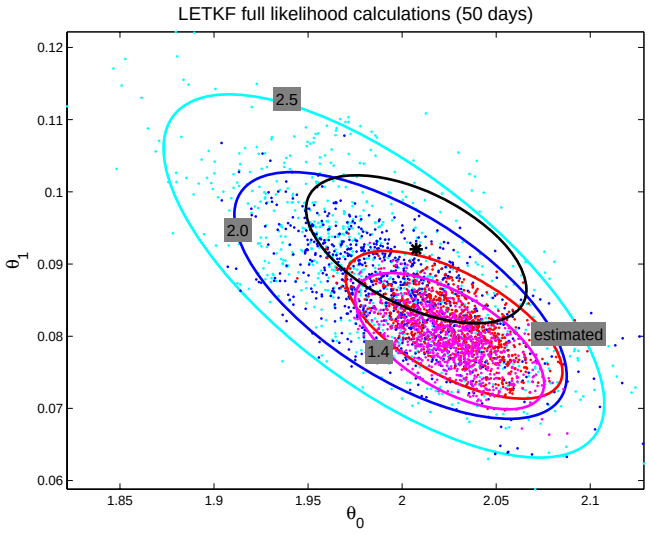


Fig. 10. Full likelihood calculations using LETKF data assimilation method. Different MCMC runs with different covariance inflation factor together with 95 % trust ellipses. Red color indicates a run, where the factor is also estimated (mean value is 1.52). Magenta, blue and cyan colors indicate a run where covariance inflation factor is fixed to 1.4, 2.0 and 2.5, respectively. Black star and ellipse is acquired from correspondent EKF likelihood calculations.

using different values for the covariance inflation parameter and a run where the inflation parameter is estimated together with the model parameters. Although there is a small bias, depending on the value of the covariance inflation parameter, the agreement with EKF calculations is rather good. Thus, deterministic ensemble filters seem to be more suitable for parameter estimation purposes.

5.2.3 State augmentation

As discussed in Sect. 4, in SA the model parameters are modeled as dynamical quantities, and the estimates do not converge to any fixed value as more observations are added. The rate at which parameter values can change from one filter time step to another is controlled by the extra tuning parameter, the model error covariance for the parameters, $\mathbf{C}_{\boldsymbol{\theta}}$. Here, we study how the SA method performs in parameter estimation and specifically how the tuning of $\mathbf{C}_{\boldsymbol{\theta}}$ affects the results.

Tuning of the parameter error covariance

In our experiments, we use a diagonal matrix as $\mathbf{C}_{\boldsymbol{\theta}}$ and keep it fixed during the runs. The model error for the state vector was fixed to its "optimal value", obtained from the likelihood experiments. In Fig. 11 we show four different runs using the EKF version of the SA method. The runs are made so that the diagonal elements of $\mathbf{C}_{\boldsymbol{\theta}}$ are taken to be 0.1 %, 1 %,10 % and 100 % of the optimal parameter values acquired from the likelihood experiment. In all cases, the SA method converged quickly to values near the optimum. The effect of the size of $\mathbf{C}_{\boldsymbol{\theta}}$ was as expected. When $\mathbf{C}_{\boldsymbol{\theta}}$ is set to be small, the method reacts slowly on new data and the parameter values take small steps. On the other hand, if $\mathbf{C}_{\boldsymbol{\theta}}$ is set large, the method allows larger deviations, but can yield unrealistic values for the parameters. Some theoretical discussion about the effect of $\mathbf{C}_{\boldsymbol{\theta}}$ is given in Appendix B.

In this example, we do not observe any systematic temporal variations in the parameter value. However, it is worth pointing out that the SA method could be useful in checking if such variations exist. Since the parameter trajectories are stationary, one could use the mean value of the trajectories as the final parameter estimate. In the current example, the mean is a good estimate, and it is also rather insensitive to the tuning of $\mathbf{C}_{\boldsymbol{\theta}}$. In general, however, the parameter trajectories cannot be interpreted in the statistical sense, since the parameter values and their variation depend entirely on the tuning of $\mathbf{C}_{\boldsymbol{\theta}}$. Thus, the SA method cannot be used for statistical inference of the parameters in the same sense as the likelihood approach.

Sensitivity to model error covariance

If, on the other hand, we keep the "parameter model error covariance" $\mathbf{C}_{\boldsymbol{\theta}}$ fixed and vary the model error covariance $\mathbf{C}_{\boldsymbol{x}}$, the effects are somewhat different than in the likelihood approach. Too small σ^2 values can cause the filter to diverge leading to unrealistic parameter estimates. Examples of runs with too large model error covariance are illustrated in Fig. 12. We observe that too large σ^2 values can cause bias to the estimates. In addition, $\mathbf{C}_{\boldsymbol{x}}$ affects the rate of change of

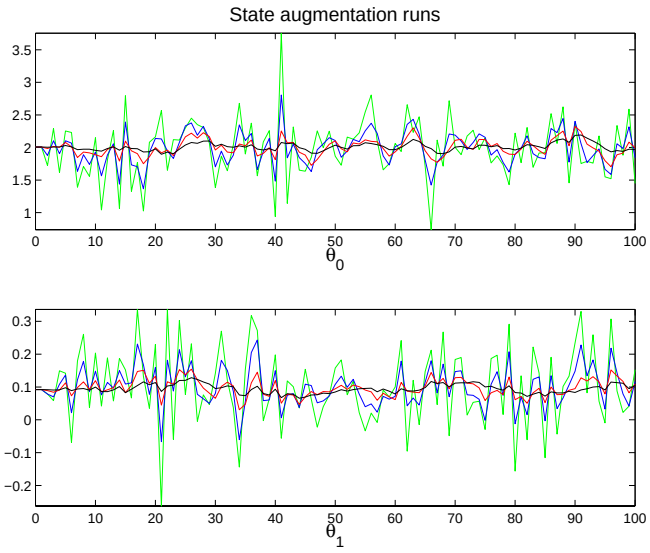


Fig. 11. Runs using the EKF version of the SA method, when the diagonal elements of $\mathbf{C}_{\boldsymbol{\theta}}$ are taken to be 0.1 % (black), 1 % (red),10 % (blue) and 100 % (green) of the optimal initial values. The effect of the size of $\mathbf{C}_{\boldsymbol{\theta}}$ was as expected. When $\mathbf{C}_{\boldsymbol{\theta}}$ is set to be small, the method reacts slowly on new data and the parameter values take small steps.

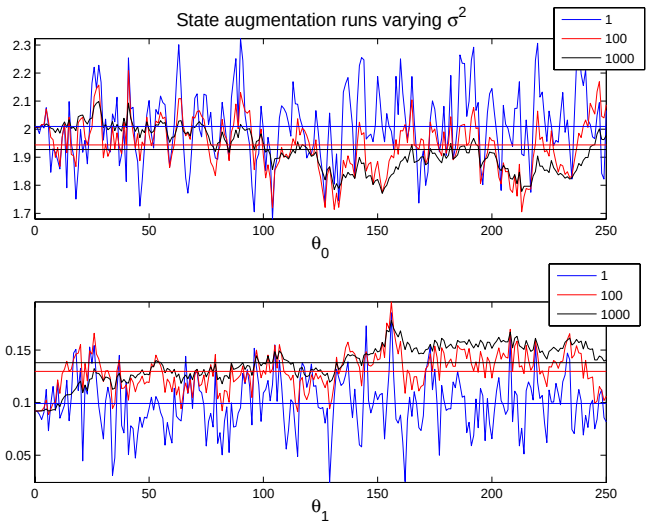


Fig. 12. Examples of too large model error: the model error variance is multiplied with 100 and 1000. Too large model error in state augmentation will cause bias in the parameter values. Note that the rate of change of the parameters becomes smaller as the variance grows, although $\mathbf{C}_{\boldsymbol{\theta}}$ is kept fixed in all runs. The straight lines represent the means of the parameter trajectories.

the parameters: the higher σ^2 is, the smaller are the parameter changes. The latter effect can be theoretically justified (see Appendix B for details).

State augmentation with EnKF

The use of ensemble based filtering methods is possible in state augmentation system. In our tests, with a large enough ensemble size, the results were similar to the EKF results. In the Lorenz system the minimum required ensemble size is roughly 50. Smaller ensemble size leads to underestimation of the covariances and can cause the filter to diverge. We note that state augmentation does not have the problem with the stochasticity of EnKF, which was encountered in the likelihood approach.

6 Remarks and discussion

6.1 Applicability to large scale systems

The discussed parameter estimation methods differ in their applicability to large scale systems like NWP and climate models. The summary statistics based approaches are computationally expensive although straightforward to implement: one only needs to simulate the model, and compare the selected summary statistics of the simulations and the observations. The difficulty lies in selecting the appropriate summary statistics that enable the identification of the parameters.

The state augmentation and the likelihood approaches depend on a data assimilation system, which is often available for NWP systems, but not commonly for climate models. The state augmentation method requires modifications to the assimilation method. In deterministic assimilation systems, such as the variational approximations to EKF that are often used in operational NWP systems (Rabier et al., 2000; Gauthier et al., 2007), one needs to add derivative computations with respect to the parameters. If an ensemble data assimilation method is used, such as a variant of the EnKF (see Houtekamer et al., 2005) parameter perturbations need to be added. Computationally, state augmentation is economical, since it requires only one assimilation sweep over the selected data.

The filter likelihood approach is computationally much more challenging than the state augmentation approach, since it involves many repeated filter sweeps, the number of which depends on the parameter estimation technique used. The likelihood approach requires, in addition to a data assimilation system, a method to estimate the forecast error covariance. In ensemble data assimilation systems, the covariance can be estimated from the ensemble. Variational data assimilation methods do not contain error propagation mechanisms, and it is not immediately clear how the likelihood approach can be implemented in these systems. A potential way is to obtain the covariance from ensemble prediction systems (EPS), that are built for approximating the prediction errors (Molteni et al., 1996; Palmer et al., 2005). Our preliminary tests with the low order system suggest that such EPS information could be used to approximate the likelihood approach, but verifying this is a topic of on-going research.

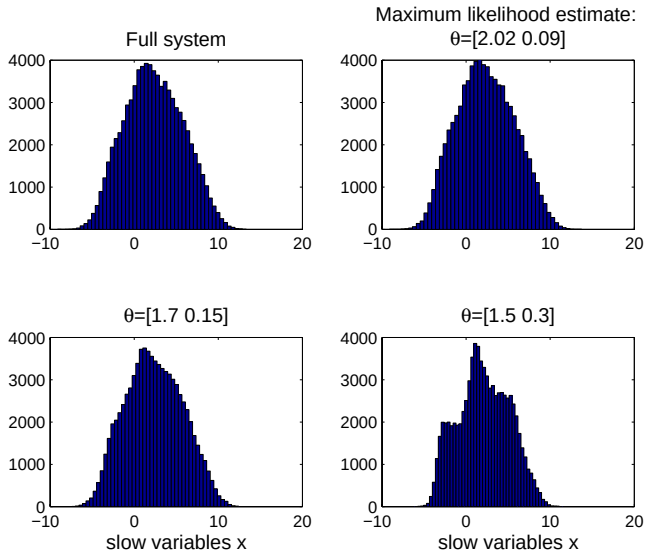


Fig. 13. Distribution of the state variables in the full Lorenz system (top left) and in the parameterized system with maximum likelihood estimate (top left) and two arbitrarily chosen "poor" parameter values (bottom row).

6.2 Climatologies with tuned parameters

In our likelihood experiment, the optimal parameter values led to improved short-range average forecast skills, as expected. Another question is related to the effect of parameter tuning on the quality of long model simulations (or "climatologies"): the tuned parameters should, in addition to improving short-range forecasts, improve climatologies, too. In Fig. 13, we compare the histograms of the state variables in the full Lorenz system and in the forecast model with different parameter values. We compare the statistics of the full system to the statistics produced by the forecast model with maximum likelihood parameter estimate and two arbitrarily chosen "poor" parameter values. We observe that, in this case, the parameters tuned with the likelihood approach produce also the correct climatologies. We also note that the overall statistics of the system can be quite good even with rather poor parameter values. This highlights the difficulties in choosing the correct likelihood for the summary statistics approach.

7 Conclusions

In this paper, we review three methods for parameter estimation in chaotic systems. In the summary statistics approach, the selected statistics computed from model simulations are compared to the same statistics calculated from observations. In the state augmentation method, unknown parameters are added to the state vector and estimated "on-line" together with the model state in a data assimilation system. In the likelihood approach, the likelihood for a parameter value is computed by running a data assimilation method "off-line" over a selected data set. All methods were studied using a modified version of the Lorenz 95 model.

Our results indicate that the summary statistics approach, albeit relatively easy to implement and compute, can have problems in properly identifying the parameters, and may lead to biased estimates. This result is supported by the previous climate model parameter estimation experiments (Järvinen et al., 2010) where simple summary statistics were not enough to uniquely identify all selected model parameters.

The state augmentation approach can work well and converge fast, if properly tuned. State augmentation contains additional tuning parameters, to which the performance of the method is somewhat sensitive: one must correctly specify the model error covariance both for the actual model states and for the parameters. The state augmentation approach is computationally feasible, since parameters are estimated "on-line" instead of repeatedly comparing model simulations to observations. The implementation of the method requires a modification to the data assimilation system. A downside of the approach is that the "static" model parameters are modeled as dynamical quantities, and one needs to accept the fact that the parameter estimates change at every time step and do not converge to a fixed value. Moreover, the method does not support statistical inference of the model parameters, since the obtained parameter values depend directly on the tuning of the model error covariance.

The likelihood approach performed well in our tests. The performance of the method was somewhat sensitive to the tuning of the model error covariance, like in the state augmentation approach. The likelihood approach assumes that the parameter values are static, and allows for statistical inference of the model parameters. The method requires a data assimilation system, and a method to propagate model error statistics. This may be restrictive in large-scale systems. The computational burden is much higher than in the state augmentation approach, and may be a bottleneck when scaling up to large scale NWP and climate models. The likelihood can be implemented with ensemble data assimilation methods, but the statistical analysis may be complicated by the stochasticity introduced into the likelihood function, if random perturbations are used in the ensemble method.

Appendix A

Parameter estimation algorithms

A1 MAP estimation and Gaussian approximation

The Maximum a Posteriori (MAP) estimate can be found by maximizing the posterior density $p(\boldsymbol{\theta}|\boldsymbol{y})$ with respect to $\boldsymbol{\theta}$, or, equivalently, minimizing the negative logarithm of the posterior

$$L(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta}|\boldsymbol{y}) = -\log p(\boldsymbol{y}|\boldsymbol{\theta}) - \log p(\boldsymbol{\theta}).$$

The maximization can be done by different numerical methods (see, e.g. Nocedal and Wright, 1999). Once the estimate $\hat{\boldsymbol{\theta}} = \arg\min L(\boldsymbol{\theta})$ has been obtained, one can construct a multivariate Gaussian approximation of $p(\boldsymbol{\theta}|\boldsymbol{y})$ around the point $\hat{\boldsymbol{\theta}}$. It is well known (see, e.g. Gelman et al., 2003) that the covariance matrix at $\hat{\boldsymbol{\theta}}$ can be approximated by the inverse Hessian of the negative logarithm of the posterior:

$$\mathrm{Cov}(\boldsymbol{\theta}) \approx \mathbf{H}(\hat{\boldsymbol{\theta}})^{-1},$$

where the Hessian matrix $\mathbf{H}(\hat{\boldsymbol{\theta}})$ contains the second derivatives of the negative logarithm of the likelihood, evaluated at $\hat{\boldsymbol{\theta}}$:

$$\mathbf{H}_{ij}(\hat{\boldsymbol{\theta}}) = \left.\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

The Hessian can be calculated analytically or numerically. In our examples, we have used a standard finite difference approximation applying the central difference formula (Nocedal and Wright, 1999).

A2 MCMC sampling

In principle, the Bayes formula, see Eq. (3), solves the parameter estimation problem in a fully probabilistic sense. However, the problem of calculating the integral of the normalizing constant is faced. This integration is often a formidable task, even for only moderately high number of parameters in a nonlinear model, and direct application of the Bayes formula is intractable for all but trivial nonlinear cases. The MCMC methods provide a tool to handle this problem. They generate a sequence of parameter values $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots \boldsymbol{\theta}_N$, whose empirical distribution approximates the true posterior distribution for large enough sample size N.

In many MCMC methods, instead of sampling directly from the true distribution, one samples from an artificial *proposal* distribution. Combining the sampling with a simple accept/reject procedure, the posterior can be correctly approximated. The simplest MCMC method is the *Metropolis algorithm* (Metropolis et al., 1953):

1. Initialize by choosing a starting point $\boldsymbol{\theta}_1$.
2. Choose a new candidate $\hat{\boldsymbol{\theta}}$ from a suitable proposal distribution $q(.|\boldsymbol{\theta}_n)$ that may depend on the previous point of the chain.
3. *Accept* the candidate with probability

$$\alpha(\boldsymbol{\theta}_n, \hat{\boldsymbol{\theta}}) = \min\left(1, \frac{\pi(\hat{\boldsymbol{\theta}})}{\pi(\boldsymbol{\theta}_n)}\right).$$

   If rejected, repeat the previous point in the chain. Go back to step 2.

So, points with $\pi(\hat{\boldsymbol{\theta}}) > \pi(\boldsymbol{\theta}_n)$, i.e., steps "uphill", are always accepted. But also points with $\pi(\hat{\boldsymbol{\theta}}) < \pi(\boldsymbol{\theta}_n)$, i.e. steps "downhill", may be accepted, with probability that is given by the *ratio* of the π values. In practice, this is done by generating a uniformly distributed random number $u \in [0,1]$ and accepting $\hat{\boldsymbol{\theta}}$ if $u \leq \pi(\hat{\boldsymbol{\theta}})/\pi(\boldsymbol{\theta}_i)$. Note that only the ratios of π at consecutive points are needed, so the main problem of calculation the normalizing constant is circumvented, since the constant cancels out.

However, the choice of the proposal distribution may still pose a problem. It should be chosen so that the "sizes" of the proposal q and target distributions suitably match. This may be difficult to achieve, and an unsuitable proposal can leadt to inefficient sampling. For simple cases, the proposal might be relatively easy to find by some hand-tuning. However, the "size" of the proposal distribution is not a sufficient specification. Especially In higher dimensions, the shape and orientation of the proposal are crucial. The most typical proposal is a multi–dimensional Gaussian (Normal) distribution. In the *random walk* version, the center point of the Gaussian proposal is chosen to be the current point of the chain. The task then is to find a covariance matrix that produces efficient sampling.

Several efficient adaptive methods have been recently proposed, for example, the adaptive Metropolis (AM) algorithm (Haario et al., 2001). In adaptive MCMC, one uses the sample history to automatically tune the proposal distribution "on-line" as the sampling proceeds. In AM, the empirical covariance from the samples obtained so far is used as the covariance of a Gaussian proposal. In this paper, a variant of AM called the delayed rejection adaptive Metropolis (DRAM, Haario et al., 2006) is used for all sampling tasks.

A3 Importance sampling

Some methods considered here use a likelihood which depends on initial conditions, random seeds and other settings, and the estimated likelihood is therefore random. For such methods, we used the following importance sampling procedure for estimating the parameters. The likelihood was computed for a set of candidate parameter values $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$ which were drawn from an *importance function* $g(\boldsymbol{\theta})$. The posterior distribution of the parameters was evaluated by weighting each sample according to their likelihood values with respect to the importance function:

$$w_i = p(z \mid \boldsymbol{\theta}_i)/g(\boldsymbol{\theta}_i).$$

One can now compute the required statistics using samples $\boldsymbol{\theta}_i$ with weights w_i. Here, we evaluated the weighted posterior mean

$$\bar{\boldsymbol{\theta}} = \frac{1}{W}\sum_{i=1}^{N} w_i \boldsymbol{\theta}_i$$

with $W = \sum_{i=1}^{N} w_i$ and the weighted covariance matrix

$$\mathbf{C}_{\boldsymbol{\theta}} = \frac{1}{W}\sum_{i=1}^{N} w_i (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})^T (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})$$

to approximate the confidence intervals for the closure parameters.

Appendix B

The effect of the model error covariance matrix in the state augmentation method

Here we study the effect of the block diagonal error covariance matrix

$$\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{C}_{\boldsymbol{x}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\boldsymbol{\theta}} \end{bmatrix}$$

in the state augmentation set up, that is when we assume that the errors in the state and the parameters are uncorrelated. For notational convenience, we do not use the time index k. Naturally, we do not observe the model parameters and the observation operator is

$$\tilde{\mathbf{K}} = \begin{bmatrix} \mathbf{K} & \mathbf{0} \end{bmatrix},$$

where $\mathbf{K}$ is the original observation operator.

The Kalman gain matrix $\mathbf{G}$, that defines how much the prior state and covariance are changed by an observation, can be written as

$$\begin{aligned} \mathbf{G} &= \mathbf{C}^p \tilde{\mathbf{K}}^T (\tilde{\mathbf{K}} \mathbf{C}^p \tilde{\mathbf{K}}^T + \mathbf{C}^{\mathrm{e}})^{-1} \\ &= \mathbf{M}^{\boldsymbol{\theta}} \mathbf{C}^{\mathrm{est}}_{k-1} \mathbf{M}^{\boldsymbol{\theta}^T} \tilde{\mathbf{K}}^T (\tilde{\mathbf{K}} \mathbf{M}^{\boldsymbol{\theta}} \mathbf{C}^{\mathrm{est}}_{k-1} \mathbf{M}^{\boldsymbol{\theta}^T} \tilde{\mathbf{K}}^T \\ &\quad + \tilde{\mathbf{K}} \mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}} \tilde{\mathbf{K}}^T + \mathbf{C}^{\mathrm{e}})^{-1} \\ &\quad + \mathbf{C}_{\boldsymbol{x},\theta} \tilde{\mathbf{K}}^T (\tilde{\mathbf{K}} \mathbf{M}^{\theta} \mathbf{C}^{\mathrm{est}}_{k-1} \mathbf{M}^{\boldsymbol{\theta}^T} \tilde{\mathbf{K}}^T \\ &\quad + \tilde{\mathbf{K}} \mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}} \tilde{\mathbf{K}}^T + \mathbf{C}^{\mathrm{e}})^{-1}. \end{aligned} \tag{B1}$$

Our augmented model error covariance matrix $\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}}$ appears in the gain matrix only as $\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}} \tilde{\mathbf{K}}^T$. Now,

$$\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}} \tilde{\mathbf{K}}^T = \begin{bmatrix} \mathbf{C}_{\boldsymbol{x}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\boldsymbol{\theta}} \end{bmatrix} \begin{bmatrix} \mathbf{K}^T \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\boldsymbol{x}} \mathbf{K}^T \\ \mathbf{0} \end{bmatrix}.$$

This means that the parameter part $\mathbf{C}_{\boldsymbol{\theta}}$ of the model error covariance matrix has no (direct) effect on the gain matrix. Hence, it would be the same as it would be directly inserted to its place in the posterior error covariance matrix. Especially this can be noted in the first round of the state augmentation: the selected matrix $\mathbf{C}_{\boldsymbol{\theta}}$ has no effect on parameters or state.

From the expansion (B1) of $\mathbf{G}$ we can note that only the first term $\mathbf{M}^{\boldsymbol{\theta}} \mathbf{C}^{\mathrm{est}}_{k-1} \mathbf{M}^{\boldsymbol{\theta}^T} \tilde{\mathbf{K}}^T (\tilde{\mathbf{K}} \mathbf{M}^{\boldsymbol{\theta}} \mathbf{C}^{\mathrm{est}}_{k-1} \mathbf{M}^{\boldsymbol{\theta}^T} \tilde{\mathbf{K}}^T + \tilde{\mathbf{K}} \mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}} \tilde{\mathbf{K}}^T + \mathbf{C}^{\mathrm{e}})^{-1}$ affects the parameter part of the gain matrix, since the second term has a multiplier $\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}} \tilde{\mathbf{K}}^T$. In that term the model error term $\mathbf{C}_{\boldsymbol{x},\boldsymbol{\theta}}$ appears only in the inverse, so if we will increase $\mathbf{C}_{\boldsymbol{x}}$ in the experiments, it will cause a smaller rate of change to the parameters. In addition to the inverse, the previous posterior covariance matrix $\mathbf{C}^{\mathrm{est}}_{k-1}$ appears also in the "numerator". Hence, the effect of increasing $\mathbf{C}_{\boldsymbol{\theta}}$ will saturate at some point.

Acknowledgements. Martin Leutbecher from ECMWF is gratefully acknowledged for his support and help in designing the numerical experiments carried out with the stochastic Lorenz-95 system. The research has been supported by the Academy of Finland (project numbers 127210, 132808, 133142 and 134999) and the Centre of Excellence in Inverse Problems Research.

Edited by: W. Hsieh
Reviewed by: two anonymous referees

References

Annan, J. and Hargreaves, J.: Efficient estimation and ensemble generation in climate modelling, Phil. Trans. R. Soc. A, 365, 2077–2088, doi:10.1098/rsta.2007.2067, 2007.

Annan, J. D., Lunt, D. J., Hargreaves, J. C., and Valdes, P. J.: Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter, Nonlin. Processes Geophys., 12, 363–371, doi:10.5194/npg-12-363-2005, 2005.

Auvinen, H., Bardsley, J. M., Haario, H., and Kauranne, T.: Large-Scale Kalman Filtering Using the Limited Memory BFGS Method, Electron. T. Numer. Ana., 35, 217–233, 2009.

Auvinen, H., Bardsley, J., Haario, H., and Kauranne, T.: The variational Kalman filter and an efficient implementation using limited memory BFGS, Int. J. Numer. Meth. Fl., 64, 314–335, 2010.

Bishop, C. M.: Pattern Recognition and Machine Learning, Information Science and Statistics, Springer, New York, 2nd Edn., 2006.

Bonavita, M., Torrisi, L., and Marcucci, F.: The ensemble Kalman filter in an operational regional NWP system: preliminary results with real observations, Q. J. Roy. Meteor. Soc., 134, 1733–1744, doi:10.1002/qj.313, 2008.

Cappe, O., Godsill, S., and Moulines, E.: An overview of existing methods and recent advances in sequential Monte Carlo, Proceedings of IEEE, 95, 899–924, doi:10.1109/JPROC.2007.893250, 2007.

Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., Guillemaud, T., and Estoup, A.: Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation, Bioinformatics, 24, 2713–2719, 2008.

DelSole, T. and Yang, X.: State and Parameter Estimation in Stochastic Dynamical Models, Physica D, 239, 1781–1788, 2010.

Dowd, M.: Estimating parameters for a stochastic dynamic marine ecological system, Environmetrics, 22, 501–515, doi:10.1002/env.1083, 2011.

Evensen, G.: Data assimilation: The ensemble Kalman filter, Springer, 2007.

Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S., and Morneau, J.: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada, Mon. Weather Rev., 135, 2339–2354, 2007.

Gelman, A., Carlin, J., Stern, H., and Rubin, D.: Bayesian Data Analysis, Chapman & Hall, 2nd Edn., 2003.

Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, Bernoulli, 7, 223–242, 2001.

Haario, H., Laine, M., Mira, A., and Saksman, E.: DRAM: Efficient adaptive MCMC, Stat. Comput., 16, 339–354, doi:10.1007/s11222-006-9438-0, 2006.

Houtekamer, P., Herschel, L., Mitchell, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B.: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations, Mon. Weather Rev., 133, 604–620, 2005.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D, 230, 112–126, 2007.

Ionides, E., Breto, C., and King, A.: Inference for nonlinear dynamical systems, Proc. Nat. Aca. Sci., 103, 18438–18443, 2006.

Jackson, C. S., Sen, M. K., Huerta, G., Deng, Y., and Bowman, K. P.: Error Reduction and Convergence in Climate Prediction, J. Climate, 21, 6698–6709, doi:10.1175/2008JCLI2112.1, 2008.

Järvinen, H., Räisänen, P., Laine, M., Tamminen, J., Ilin, A., Oja, E., Solonen, A., and Haario, H.: Estimation of ECHAM5 climate model closure parameters with adaptive MCMC, Atmos. Chem. Phys., 10, 9993–10002, doi:10.5194/acp-10-9993-2010, 2010.

Järvinen, H., Laine, M., Solonen, A., and Haario, H.: Ensemble prediction and parameter estimation system: the concept, Q. J. Roy. Meteor. Soc., doi:10.1002/qj.923, published online, 2011.

Kalman, R.: A New Approach to Linear Filtering and Prediction Problems, Transactions of the ASME – Journal of Basic Engineering, Series D, 82, 35–42, 1960.

Kitagawa, G.: A self-organizing state space model, J. Am. Stat. Assoc., 93, 1203–1215, 1998.

Kivman, G. A.: Sequential parameter estimation for stochastic systems, Nonlin. Processes Geophys., 10, 253–259, doi:10.5194/npg-10-253-2003, 2003.

Laine, M., Solonen, A., Haario, H., and Järvinen, H.: Ensemble prediction and parameter estimation system: the method, Q. J. Roy. Meteor. Soc., published online, doi:10.1002/qj.922, 2011.

Leutbecher, M.: Predictability and Ensemble Forecasting with Lorenz-95 systems, Lecture notes, ECMWF meteorological training course on Predictability, diagnostics and seasonal forecasting, available at: http://www.ecmwf.int/newsevents/training/meteorological_presentations/pdf/PR/Practice_L95.pdf, 2010.

Lorenz, E.: Predictability: A problem partly solved, Proceedings of the Seminar on Predicability, European Center on Medium Range Weather Forecasting, 1, 1–18, 1995.

Marjoram, P. and Tavaré, S.: Modern computational approaches for analysing molecular genetic variation data, Nat. Rev. Genet., 7, 759–770, 2006.

Martin, G. M., Johnson, D. W., and Spice, A.: The measurement and parameterization of effective radius of droplets in warm stratocumulus, J. Atmos. Sci., 51, 1823–1842, 1994.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E.: Equations of State Calculations by Fast Computing Machines, J. Chem. Phys., 21, 1087–1092, 1953.

Molteni, F., Buizza, R., Palmer, T., and Petroliagis, T.: The ECMWF Ensemble Prediction System: Methodology and validation, Q. J. Roy. Meteor. Soc., 122, 73–119, 1996.

Nocedal, G. and Wright, S.: Numerical Optimization, Springer, Berlin, 1999.

Ott, E., Hunt, B., Szunyogh, I., Zimin, A., Kostelich, E., Corazza, M., Kalnay, E., Pati, D., and Yorke, J.: A local ensemble Kalman filter for atmospheric data assimilation, Tellus A, 56, 415–428, 2004.

Palmer, T., Shutts, G., Hagedorn, R., Doblas-Reyes, F., Jung, T., and Leutbecher, M.: Representing Model Uncertainty in Weather and Climate Prediction, Ann. Rev. Earth Planet. Sci., 33, 163–193, 2005.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A.: The ECMWF implementation of four dimensional variational assimilation, Part I: Experimental results with simplified physics, Q. J. Roy. Meteor. Soc., 126, 1143–1170, 2000.

Rougier, J. C.: Efficient Emulators for Multivariate Deterministic Functions, J. Comput. Graphic. Stat., 17, 827–843, 2008.

Sexton, D., Murphy, J., Collins, M., and Webb, M.: Multivariate Prediction Using Imperfect Climate Models Part I: Outline of Methodology, Clim. Dynam., 1–30, doi:10.1007/s00382-011-1208-9, 2011.

Shapiro, A., Dentcheva, D., and Ruszczynski, A.: Lectures on Stochastic Programming: Modeling and Theory, MOS-SIAM Series on Optimization 9, Soc. Ind. Appl. Math., 2009.

Singer, H.: Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood versus extended Kalman filter and Ito-Taylor expansion, J. Comput. Graph. Stat., 11, 972–995, 2002.

Tippett, M., Anderson, J., Bishop, G., Hamill, T., and Whitaker, J.: Ensemble Square Root Filters*, Mon. Weather Rev., 131, 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2, 2003.

Villagran, A., Huerta, G., Jackson, C. S., and Sen, M. K.: Computational Methods for Parameter Estimation in Climate Models, Bayesian Analysis, 3, 823–850, doi:10.1214/08-BA331, 2008.

Whitaker, J. and Hamill, T.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, 2002.

Wilks, D.: Effects of stochastic parametrizations in the Lorenz '96 system, Q. J. Roy. Meteor. Soc., 131, 389–407, doi:10.1256/qj.04.03, 2005.

Zupanski, M.: Maximum Likelihood Ensemble Filter: Theoretical Aspects, Mon. Weather Rev., 133, 1710–1726, 2005.

Publ. III

A dilemma of the uniqueness of weather and climate model closure parameters

Hakkarainen, J., Solonen, A., Ilin, A., Susiluoto, J., Laine, M., Haario, H., and Järvinen, H., *Tellus A 65* (2013), doi:10.3402/tellusa.v65i0.20147.

Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

A dilemma of the uniqueness of weather and climate model closure parameters

By JANNE HAKKARAINEN[1,2]*, ANTTI SOLONEN[1,2], ALEXANDER ILIN[3], JOUNI SUSILUOTO[1,4], MARKO LAINE[1], HEIKKI HAARIO[2] and HEIKKI JÄRVINEN[4], [1]*Finnish Meteorological Institute, Helsinki, Finland;* [2]*Lappeenranta University of Technology, Lappeenranta, Finland;* [3]*Aalto University School of Science, Espoo, Finland;* [4]*University of Helsinki, Helsinki, Finland*

(Manuscript received 21 November 2012; in final form 19 March 2013)

ABSTRACT

Parameterisation schemes of subgrid-scale physical processes in atmospheric models contain so-called closure parameters. Their precise values are not generally known; thus, they are subject to fine-tuning for achieving optimal model performance. In this article, we show that there is a dilemma concerning the optimal parameter values: an identical prediction model formulation can have two different optimal closure parameter value settings depending on the level of approximations made in the data assimilation component of the prediction system. This result tends to indicate that the prediction model re-tuning in large-scale systems is not only needed when the prediction model undergoes a major change, but also when the data assimilation component is updated. Moreover, we advocate an accurate albeit expensive method based on so-called filter likelihood for the closure parameter estimation that is applicable in fine-tuning of both prediction model and data assimilation system parameters. In this article, we use a modified Lorenz-95 system as a prediction model and extended Kalman filter and ensemble adjustment Kalman filter for data assimilation. With this setup, we can compute the filter likelihood for the chosen parameters using the output of the two versions of the Kalman filter and apply a Markov chain Monte Carlo algorithm to explore the parameter posterior distributions.

Keywords: model tuning, Markov chain Monte Carlo, likelihood, filter formulation

1. Introduction

Long-term improvements in atmospheric general circulation models (GCMs) used in numerical weather prediction and climate simulations originate from gradually refined representations of atmospheric phenomena, especially of the subgrid-scale physical processes. In these models, the processes are represented by parameterisation schemes where the subgrid-scale variability is expressed using model variables of the resolved scales. Moreover, the schemes contain so-called closure parameters. The purpose of these model parameters is to encapsulate some atmospheric processes or properties which are not affordable for explicit modelling. The development process of the physical parameterisation schemes involves basically two steps: derivation of individual parameterisation schemes, and constraining of the schemes with large-scale observations (Lohmann et al., 2007). Here, large-scale observations means observations that are available for model tuning at the GCM scale. In the first step, dedicated laboratory measurements or focused field measurements campaigns are used, if a scheme cannot be derived from the first principles. The second step accounts for the fact that individual schemes are often valid for small scales, while in the context of GCMs, the schemes are applied in a relatively coarse resolution over a wide range of spatial and temporal scales. Thus, the schemes of interest have to be tuned such that the model indeed simulates the key atmospheric quantities within the uncertainty of some large-scale observational constraint. In practice, the closure (or, tuning) parameters provide the necessary degrees-of-freedom so that a realistic model response can be obtained.

In GCMs, a comprehensive set of physical parameterisation schemes are assembled together and embedded in the solver of the atmospheric dynamics. Time-space truncation and numerical approximations render, however, all

*Corresponding author.
email: janne.hakkarainen@fmi.fi

Citation: Tellus A 2013, **65**, 20147, http://dx.doi.org/10.3402/tellusa.v65i0.20147

model components to some extent imperfect. These model imperfections lead to systematic and random simulation errors. The systematic errors appear as a climate drift in the model simulations. The geographical patterns of these errors tend to develop very early in the simulation (Jung, 2005), and the so-called systematic initial tendency errors can be used for diagnosis and attribution of model errors to some particular atmospheric processes (Klinker and Sardeshmukh, 1992). In fact, there are attempts to estimate the initial tendency model errors and thus parameterise unrepresented physical phenomena in order to reduce systematic model errors (Kaas et al., 1999; D'Andrea and Vautard, 2000). Clearly, in the model simulations, there are multitudes of errors with different sources and the errors are in mutual non-linear interaction. Thus, it is not a surprise that it is very hard to improve the general model performance by improving the realism of representation of individual physical parameterisations (Jakob, 2010). Therefore, for a given model formulation, it is natural to strive towards optimal predictive skill by tuning the available degrees-of-freedom, that is, the closure parameter values. The estimation techniques range from trial-and-error to sophisticated parameter estimation techniques (e.g. Severijns and Hazeleger, 2005; Kunz et al., 2008; Järvinen et al., 2010; Neelin et al., 2010). The closure parameter values corresponding to the optimal predictive skill naturally depend on the level of approximation of the prediction system.

There is thus a dilemma on the uniqueness of weather and climate model closure parameters: one can either anchor the parameter values to the observable truth obtained, for example, via measurements from focused field campaigns, or tune the parameter values to account for the imperfections in the prediction system. It is qualitatively obvious that these two do not necessarily coincide and a practical optimum is a compromise between the two. In this article, we demonstrate that this dilemma exists. An outline of the demonstration is as follows. We use a low-order prediction system where synthetic observations are created using a modified version of the Lorenz-95 model (Lorenz, 1995; Wilks, 2005). As a prediction model, we use the standard Lorenz-95 model where the subgrid-scale effects on resolved scales are represented by a linear bulk parameterisation containing two closure parameters. Thus, the prediction model is imperfect. Data assimilation of the synthetic observations is performed using either extended Kalman filter (EKF) (Kalman, 1960) or ensemble adjustment Kalman filter (EAKF) (Anderson, 2001). We observe that the optimal prediction model corresponds to a unique but different parameter setting depending on the choice of the data assimilation component. The conclusion is that the dilemma exists.

Since the optimal closure parameter values depend on the implementation of the prediction system, such as the choice of the data assimilation algorithm, methodology for tuning the parameters specifically for each prediction system is needed. Here, we present such a method, using a combination of the so-called filter likelihood approach, where the likelihood is formulated based on the output of the filtering methods (Singer, 2002; Hakkarainen et al., 2012) and a Markov chain Monte Carlo algorithm (Haario et al., 2001, MCMC). We show that the approach yields parameter values that are close to the ones corresponding to optimal filter accuracy. In addition, we demonstrate how the approach can be implemented via the Data Assimilation Research Testbed software environment (Anderson et al., 2009, DART).

2. Methods and experimentation

In this section, an introduction to parameter estimation using the filter likelihood approach is given. It is then tested using the modified Lorenz-95 model.

2.1. On state estimation methods

One of the most common techniques for state estimation is Kalman filtering (Kalman, 1960). The basic idea in all filters is to obtain the posterior state estimate x_k^{est} and some error statistics, typically the error covariance matrix C_k^{est}, given the prior information x_k^{p} and the observations y_k at time step k.

The posterior state estimate can be transported to the next filtering time step to become the new prior state by using the forecast model. For the state vector, this is usually straightforward, but the fundamental question in filtering is how to transport the uncertainty in time. In EKF, this problem is solved by calculating the prior covariance matrix C_k^p at time k as

$$C_k^p = M_k C_{k-1}^{\text{est}} M_k^{\text{T}} + Q_k, \tag{1}$$

where Q_k is the model error covariance matrix and M_k is the tangent-linear forecast model. If the dimension of the state space is high, such as in large-scale atmospheric models, eq. (1) becomes impossible to compute without approximations.

One approximative approach is ensemble filtering, where the uncertainty is transported to the next filtering time step by transporting an ensemble of states using the forecast model. The difficulty (and variety) of the ensemble methods lies in how to update the prior ensemble in the analysis step. There are many alternatives, for instance the ensemble adjustment Kalman filter (Anderson, 2001, 2010), which belongs to the family of the so-called ensemble square root

filters (Tippett et al., 2003). In this study, the EAKF was chosen because it is a *deterministic* ensemble based filter, that is, it does not include randomisation of the ensembles. The non-randomness is important, since in Section 2.3 we want to create a likelihood function based on the filtering outputs and the randomness in the likelihood would complicate the parameter inference (Dowd, 2011; Hakkarainen et al., 2012). Also, EAKF was chosen because it is conveniently available in the data assimilation research testbed (Anderson et al., 2009, DART) and widely applied (e.g. Schwartz et al., 2011; Torn and Davis, 2012).

In EKF, the modelling errors of the forecast model are taken into account via the error covariance matrix Q_k. In EAKF (and in ensemble square root filtering in general), the neglect of unknown model errors can lead to overconfidence in the prior state and hence ignorance of the observations in the analysis, which can lead to filter divergence. This is effectively circumvented by introducing covariance inflation factors (Anderson and Anderson, 1999). In addition, the sampling errors related to the limited ensemble size (undersampling), are mostly removed by localisation (Hamill et al., 2001), where observations affect only nearby grid points.

2.2. On parameter estimation

Parameter estimation techniques developed for the atmospheric models can be divided in two categories: *online* and *offline* methods. In online methods, it is assumed that model parameters are not static quantities, but can evolve adaptively, for example, as a part of a state estimation process (e.g. Annan et al., 2005) and are thus computationally relatively cheap to implement. Typically, a statistical interpretation is lacking in online methods, since the rate of change of the parameters is controlled by the user. In this article, these online methods are not discussed further. Instead we use offline methodology, where a predetermined (training) set of observations y is used for tuning the parameter values.

In Bayesian methodology (e.g. Gelman et al., 2003), the knowledge about the unknown parameters is inferred from the posterior distribution:

$$p(\theta|y) \propto p(\theta)p(y|\theta), \tag{2}$$

which is evaluated using the prior $p(\theta)$ and the likelihood $p(y|\theta)$. The prior contains the information that we have about the parameters based on the accumulated information from the past. The likelihood function specifies how plausible the observed data are given model parameter values. Therefore, defining a proper likelihood function is the central question in parameter estimation.

The parameter estimation problem is often written in a (cost) functional form with parameters as arguments. Different numerical methods can be used for optimising the cost function and point estimates, such as the Maximum a Posteriori (MAP) estimate, can be obtained. In addition, Monte Carlo methods like MCMC can be used for producing samples from the posterior distribution instead of a single point estimate, and studying the uncertainty in the parameter values. For large-scale applications, applying MCMC is challenging, since the methods can involve thousands of repeated likelihood evaluations, but ways to improve the efficiency of MCMC for high-CPU models have been recently developed (Solonen et al., 2012).

2.3. Likelihood via filtering methods

In this section, the filter likelihood concept is discussed in the context of the extended Kalman filter and the ensemble adjustment Kalman filter. The approach is known in the parameter estimation of stochastic models (e.g. Singer, 2002), but less studied in connection with deterministic, chaotic systems (Hakkarainen et al., 2012). For the sake of completeness, the derivation of the likelihood computation is briefly reviewed in the Appendix.

When EKF is considered, the filter likelihood formula, i.e., the likelihood for observing $y_{1:n}$ given the parameters θ, can be written as

$$\begin{aligned} p(y_{1:n}|\theta) &= p(y_1|\theta)\prod_{k=2}^{n} p(y_k|y_{1:k-1},\theta) \\ &= \prod_{k=1}^{n} \exp\left(-\frac{1}{2} r_k^T (C_k^y)^{-1} r_k\right)(2\pi)^{-d/2}|C_k^y|^{-1/2} \\ &\propto \exp\Bigg(-\frac{1}{2}\sum_{k=1}^{n} \underbrace{\left[r_k^T (C_k^y)^{-1} r_k + \log|C_k^y|\right]}_{=J_k}\Bigg), \end{aligned} \tag{3}$$

where $r_k = y_k - \mathcal{H}(x_k^p)$ is the innovation vector and $C_k^y = H_k C_k^p H_k^T + R_k$ is its error covariance matrix at time k. Operator $|\cdot|$ denotes the matrix determinant. In filtering, $\mathcal{H}$ is the observation operator that maps from the state space to the observation space and H is its linearization. The prior state x_k^p is given by the model, the prior covariance matrix C_k^p is given by the Kalman filtering formulas [eq. (1)] and R_k is the observation error covariance matrix.

Essentially, the filter likelihood formula [eq. (3)] depends on the summation over the cost functions J_k, where the normalising term $\log|C_k^y|$ has to be accounted for, because it implicitly depends on the parameters θ. The dependence comes from eq. (1) via the tangent-linear forecast model. It should be noted that eq. (3) itself does not require that x_k^p and its covariance matrix C_k^p are calculated using EKF. In principle, any state estimation method that produces these prior estimates could be used.

In EAKF, as implemented in the DART system, the discrepancy between the prior ensemble and the observations are calculated in the observation space for every single observation sequentially. The update of the ensemble and the weights are also calculated sequentially in the observation space. Thus, it is natural to adopt this approach also in the filter likelihood calculations. The cost function part of the filter likelihood formula [eq. (3)] can be calculated sequentially as a sum

$$J_k = \sum_{l=1}^{L} \frac{(y_l - \overline{z_l^p})^2}{(\sigma_{y_l}^2 + \sigma_{z_l^p}^2)} + \log(\sigma_{y_l}^2 + \sigma_{z_l^p}^2), \tag{4}$$

where y_l is an individual observation and $\overline{z_l^p}$ is the individual prior ensemble mean in observation space and $\sigma_{y_l}^2$ and $\sigma_{z_l^p}^2$ are their variances, respectively. The prior ensemble mean $\overline{z_l^p}$ depends implicitly on y_{l-1}, because of the sequential update. The summation goes through all the individual observations $y_{1:L,k}$ at time k. The DART system does not provide the above computation, but including these in the DART code can be easily implemented.

2.4. Model and data

To explore the filter likelihood calculation, a modified version of the Lorenz-95 system is used (Lorenz, 1995; Wilks, 2005). The full system consists of two interrelated variables X and Y, and it is written as

$$\frac{dX_i}{dt} = -X_{i-1}(X_{i-2} - X_{i+1}) - X_i + F - \frac{hc}{b} \sum_{j=J(i-1)+1}^{Ji} Y_j \tag{5}$$

$$\frac{dY_j}{dt} = -cbY_{j+1}\left(Y_{j+2} - Y_{j-1}\right) - cY_j + \frac{c}{b}F_Y + \frac{hc}{b}X_{1+\lfloor \frac{j-1}{J} \rfloor} \tag{6}$$

where $i = 1,\ldots, I$ and $j = 1,\ldots, JI$. That is, each of the 'slow' state variables X_i are forced by a sum of the additional fast variables Y_j. The fast variables have dynamics similar to the slow variables, but they are also coupled with the slow variables. In the model, cyclic boundary conditions, i.e., $X_{I+1} = X_1$ and $Y_{JI+1} = Y_1$, are used. In this paper, values $I = 40$, $J = 8$, $F = F_Y = 10$, $h = 1$ and $c = b = 10$, adopted from Leutbecher (2010), are used.

The parametrized forecast model for the full system reads

$$\frac{dX_i}{dt} = -X_{i-1}(X_{i-2} - X_{i+1}) - X_i + F - g(X_i, \theta), \tag{7}$$

where $g(X_i, \theta) = \theta_0 + \theta_1 X_i$ is the parameterization in which the effect of the missing 'sub-grid' scale fast variables Y_j are modeled using only the single, 'resolved', local variable X_i.

The aim of the experiment is to tune the parameters $\theta = (\theta_0,\ \theta_1)$ given a synthetic set of observations $y_{1:n}$, computed by the full set of equations in (5) and (6) with additive noise, using the filter likelihood technique. In the experiment, one filtering step ('24 h') is 8 model integration steps ('3 h'). Here the number of assimilation steps $n = 100$ ('days'), and hence during each run the model is integrated 100×8 times. The experiment is adopted from Hakkarainen et al. (2012) and a more comprehensive introduction is given there.

2.5. Experiment setup

In the experiment, the parameter values are sampled from the posterior distribution {consisting of filter likelihood [eq. (3)] and uninformative prior} using an adaptive Metropolis algorithm (Haario et al., 2001). The chain length is selected to be 3000 and the first 500 samples (the burn-in period) are discarded. The filter likelihood MCMC experimentation is synthesised with the following pseudo-algorithm, where the superscript m on θ denotes MCMC chain index:

Step 0:
Initialise the MCMC run and select θ^1.
Step 1:
Propose a new candidate $\hat{\theta}$ for the parameters.
Step 2:
Run the state estimation system using $\hat{\theta}$ and evaluate formula [eq. (3)] during the run.
Step 3:
Accept $\hat{\theta}$ with probability $\min\left(1, \frac{p(y_{1:n}|\widehat{\theta})p(\widehat{\theta})}{p(y_{1:n}|\theta^m)p(\theta^m)}\right)$.
Step 4:
If $\hat{\theta}$ is accepted set $\theta^{m+1} = \hat{\theta}$, else set $\theta^{m+1} = \theta^m$.
Step 5:
Set $m = m+1$ and go to step 1 until m is equal to the chain length.

In our implementation, we use an uninformative (flat) prior, which means that the prior is a constant, $p(\theta)\ \alpha 1$. Then, the prior term cancels out in Step 3 of the above algorithm, since $p(\widehat{\theta})/p(\theta^m) = 1$.

For readers not familiar with MCMC, a short introduction is in order. The MCMC algorithm works by generating candidate values from a *proposal distribution*, which in our case is Gaussian centred at the current parameter value. The candidates are then either accepted or rejected according to a simple rule that contains the ratio of the posterior densities at the proposed point and at the current point, see Step 3 of the above algorithm. One can see that moves 'upward' (to a point with higher posterior density) are always accepted. However, also moves downward can be accepted, and the resulting random walk ends up exploring the posterior distribution without converging to a single point. The algorithm is a standard tool in statistical

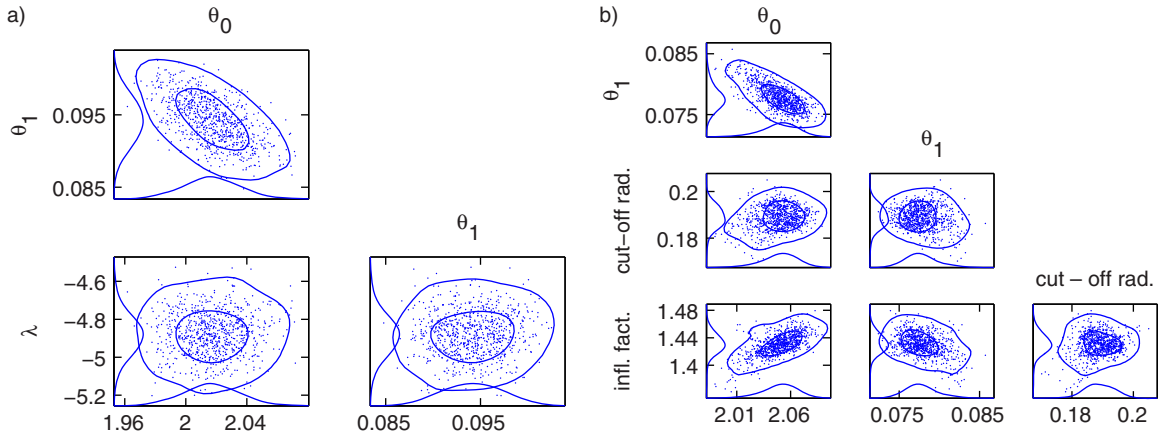


Fig. 1. Pairwise MCMC samples using EKF (left panel) and EAKF (right panel) version of the filter likelihood formula [eq. (3)]. In addition to the model parameters θ_0 and θ_1, filter related parameters are sampled too. Contours illustrate 95% and 50% probability regions. Also, the one-dimensional marginal densities are illustrated.

analysis and can be shown to yield samples from the correct target density. Here, we apply an adaptive version of the algorithm Haario et al. (2001), where the proposal distribution is automatically tuned to better match the posterior distribution. For more details on MCMC, see Robert and Casella (2005).

Using the filter likelihood [eq. (3)] for MCMC, we are able to estimate both the model parameters θ and the filter-specific tuning parameters. In the experiments, for the EKF version, the model error covariance matrix is parameterised as follows

$$Q = \exp(\lambda) I, \tag{8}$$

where I is an identity matrix. The parameter λ is tuned together with the model parameters θ_1 and θ_1. In the EAKF version of the filter likelihood, the prior covariance inflation factor and the cut-off radius related to localisation (Gaspari and Cohn, 1999), defined in the DART setup, are tuned together with the model parameters θ.

As a validation metric for the parameters, we use the average root mean squared error (RMSE) defined as

$$r(\theta) = \frac{1}{n}\sum_{k=1}^{n}\sqrt{\frac{1}{I}\sum_{i=1}^{I}(x_{i,k}^{\mathrm{est}}(\theta) - x_{i,k}^{\mathrm{true}})^2} \tag{9}$$

where x_k^{true} is the true state and x_k^{est} is the posterior estimate. We evaluate RMSE on a fixed grid for θ_1 and θ_2 keeping the filter-specific parameters fixed.

2.6. Results

The experiments are started by a step, where all parameters (i.e. filter parameters and the model parameters) are estimated together with MCMC. In Fig. 1, pairwise MCMC samples using EKF (left panel) and EAKF (right panel) are shown. It can be seen, that the marginal distributions are rather Gaussian and all parameters are identified. Based on the mean values of the MCMC chains, values -4.8824, 1.4310 and 0.1889 were chosen for λ, the prior inflation factor and the cut-off radius, respectively.

In the second step, to further underline the difference in the model parameter estimates in the usual situation where the filter parameters are fixed, we calculate the conditional θ distributions using MCMC by fixing the filter parameters to the mean values found in the previous step. For the validation purposes, the RMS error [eq. (9)] is evaluated on a fixed grid to explore the parameter space for 'all' possible pairs of θ_0 and θ_1.

The results are illustrated in Fig. 2, where the 50% and 95% probability contours calculated from the EKF and EAKF MCMC chains are plotted using solid and dashed lines, respectively. The colours on the background indicate the average RMSE values [eq. (9)] for EKF (left panel) and EAKF (right panel). We observe that the closure parameter values that correspond to the lowest average RMSE values for EKF and EAKF are not the same. This implies that the optimal parameter values for EKF are not optimal for EAKF and vice versa. This shows that there is a dilemma of the uniqueness of the model closure parameters: an identical model formulation can have more than one optimal parameter value combination. This can be explained by the fact that the two data assimilation components have different systematic approximation errors, and these errors are compensated, to some extent, by the different parameter values in the forecast model. In this sense, the forecast model parameter values should be considered as filter-dependent quantities. The consequences of this result are discussed below in Section 3.

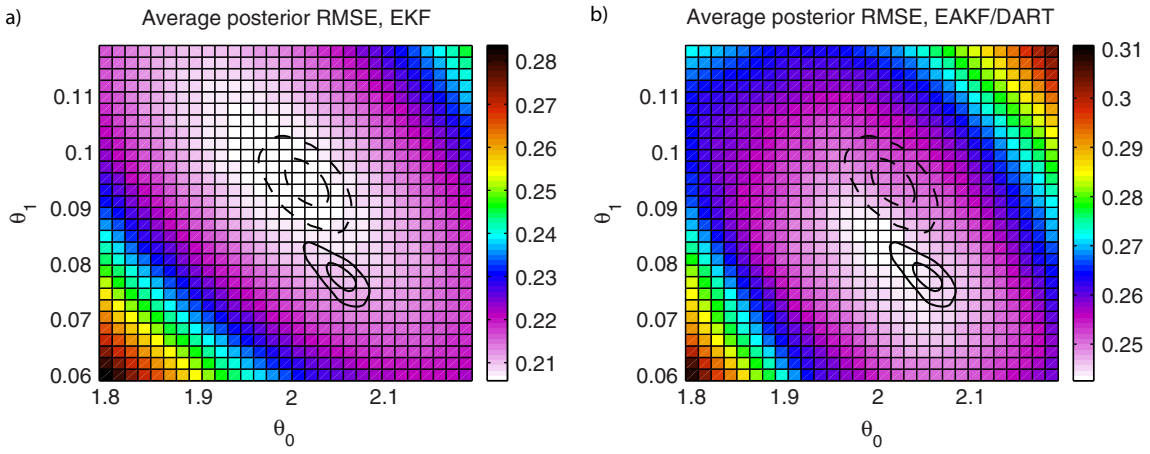


Fig. 2. Average posterior RMSE values [eq. (9)] when using EKF (left panel) and EAKF (right panel). Lower values indicate better performance. Solid and dashed lines indicate the 50% and 95% probability regions as obtained with MCMC in case of EAKF and EKF, respectively. RMSE values are depicted by the background colour scheme. Note the difference in colour scales.

Figure 2 also depicts that the filter likelihood method captures the optimal parameter values correctly: the lowest posterior RMSE values are close to the obtained posterior distributions. In addition, it can be seen that the results produced by the EKF and EAKF versions of the filter likelihood technique are similar, but not identical. The posterior distribution obtained using EKF is closer to the parameter values that yield low RMS error for EKF. And vice versa: the parameters estimated using EAKF are closer to the optimal values for EAKF than the ones obtained by EKF. That is, applying the filter likelihood technique seems to provide rather good filter-specific parameter estimates. In this study, the RMSE values produced by EKF are lower than those produced by EAKF (see the RMSE values at the background of Fig. 2), which was expected. In addition, it can be noted, that the choice of data assimilation component has a much larger effect on the RMSE than the exact values of the model parameters.

3. Discussion and conclusions

In this article, we presented a numerical experiment with the purpose to simulate the prediction model parameter estimation problem in chaotic systems. We studied a simplified prediction system, which consists of a modified Lorenz-95 system as the prediction model, where the net effect of fast variables is represented by a two-parameter linear scheme. Two versions of the Kalman filter were used as different data assimilation components. We have shown that depending on the version of data assimilation that is applied, either the extended Kalman filter or the ensemble adjustment Kalman filter, the optimal prediction model tuning parameters corresponding to the highest system accuracy have two distinct optima. Accuracy is measured here as the RMS errors of the state estimate against the known truth. Replacement of EKF with EAKF implies a major change to the prediction system and it is only natural that the prediction model parameters need re-tuning.

In addition, we have demonstrated a method that can be used to estimate the closure parameters for each prediction system component separately. The approach is based on computing the likelihood using the output of the data assimilation system. We also demonstrate how Markov chain Monte Carlo sampling methods can be used on top of this likelihood formulation. The method recovers the optimal forecast model parameter values for both data assimilation systems. In addition, both the forecast model parameters and the data assimilation related parameters can be estimated. In our case, the data assimilation related parameters were the model error covariance matrix in EKF, and covariance inflation and localisation parameters in EAKF.

The models applied in operational weather forecasting are far more complex than the system applied here. If we assume that the behaviour discovered here scales up, it has some consequences for tuning of large-scale systems. Traditionally, model releases are carefully fine-tuned and parallel tested associated with forecast model changes. Changes in data assimilation are, however, typically considered independent of forecast model fine-tuning. According to the result presented here, this should not be the case. If the data assimilation component undergoes a major revision but the prediction model remains the same, then the prediction model tuning corresponds to the previous

version of the data assimilation component and may thus be suboptimal. In this case, the correct procedure would be to consider model tuning, too.

Finally, this work considered a low-order prediction system. Our future work will be directed towards quantifying how this result scales-up to more realistic systems. We will use the ECHAM6 prediction model together with the DART toolbox to study this question.

4. Acknowledgements

We would like to acknowledge the National Center for Atmospheric Research for making the Data Assimilation Research Testbed available on-line. The research has been financially supported by the Academy of Finland (project numbers 127210, 132808, 133142, 134935 and 134999) and the Centre of Excellence in Inverse Problems Research. Additionally, this work was funded by the European Commission's 7th Framework Programme, under Grant Agreement number 282672, EMBRACE project, and the Nessling Foundation.

5. Appendix

A.1. Derivation of the filter likelihood formula

We present the derivation of the filter likelihood formula [eq. (3)], briefly. A more comprehensive discussion is given in Hakkarainen et al. (2012).

Let us consider the following general state space model.

$$x_k \sim p(x_k|x_{k-1}, \theta) \tag{A1}$$

$$y_k \sim p(y_k|x_k) \tag{A2}$$

$$\theta \sim p(\theta) \tag{A3}$$

of state x_k, observations y_k and parameters θ, respectively. Our goal is to find the posterior distribution $p(\theta|y_{1:n})$. In filtering, the previous state estimate can be transported to the next time step's prior by using the prediction model

$$p(x_k|y_{1:k-1}, \theta) = \int p(x_k|x_{k-1}, \theta) \times p(x_{k-1}|y_{1:k-1}, \theta) dx_{k-1} \tag{A4}$$

and in the analysis step the prior is updated by y_k using the Bayes' formula

$$p(x_k|y_{1:k}, \theta) \propto p(y_k|x_k, \theta) p(x_k|y_{1:k-1}, \theta). \tag{A5}$$

The predictive distribution for next observations can be obtained by

$$p(y_k|y_{1:k-1}, \theta) = \int p(y_k|x_k, \theta) p(x_k|y_{1:k-1}, \theta) dx_k. \tag{A6}$$

Using Bayes' formula and the chain rule we obtain

$$\begin{aligned} p(\theta|y_{1:n}) \propto p(y_{1:n}|\theta)p(\theta) = {} & p(y_n|y_{1:n-1}, \theta) \\ & \times p(y_{n-1}|y_{1:n-2}, \theta) \ldots p(y_2|y_1, \theta) \\ & \times p(y_1|\theta)p(\theta), \end{aligned} \tag{A7}$$

where $p(y_k|y_{1:k-1}, \theta)$ can be calculated using eq. (A6).

In extended Kalman filtering the predictive distribution is

$$y_k|y_{1:k-1}, \theta \sim N(\mathcal{H}(x_k^p), C_k^y). \tag{A8}$$

Now, applying this to eq. (A8) we can obtain eq. (3).

References

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N. and co-authors. 2009. The data assimilation research testbed: a community facility. *B. Am. Meteorol. Soc.* **90**(9), 1283–1296. DOI: 10.1175/2009BAMS2618.1.

Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**(12), 2884–2903. DOI: 10.1175/1520-0493(2001)129 <2884:AEAKFF >2.0.CO.

Anderson, J. L. 2010. A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.* **138**(11), 4186–4198. DOI: 10.1175/2010MWR3253.1.

Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**(12), 2741–2758. DOI: 10.1175/1520-0493(1999)127 <2741:AMCIOT >2.0.CO.

Annan, J. D., Lunt, D. J., Hargreaves, J. C. and Valdes, P. J. 2005. Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. *Nonlin. Processes Geophys.* **12**(3), 363–371. DOI: 10.5194/npg-12-363-2005.

D'Andrea, F. and Vautard, R. 2000. Reducing systematic errors by empirically correcting model errors. *Tellus A* **52**(1), 21–41. DOI: 10.1034/j.1600-0870.2000.520103.x.

Dowd, M. 2011. Estimating parameters for a stochastic dynamic marine ecological system. *Environmetrics* **22**(4), 501–515. DOI: 10.1002/env.1083.

Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteorol. Soc.* **125**(554), 723–757. DOI: 10.1002/qj.49712555417.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. 2003. *Bayesian Data Analysis*. 2nd ed. Chapman & Hall, Boca Raton, FL.

Haario, H., Saksman, E. and Tamminen, J. 2001. An adaptive Metropolis algorithm. *Bernoulli* **7**(2), 223–242.

Hakkarainen, J., Ilin, A., Solonen, A., Laine, M., Haario, H. and co-authors. 2012. On closure parameter estimation in chaotic systems. *Nonlin. Processes Geophys.* **19**(1), 127–143. DOI: 10.5194/npg-19-127-2012.

Hamill, T. M., Whitaker, J. S. and Snyder, C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**(11), 2776–2790. DOI: 10.1175/1520-0493(2001)129 <2776:DDFOBE >2.0.CO.

Jakob, C. 2010. Accelerating progress in global atmospheric model development through improved parameterizations: challenges, opportunities, and strategies. *B. Am. Meteorol. Soc.* **91**(7), 869–875. DOI: 10.1175/2009BAMS2898.1.

Järvinen, H., Räisänen, P., Laine, M., Tamminen, J., Ilin, A. and co-authors. 2010. Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmos. Chem. Phys.* **10**(20), 9993–10002. DOI: 10.5194/acp-10-9993-2010.

Jung, T. 2005. Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Quart. J. Roy. Meteorol. Soc.* **131**(607), 1045–1073. DOI: 10.1256/qj.04.93.

Kaas, E., Guldberg, A., May, W. and Déqué, M. 1999. Using tendency errors to tune the parameterisation of unresolved dynamical scale interactions in atmospheric general circulation models. *Tellus A* **51**(5), 612–629. DOI: 10.1034/j.1600-0870.1999.00006.x.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng. Ser. D.* **82**, 35–42.

Klinker, E. and Sardeshmukh, P. D. 1992. The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.* **49**(7), 608–627. DOI: 10.1175/1520-0469(1992)049 <0608:TDOMDI >2.0.CO.

Kunz, T., Fraedrich, K. and Kirk, E. 2008. Optimisation of simplified GCMs using circulation indices and maximum entropy production. *Clim. Dynam.* **30**, 803–813. DOI: 10.1007/s00382-007-0325-y.

Leutbecher, M. 2010. Predictability and ensemble forecasting with Lorenz-95 systems, Lecture notes. ECMWF meteorological training course on Predictability, diagnostics and seasonal forecasting. Online at: http://www.ecmwf.int/newsevents/training/meteorological_presentations/pdf/PR/Practice_L95.pdf

Lohmann, U., Quaas, J., Kinne, S. and Feichter, J. 2007. Different approaches for constraining global climate models of the anthropogenic indirect aerosol effect. *B. Am. Meteorol. Soc.* **88**(2), 243–249. DOI: 10.1175/BAMS-88-2-243.

Lorenz, E. N. 1995. Predictability: a problem partly solved. In: *Proceedings of the Seminar on Predicability, European Center on Medium Range Weather Forecasting* **1**, 1–18. Online at: http://www.ecmwf.int/publications/library/do/references/show?id=87423

Neelin, J. D., Bracco, A., Luo, H., McWilliams, J. C. and Meyerson, J. E. 2010. Considerations for parameter optimization and sensitivity in climate models. *Proc. Natl. Acad. Sci.* **107**(50), 21349–21354. DOI: 10.1073/pnas.1015473107.

Robert, C. P. and Casella, G. 2005. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Secaucus, NJ.

Schwartz, C. S., Liu, Z., Chen, Y. and Huang, X.-Y. 2011. Impact of assimilating microwave radiances with a limited-area ensemble data assimilation system on forecasts of typhoon morakot. *Wea. Forecasting* **27**(2), 424–437. DOI: 10.1175/WAF-D-11-00033.1.

Severijns, C. A. and Hazeleger, W. 2005. Optimizing parameters in an atmospheric general circulation model. *J. Climate* **18**(17), 3527–3535. DOI: 10.1175/JCLI3430.1.

Singer, H. 2002. Parameter estimation of nonlinear stochastic differential equations: simulated maximum likelihood versus extended Kalman filter and Itô–Taylor expansion. *J. Comput. Graph. Stat.* **11**, 972–995. DOI: 10.1198/106186002808.

Solonen, A., Ollinaho, P., Laine, M., Haario, H., Tamminen, J. and co-authors. 2012. Efficient MCMC for climate model parameter estimation: parallel adaptive chains and early rejection. *Bayesian Anal.* **7**(3), 715–736. DOI: 10.1214/12-BA724.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**(7), 1485–1490. DOI: 10.1175/1520-0493(2003)131 <1485:ESRF >2.0.CO.

Torn, R. D. and Davis, C. A. 2012. The influence of shallow convection on tropical cyclone track forecasts. *Mon. Wea. Rev.* **140**(7), 2188–2197. DOI: 10.1175/MWR-D-11-00246.1.

Wilks, D. 2005. Effects of stochastic parametrizations in the Lorenz 96 system. *Quart. J. Roy. Meteorol. Soc.* **131**(606), 389–407. DOI: 10.1256/qj.04.03.

Publ. IV

Direct comparisons of GOMOS and SAGE III NO_3 vertical profiles

Hakkarainen, J., Tamminen, J., Moore, J. R., and Kyrölä, E., *Atmos. Meas. Tech. 5*, 7 (2012), 1841–1846, doi:10.5194/amt-5-1841-2012.

Atmos. Meas. Tech., 5, 1841–1846, 2012
www.atmos-meas-tech.net/5/1841/2012/
doi:10.5194/amt-5-1841-2012

Atmospheric Measurement Techniques

Direct comparisons of GOMOS and SAGE III NO_3 vertical profiles

J. Hakkarainen[1,2], J. Tamminen[1], J. R. Moore[3], and E. Kyrölä[1]

[1]Finnish Meteorological Institute, Earth Observation, Helsinki, Finland
[2]Lappeenranta University of Technology, Lappeenranta, Finland
[3]Science Systems and Applications, Inc., Hampton, Virginia, USA

Correspondence to: J. Hakkarainen (janne.hakkarainen@fmi.fi)

Received: 8 December 2011 – Published in Atmos. Meas. Tech. Discuss.: 13 February 2012
Revised: 21 June 2012 – Accepted: 10 July 2012 – Published: 31 July 2012

Abstract. In this paper, we present the first global comparisons between the two unique satellite-borne data sets of nitrogen trioxide (NO_3) vertical profiles retrieved from the GOMOS (Global Ozone Monitoring by the Occultation of Stars) stellar occultations and the SAGE III (Stratospheric Aerosol and Gas Experiment) lunar occultations. The comparison results indicate that between the altitudes 25 km and 45 km the median difference between these two data sets is within ±25 %. The study of zonal median profiles shows that the climatologies calculated from GOMOS and SAGE III profiles are comparable and represent the same features in all latitude bands. No clear systematic differences are observed. The median profiles are closest in the tropics and slightly deviating at high latitudes.

1 Introduction

The radical nitrate NO_3 is important in the stratospheric nighttime photochemistry. It is chemically coupled to nitrogen oxides ($NO_x = NO + NO_2$), whose reactions in the middle atmosphere form the primary catalytic ozone destruction cycle (Marchand et al., 2004). NO_3 has strong diurnal variations, and during sunrise and sunset photolysis destroys NO_3 extremely quickly in the presence of sunlight. During the nighttime, in the absence of heterogeneous processes, the NO_3 chemistry scheme is believed to be relatively simple with three reactions:

$$NO_2 + O_3 \rightarrow NO_3 + O_2 \quad \text{(R1)}$$

$$NO_3 + NO_2 + M \rightarrow N_2O_5 + M \quad \text{(R2)}$$

$$N_2O_5 + M \rightarrow NO_3 + NO_2 + M \quad \text{(R3)}$$

NO_3 is mainly produced by Reaction (R1) of NO_2 and O_3. The sink of NO_3 is the Reaction (R2) with NO_2 which produces N_2O_5, which during the polar winter and spring reacts on the surface of stratospheric sulphate aerosol heterogeneously to form HNO_3 and polar stratospheric clouds (Amekudzi et al., 2005). The thermal decomposition of N_2O_5 (Reaction R3) is an additional source of NO_3 (Marchand et al., 2004).

Historically, NO_3 has been observed by ground-based lunar measurements, and the first measurements of NO_3 were published by Noxon et al. (1978). In addition to ground-based measurements, balloon-borne measurements to observe the vertical structures have been made using stellar and lunar occultations (Naudet et al., 1981; Renard et al., 1996). Recently, NO_3 slant-column densities have been observed through sunrise and sunset using limb-scattered solar light measured by OSIRIS (McLinden and Haley, 2008).

Due to the strong diurnal variation of NO_3, in practice existing only during nighttime and being undetectable during daytime, there are only a few data sets of satellite-borne NO_3 profiles available. The GOMOS (Global Ozone Monitoring by the Occultation of Stars) instrument has provided a long data set of simultaneous NO_2 and NO_3 observations since August 2002 (Hauchecorne et al., 2005; Kyrölä et al., 2010a). While GOMOS uses stellar light as a light source to measure the vertical distributions in the atmosphere, SAGE III (Stratospheric Aerosol and Gas Experiment) and SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY) employ the lunar occultation technique to observe NO_3.

The strong diurnal variation makes the validation challenging, and many previous validations of the satellite-borne NO_3 profiles include models and/or a chemical data

assimilation schemes (e.g. Marchand et al., 2004; Amekudzi et al., 2007). In the study of Marchand et al. (2004), the self-consistency of GOMOS NO_3, NO_2 and O_3 was verified. Their results also indicated that there is no substantial bias in GOMOS NO_3 data. In the study of Renard et al. (2008), four GOMOS NO_3 profiles were compared against balloon-borne observations as a "one-shot" validation exercise.

To our knowledge, this is the first paper where the satellite-borne NO_3 profiles are directly compared with each other.

2 Data sets and methods

2.1 GOMOS data

The GOMOS (Global Ozone Monitoring by the Occultation of Stars) instrument was launched on 1 March 2002 by the European Space Agency on board the ENVISAT platform (Bertaux et al., 2010). Since August 2002, GOMOS has provided more than 850 000 individual vertical profiles of ozone, NO_2, NO_3 and other species. About half of the occultations are made during nighttime. In the beginning of the mission, the instrument made about 400–500 occultations in a day, but, due to the instrumental problems leading to the reduced viewing angle, the number of occultations has been about 200–300 since January 2005. GOMOS nighttime observations are made during the ascending path of ENVISAT, and the local time is approximately equal to the local hour of the ascending node, 22:00 LT (Bertaux et al., 2010). GOMOS tangent point local times cover 1.5 h at the Equator and 3 h at mid-latitudes (Kyrölä et al., 2010a).

The GOMOS inversion has been split in two parts: the spectral inversion and the vertical inversion (Kyrölä et al., 2010b). In the first part, horizontally integrated line densities of O_3, NO_2, NO_3 and aerosols are retrieved simultaneously using a combination of absolute and differential cross sections. In the second part, NO_3 profiles are retrieved from these horizontally integrated line densities at different tangent altitudes. In the latter part, Tikhonov regularization is applied to compensate low signal-to-noise ratio. The vertical resolution of the NO_3, with the current regularization parameter, is 4 km while the sampling resolution (which corresponds to both tangent height resolution and the vertical grid of the product) is smaller (0.5–1.7 km). Besides a smoothing requirement used in Tikhonov regularization, the GOMOS inversion does not use any a priori information of NO_3 profiles.

As GOMOS uses stellar light as a light source, the quality of the measurements and the observations varies from star to star. In the study of Tamminen et al. (2010), GOMOS data were characterized. They concluded that NO_3 can be observed in the 25–45 km altitude range with a precision of 20–40 % with the bright and medium bright stars, and noted also that the cool stars are slightly more favorable for the NO_3 retrieval. In this work, we study only the NO_3 profiles that have been retrieved using stars brighter than 1.9 magnitude and cooler than 15 000 K. We have also screened the GOMOS data so that the solar zenith angle is higher than 107°. We use the GOMOS data version IPF 5.0.

In Fig. 1, we show an example of the GOMOS spectral fit at 40 km and the mean GOMOS spectral fit calculated from the tangent heights between 30 and 45 km. In Fig. 1, the NO_3 absorption features located at 623 nm and 662 nm are clearly visible. The GOMOS star used in this occultation is Alpha Carinae (GOMOS star number 2), which is optimal for measuring NO_3. It is a bright and cool star, and thus the signal-to-noise ratio around the NO_3 absorption bands is high. The GOMOS spectral fitting window is from 248 nm to 690 nm, but in Fig. 1 we have concentrated on area where the NO_3 absorption features are located.

2.2 SAGE III data

SAGE III continued the heritage of SAGE I (1979–1981) and SAGE II (1984–2005) by measuring ozone, nitrogen dioxide, water vapor, and aerosol extinctions by solar occultation technique (McCormick et al., 1989) and additionally performed new nocturnal measurements of ozone, NO_2, and NO_3 using lunar occultation technique (SAGE III ATBD Team, 2002). SAGE III was launched 10 December 2001 on board a Russian Meteor-3M spacecraft, and it recorded data between 7 May 2002–26 October 2005 in lunar occultation mode.

The current SAGE III lunar data set version 3.0 includes 583, 717, 959 and 302 vertical profiles for the years 2002, 2003, 2004 and 2005, respectively. The v3.0 data set contains nocturnal vertical profiles of ozone, NO_2, and NO_3 with near-global coverage. Approximately 32 % of SAGE III lunar measurements occurred between 23:00 and 24:00 local solar time at the tangent point, and approximately 49 % of the measurements occurred in polar regions ($|\text{latitude}| > 60°$). Measurements were attempted when the lunar phase was 40 % or greater and the solar zenith angle greater than 95 degrees. Algorithms for altitude registration, refraction, and data binning were derived using techniques similar to the SAGE II (Chu et al., 1989) and SAGE III solar processing (Chu and Veiga, 1998). The gas species retrieval algorithms were developed prior to launch using a complex forward simulation model. The simulation incorporated a solar spectrum over the SAGE III wavelengths reflected by a modeled lunar disk with variable albedo, ray tracing through the atmosphere, and the effects of Rayleigh scattering and absorption by molecular gases and aerosols. Briefly, the retrieval procedure used a differential optical absorption spectroscopy algorithm (Platt et al., 1979) to compute line-of-sight column densities of gas species and then performed an inversion using an onion peel algorithm to compute number density concentrations.

The vertical resolution of the SAGE III NO_3 profiles is 1 km, and the data are given on a 0.5 km fixed grid between

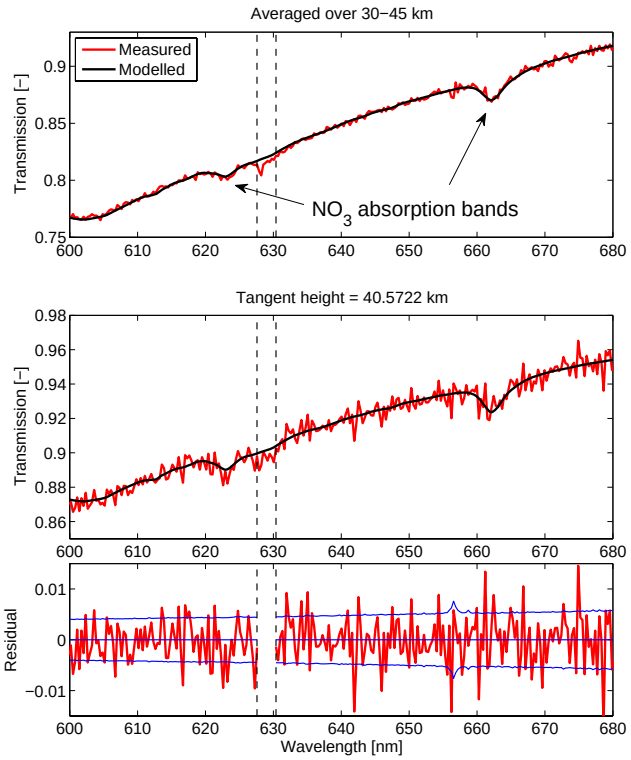


Fig. 1. An example of the GOMOS spectral fit. Uppermost panel shows GOMOS mean spectral fit calculated from the tangent heights between 30 and 45 km. Middle panel shows spectral fit at 40 km, and the lowermost panel is the residual at 40 km. The example is the same as in Fig. 4. NO_3 absorption features are clearly visible. Note that in the GOMOS Level 2 processing the spectral range 627.9–630.1 nm, covering the red line of atomic oxygen and O_2 absorption band, has been flagged and this area is not used in the retrieval (Kyrölä et al., 2010b; Tamminen et al., 2010).

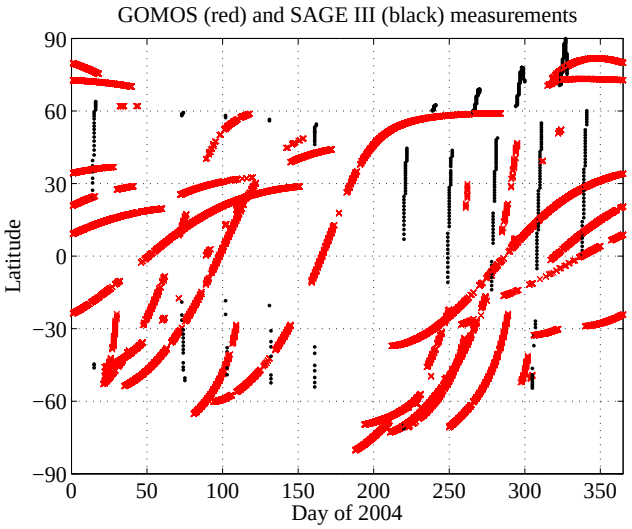


Fig. 2. GOMOS (red) and SAGE III (black) measurements in the year 2004. The GOMOS data have been screened, so that the solar zenith angle is higher than 107° and the stars used are brighter than 1.9 magnitude and cooler than 15 000 K. The SAGE III data have been screened, so that the solar zenith angle is higher than 107°.

20–60 km. The errors associated with the SAGE III NO_3 observations are provided in the data products and are 20–50 % between 25 km and 45 km. For a detailed description of the instrument, lunar processing algorithms, and cross section data, see SAGE III ATBD Team (2002). In this study, SAGE III data have been screened, so that the solar zenith angle is higher than 107°.

2.3 Collocations and comparisons

These two data sets provide unique information on the nighttime NO_3 profiles measured from the Earth's limb. Temporal overlap of the data sets starts in August 2002 and continues until the end of 2005. Still, finding suitable collocations between GOMOS and SAGE III is quite difficult, because the two data sets are not homogeneously distributed (see Fig. 2).

In practice, we need to find a compromise between spatiotemporal limits and statistical representativeness. In order to find matches between GOMOS and SAGE III, we set the maximum latitudinal and longitudinal difference to be 2 and 5 degrees, respectively. For the temporal difference, we allowed the measurements to be 24 h away from each other, and at the same time set the local hour difference to be less than 2 h. With these criteria, we found 5, 23 and 8 collocated pairs for the years 2002, 2003 and 2004, respectively. For the year 2005, we did not find any matches, mainly due to the fact that SAGE III measured most of its data during a time when GOMOS was suffering from a technical anomaly. If we allow the temporal difference to be one week, still demanding the local hour difference to be less than 2 h, we can find 115 matches instead of 36. In cases where multiple GOMOS matches were found to an individual SAGE III profile, we selected the one that had the smallest time difference. The spatial distribution of these 115 matches is shown in Fig. 3. When 36 collocations are considered, the mean local time difference (GOMOS-SAGE III) is 18 min. When 115 collocations are considered, the mean difference is 7 min.

For statistical analysis, we use a symmetrically normalized GOMOS to SAGE III difference defined as

$$f(g,s) = 200 \times \frac{g-s}{g+s}\,[\%], \tag{1}$$

where g is GOMOS and s is SAGE III. Because GOMOS and SAGE III profiles possess different vertical resolutions, GOMOS averaging kernels are applied to the SAGE III profiles.

The expected variance of the difference (Eq. 1) is approximated by

$$\sigma_f^2 \approx \left|\frac{\partial f}{\partial g}\right|^2 \sigma_g^2 + \left|\frac{\partial f}{\partial s}\right|^2 \sigma_s^2, \tag{2}$$

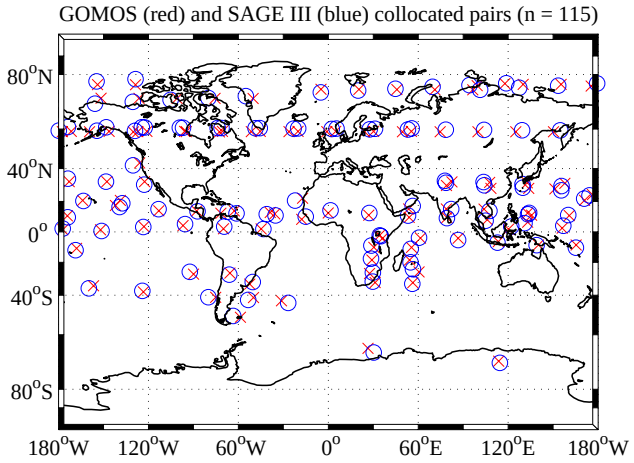


Fig. 3. Spatial distribution of 115 collocated GOMOS-SAGE III pairs (red cross and blue ring) from the years 2002–2004.

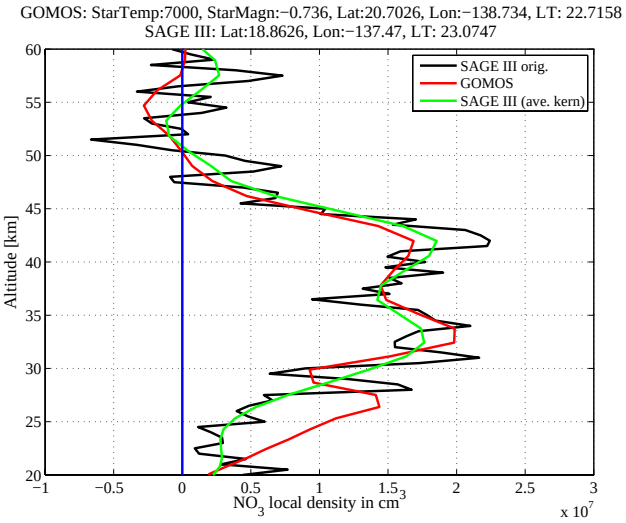


Fig. 4. An individual match between GOMOS and SAGE III showing good agreement. SAGE III profile is plotted before and after the GOMOS averaging kernel was applied. Figure 4 illustrates two-peaked NO_3 profile.

where σ_g and σ_s are the error estimates reported by GOMOS and SAGE III products, respectively. Eq. (2) makes the standard assumption that GOMOS and SAGE III errors are uncorrelated.

3 Results

Herein, we seek to verify the NO_3 measurement accuracies by the direct comparison of GOMOS and SAGE III NO_3 observations.

In Fig. 4 an individual match, illustrating a two-peaked NO_3 profile and showing good agreement between the products, is shown. SAGE III profile is plotted before and after the application of the GOMOS averaging kernel. The example is the same as in Fig. 1.

In Fig. 5, we show the statistics calculated from 36 collocated pairs found from the years 2002–2004. The black solid line is the median of the individual differences, and the black dashed lines correspond to median $\pm$ interquartile deviations. The green horizontal lines represent the 95 % confidence limits ($\pm$ standard error of the mean $\times$ 1.96). For calculating the mean and standard deviation, we neglected the differences where the distance between the value and the median value is higher than $3 \times 1.4826 \times$ the median absolute deviation in order to exclude clear outliers. This is approximately the same as rejecting the data outside 3σ limits.

From Fig. 5, we observe quite large deviations and median values oscillating between $\pm$25 %. We can observe a positive bias pattern of some 10 % below 40 km. It is however small compared to the variability. Above 50 km, the median of GOMOS to SAGE III differences increases up to 100 %. These findings are also valid for the years 2003 and 2004 separately. For the year 2002, the structure of the differences is very noisy, mainly due to the fact that we have only 5 collocated pairs (not shown).

In Fig. 5 we also show the mean $\pm$ square root of the median of the individual variances as defined in Eq. (2). When these blue dashed lines are compared against the mean $\pm$ standard deviation (the green dashed lines), we note that the expected errors of the differences are consistent with the observed differences around 40 km, where NO_3 typically peaks. The uncertainties are underestimated below 33 km and overestimated above 42 km.

The statistics from 115 collocated pairs, where the temporal difference is allowed to be one week, are shown in Fig. 6. Again, we can observe a small positive bias of some 10 % below 40 km. From the interquartile deviation, we can clearly see that the spread between the observations starts growing below 30 km and above 45 km. One can observe that the general structures of the medians in Figs. 5 and 6 are similar. We can also see in Fig. 6 that the expected and observed errors are consistent with each other between the altitudes 33 km and 42 km.

In addition to the differences of the collocated pairs, we also compared the zonal medians in different latitude bands in 2004. In order to make these zonal medians more comparable, we concentrated on three month periods where the distributions inside the latitude bands are closest. Based on Fig. 2 for latitude band 60° S–30° S, we selected months 3–5 and for other latitude bands, we selected months 10–12. For the latitude band 90° S–60° S, we selected only one month (October) because after the screening there were only 6 SAGE III profiles.

The results from these six different latitude bands are shown in Fig. 7. We can see similarities with these profiles. In the tropics, the profiles are almost identical and they deviate slightly at high latitudes, while still representing the same

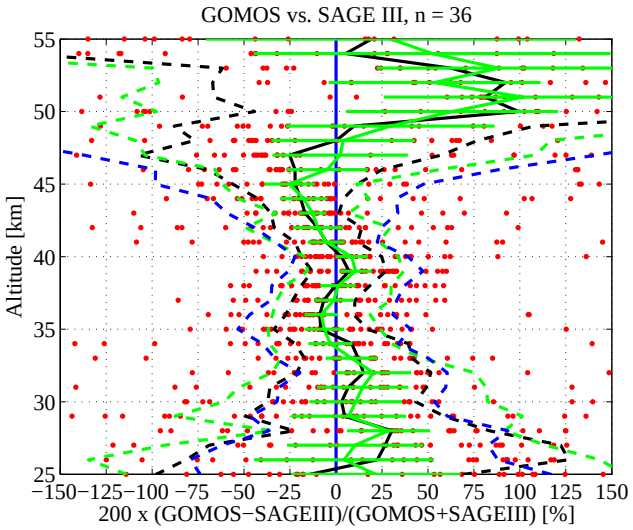


Fig. 5. The statistics of the collocated GOMOS-SAGE III pairs from the years 2002–2004. The black solid line is the median of the individual differences, and the black dashed lines correspond to median ± interquartile deviation. The green solid line is the median filtered mean, and the green horizontal lines represent its 95 % confidence limits. The green dashed lines correspond to mean ± standard deviation, and the blue dashed lines correspond to mean ± square root of the median variance as defined in Eq. (2).

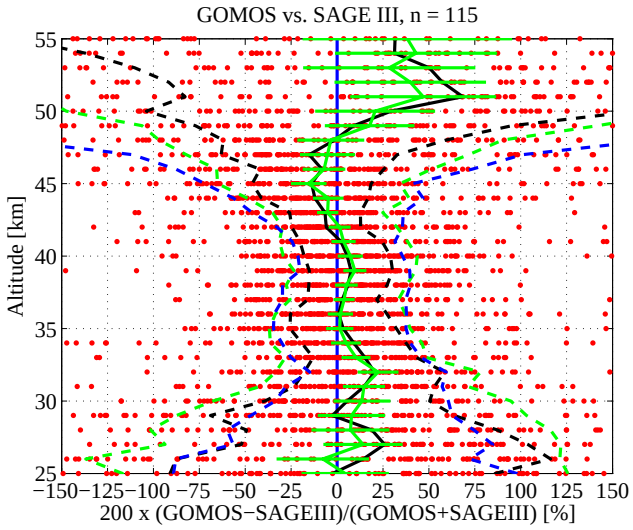


Fig. 6. Same as Fig. 5., except the temporal difference between the measurements is allowed to be as much as one week.

features. We must note that the number of profiles (see the caption of Fig. 7) that is used to calculate the medians is far from equal. Also, the temporal and spatial sampling of the observations varies, although we have concentrated on areas where the distributions are closest. Still, these comparisons confirm that GOMOS and SAGE III nighttime NO_3 climatologies agree well with each other and we do not observe any clear systematic differences between them.

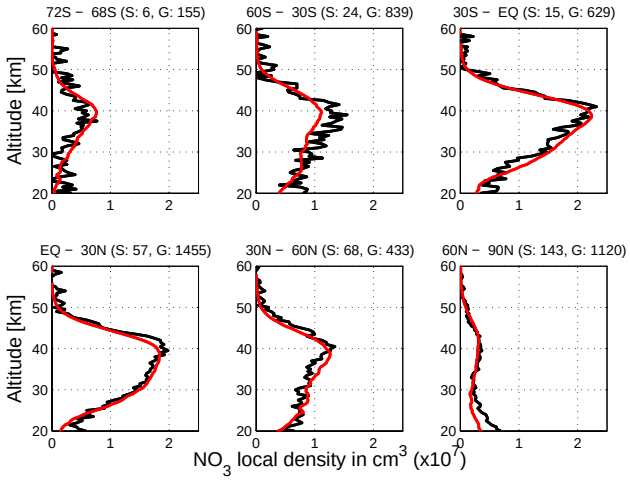


Fig. 7. GOMOS (red) and SAGE III (black) zonal medians in six different latitude bands in 2004. The zonal medians are calculated from three month periods where the spatial and temporal distributions of the measurements inside the latitude band are closest. For latitude band 90° S–60° S, the median is calculated only from one month period (October). The numbers after S and G in titles indicate from how many profiles the median was calculated for SAGE III and GOMOS, respectively.

4 Conclusions and remarks

In this work, we compared GOMOS and SAGE III NO_3 vertical profile data sets, retrieved using stellar and lunar occultation techniques, respectively. Statistical analysis of the limited amount of collocated pairs indicates a good overall agreement between GOMOS and SAGE III. Between the altitudes 25 km and 45 km, the median difference between these two data sets is within ±25 %. From the zonal median profiles, we can see reasonable agreement, showing that the climatological median profiles are comparable. The agreement is at its best in the tropics and slightly worse in other latitude bands. We expect that the better agreement in the tropics is due to more stable atmospheric conditions and more equal sampling of the instruments.

The expected error values of the differences based on the error estimates reported in the data products are consistent with the observed standard deviations between the altitudes 33 km and 42 km. Below 33 km, the observed standard deviations are underestimated. We expect that one reason for this is that the reported GOMOS error estimates are too low. In the next processing version (IPF 6.01), the GOMOS error estimates are expected to be improved and slightly higher for NO_3.

It is worth noting that, despite the noise and other limitations, these two data sets are the only publicly available data sets of NO_3 vertical profiles, leaving GOMOS the sole data set since 2005, when the last SAGE III data were recorded. Vertical profiles of NO_3 are also retrieved from ENVISAT/SCIAMACHY lunar occultations (Amekudzi et al., 2005),

but unfortunately the limited scientific data set did not provide useful information (no collocations found) for comparison with the GOMOS NO_3 vertical profiles. A copy of the SAGE III instrument will be deployed on the International Space Station (ISS) in 2014. It will continue to record data in the lunar occultation mode, and hence it will provide NO_3 vertical profiles.

Acknowledgements. The authors would like to thank the GOMOS and SAGE III teams for providing the data. This work is partly supported by Academy of Finland's MIDAT project. Mr. Moore is supported by the NASA Contract NNL11AA00B through Science Systems and Applications, Inc. The SAGE III data sets are publicly available through the NASA Langley Atmospheric Science Data Center (ASDC). The GOMOS data sets are available via ESA website and registration is required.

Edited by: C. von Savigny

References

Amekudzi, L. K., Sinnhuber, B.-M., Sheode, N. V., Meyer, J., Rozanov, A., Lamsal, L. N., Bovensmann, H., and Burrows, J. P.: Retrieval of stratospheric NO_3 vertical profiles from SCIAMACHY lunar occultation measurement over the Antarctic, J. Geophys. Res., 110, D20304, doi:10.1029/2004JD005748, 2005.

Amekudzi, K. L., Bramstedt, K., Bracher, A., Rozanov, A., Bovensmann, H., and Burrows, J. P.: SCIAMACHY solar and lunar occultation: Validation of ozone, NO_2, and NO_3 profiles, in: Proc. of the Third Atmospheric Chemistry Validation of ENVISAT (ACVE-3), SP-642, ESA Publication Division, 2007.

Bertaux, J. L., Kyrölä, E., Fussen, D., Hauchecorne, A., Dalaudier, F., Sofieva, V., Tamminen, J., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Mangin, A., Blanot, L., Lebrun, J. C., Pérot, K., Fehr, T., Saavedra, L., Leppelmeier, G. W., and Fraisse, R.: Global ozone monitoring by occultation of stars: an overview of GOMOS measurements on ENVISAT, Atmos. Chem. Phys., 10, 12091–12148, doi:10.5194/acp-10-12091-2010, 2010.

Chu, W. P. and Veiga, R. E.: SAGE III/EOS, in: Proceedings of SPIE, 3501, 52–50, 1998.

Chu, W. P., McCormick, M. P., Lenoble, J., Brogniez, C., and Pruvost, P.: SAGE II inversion algorithm, J. Geophys. Res., 94, 8339–8351, 1989.

Hauchecorne, A., Bertaux, J.-L., Dalaudier, F., Cot, C., Lebrun, J.-C., Bekki, S., Marchand, M., Kyrölä, E., Tamminen, J., Sofieva, V., Fussen, D., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Mangin, A., Théodore, B., Guirlet, M., Snoeij, P., Koopman, R., Saavedra de Miguel, L., Fraisse, R., and Renard, J.-B.: First simultaneous global measurements of nighttime stratospheric NO_2 and NO_3 observed by Global Ozone Monitoring by Occultation of Stars (GOMOS)/Envisat in 2003, J. Geophys. Res., 110, D18301, doi:10.1029/2004JD005711, 2005.

Kyrölä, E., Tamminen, J., Sofieva, V., Bertaux, J. L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Guirlet, M., Fehr, T., and Saavedra de Miguel, L.: GOMOS O_3, NO_2, and NO_3 observations in 2002–2008, Atmos. Chem. Phys., 10, 7723–7738, doi:10.5194/acp-10-7723-2010, 2010a.

Kyrölä, E., Tamminen, J., Sofieva, V., Bertaux, J. L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Guirlet, M., Mangin, A., Blanot, L., Fehr, T., Saavedra de Miguel, L., and Fraisse, R.: Retrieval of atmospheric parameters from GOMOS data, Atmos. Chem. Phys., 10, 11881–11903, doi:10.5194/acp-10-11881-2010, 2010b.

Marchand, M., Bekki, S., Hauchecorne, A., and Bertaux, J.-L.: Validation of the self-consistency of GOMOS NO_3, NO_2 and O_3 data using chemical data assimilation, Geophys. Res. Lett., 31, L10107, doi:10.1029/2004GL019631, 2004.

McCormick, M. P., Zawodny, J. M., Veiga, R. E., Larsen, J. C., and Wang, P. H.: An overview of sage I and II ozone measurements, Planet. Space Sci., 37, 1567–1586, doi:10.1016/0032-0633(89)90146-3, 1989.

McLinden, C. A. and Haley, C. S.: Odin/OSIRIS observations of stratospheric NO_3 through sunrise and sunset, Atmos. Chem. Phys., 8, 5529–5534, doi:10.5194/acp-8-5529-2008, 2008.

Naudet, J. P., Huguenin, D., Rigaud, P., and Cariolle, D.: Stratospheric observations of NO_3 and its experimental and theoretical distribution between 20 and 40 km, Planet. Space Sci., 29, 707–712, 1981.

Noxon, J. F., Norton, R. B., and Henderson, W. R.: Observation of atmospheric NO_3, Geophys. Res. Lett., 5, 675–678, 1978.

Platt, U., Perner, D., and Pätz, H.: Simultaneous Measurement of Atmospheric CH_2O, O_3, and NO_2 by Differential Optical Absorption, J. Geophys. Res., 84, 6329–6335, 1979.

Renard, J.-B., Pirre, M., Robert, C., Moreau, G., Huguenin, D., and Russell III, J. M.: Nocturnal vertical distribution of stratospheric O_3, NO_2 and NO_3 from balloon measurements, J. Geophys. Res., 101, 28793–28804, doi:10.1029/96JD02012, 1996.

Renard, J.-B., Berthet, G., Brogniez, C., Catoire, V., Fussen, D., Goutail, F., Oelhaf, H., Pommereau, J., Roscoe, H. K., Wetzel, G., Chartier, M., Robert, C., Balois, J., Verwaerde, C., Auriol, F., François, P., Gaubicher, B., and Wursteisen, P.: Validation of GOMOS-Envisat vertical profiles of O_3, NO_2, NO_3, and aerosol extinction using balloon-borne instruments and analysis of the retrievals, J. Geophys. Res., 113, A02302, doi:10.1029/2007JA012345, 2008.

SAGE III ATBD Team: SAGE III Algorithm Theoretical Basis Document (ATBD) Solar and Lunar Algorithm, LaRC 475-00-109, Tech. rep., NASA Langley Res. Cent., Version 2.1, available at: http://eospso.gsfc.nasa.gov/eos_homepage/for_scientists/atbd/viewInstrument.php?instrument=14, 2002.

Tamminen, J., Kyrölä, E., Sofieva, V. F., Laine, M., Bertaux, J.-L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton-d'Andon, O., Barrot, G., Mangin, A., Guirlet, M., Blanot, L., Fehr, T., Saavedra de Miguel, L., and Fraisse, R.: GOMOS data characterisation and error estimation, Atmos. Chem. Phys., 10, 9505–9519, doi:10.5194/acp-10-9505-2010, 2010.

Publ. V

GOMOS one-step retrieval algorithm

Hakkarainen, J., Laine, M., and Tamminen, J., *Proc. SPIE 8890*, Remote Sensing of Clouds and the Atmosphere XVIII (2013), doi: 10.1117/12.2027109.

GOMOS one-step retrieval algorithm

Janne Hakkarainen[a,b], Marko Laine[a], and Johanna Tamminen[a]

[a]Earth Observation, Finnish Meteorological Institute, Helsinki, Finland;
[b]Lappeenranta University of Technology, Lappeenranta, Finland

ABSTRACT

Global Ozone Monitoring by Occultation of Stars (GOMOS) is a satellite instrument onboard the ENVISAT platform that was in operation during 2002–2012. During these years, GOMOS observed about 880 000 vertical profiles of ozone, NO_2, NO_3 and aerosols. The GOMOS measurement principle is relatively simple based on the stellar-occultation technique. In this paper, we present an alternative retrieval algorithm for processing the GOMOS measurements. The presented algorithm is based on the so-called one-step approach, where both the *spectral* and the *vertical* inversions are executed simultaneously. This approach has several attractive features. In particular, the one-step approach allows a better use of the smoothness prior information and, unlike in the operative algorithm, the prior given to one specie affects the other species too. This feature is critical when going near the detection limit, especially in the upper troposphere lower stratosphere (UTLS) region. The main challenge in the GOMOS one-step algorithm is to find the correct smoothness priors for the different species at different altitudes. In this paper, we give a technical description of the one-step retrieval algorithm and discuss the differences between this and the operative algorithm. In the case study part of this paper, we compare the one-step and the operative ozone retrievals in Arctic region during the exceptional ozone-depletion conditions in spring 2011. We show that the quality of the ozone profiles can be improved by introducing the one-step algorithm. The improvement is drastic in the lower stratosphere at 15–20 km altitude.

Keywords: GOMOS, Tikhonov regularization, Bayesian retrieval, one-step algorithm, UTLS, Arctic, ozone

1. INTRODUCTION

The satellite-borne occultation measurements of the atmospheric trace gases are always indirect, which leads to challenging atmospheric inverse problems. The retrieval problems have created a rich field of science and many techniques to retrieve atmospheric trace gases has been proposed during the last 30 years, since 1979 when the first instrument of the Stratospheric Aerosols and Gas Experiment (SAGE) family was launched. SAGE I employed the solar-occultation technique to measure the vertical structure of the atmosphere. Since then, various occultation instruments using, e.g., solar, lunar and stellar occultation have been launched. In addition to the occultation-type instruments, other limb-geometry instruments—designed, e.g., for measuring the limb-scattered solar light—do exist.

As the measurement principle and the inversion techniques vary, so do the other free parameters in the retrieval algorithms. Probably the most critical assumption is based on the prior information, since in the atmospheric inverse problems, the measurement data alone are not enough to identify the unknown atmospheric parameters. There are several ways of giving the prior information. One could, e.g., consider the positivity constrains or the smoothness requirements as priors. In Bayesian framework, the climatological priors are often used. In addition to priors, other assumptions—like, e.g, the number of used wavelengths, wavelength regions, cross-sections and models—do affect the retrieved profiles. Hence, different inversion techniques, different measurements and different kinds of a priori assumptions make their own characteristics to the retrieved profiles. In practice, this means that even from the same measurement data, with different retrieval algorithms, one can invert different kinds of profiles with different features like the vertical resolution and the valid altitude range.

Further author information: (Send correspondence to J.H.)
J.H.: E-mail: janne.hakkarainen@fmi.fi
M.L.: E-mail: marko.laine@fmi.fi
J.T.: E-mail: johanna.tamminen@fmi.fi

Remote Sensing of Clouds and the Atmosphere XVIII; and Optics in Atmospheric Propagation and Adaptive Systems XVI, edited by Adolfo Comeron, et. al., Proc. of SPIE Vol. 8890, 88900F

The ozone distribution and variability in the upper troposphere and lower stratosphere (UTLS) has a significant role in chemistry–climate interaction.[1] In recent years, several efforts to improve the ozone retrievals in the UTLS region have been proposed. E.g., in Refs. 1 and 2, the retrievals have been improved using the improved *a priori* climatology that allows the day-to-day variation and takes into account the tropopause height. In this work, we propose a targeted retrieval algorithm for limb-viewing Global Ozone Monitoring by Occultation of Stars (GOMOS) satellite instrument. The algorithm allows better use of the (smoothness) prior information and the retrievals are improved substantially in the lower stratosphere.

The outline of this paper is the following. In Section 2, GOMOS measurement principle and the operative retrieval algorithm are recalled. In Section 3, an alternative retrieval algorithm based on one-step approach is presented. In Section 4, in the case study part, the one-step and the operative ozone retrievals during the 2011 Arctic ozone loss period are compared. In Section 5, discussion about the differences between the retrieval algorithms is given. Finally, Section 6 concludes the paper.

2. GOMOS MEASUREMENTS AND THE OPERATIVE RETRIEVAL ALGORITHM

Global Ozone Monitoring by Occultation of Stars (GOMOS) is a satellite instrument onboard ENVISAT spacecraft that was launched in March 2002. It was operative until the end of ENVISAT mission in 2012. During the mission, GOMOS observed about 880 000 vertical profiles of ozone, NO_2, NO_3, aerosol extinctions and other species. About half of these measurements were made during nighttime. In this section, we recall the GOMOS measurement principle and the operative retrieval algorithm. For a more comprehensive introduction to GOMOS, see the GOMOS Algorithm Theoretical Basis Document (ATBD)* and ACP's GOMOS special issue† and the papers therein; in particular, Refs. 3–6.

The GOMOS measurement principle is relatively simple and it is based on the stellar-occultation technique. The stellar spectrums at different tangent altitudes are obtained with the sampling resolution of $0.3 - 1.7$ km. One occultation contains roughly 100 measurement spectrums at 1 416 different wavelengths in the UV–Visible wavelength region. Hence, roughly 150 000 individual measurements per occultation are obtained. To obtain the so-called transmission spectrum, the stellar spectrum is divided by the reference spectrum that is measured above the atmosphere. This can be written as

$$T_{\text{ext}}^{\text{meas}}(\lambda, l) = \frac{I(\lambda, l)}{I^{\text{ref}}(\lambda)}, \tag{1}$$

where $I(\lambda, l)$ is the stellar spectrum measured along the line-of-sight l and at wavelength λ. $I^{\text{ref}}(\lambda)$ is the reference spectrum measured above the atmosphere. The error characteristics of the measured transmission spectrum—used later to obtain the atmospheric profiles—vary strongly from star to star with the stellar brightness and temperature. In the lower atmosphere, in particular in the UTLS region, the signal-to-noise ratio is low due to increasing amount of aerosols. See Ref. 6 for discussion and examples.

The transmission can be modeled using the Beer-Lambert law

$$T_{\text{ext}}^{\text{mod}}(\lambda, l) = \exp(-\tau(\lambda, l)) \tag{2}$$

where the optical depth $\tau(\lambda, l)$ is given as

$$\tau(\lambda, l) = \sum_{j=1}^{\text{nconst}} \int_l \alpha_j(\lambda, T(s)) \rho_j(s) \, \mathrm{d}s, \tag{3}$$

where ρ_j is a local density profile and $\alpha_j(\lambda, T(s))$ is the constituent, wavelength and temperature dependent cross-section. The integration is done along the line-of-sight. This model can be reduced using the so-called effective cross-sections that are assumed to be constant along the integration. This reduced model can be written as

$$\tau(\lambda, l) = \sum_{j=1}^{\text{nconst}} \alpha_j^{\text{eff}}(\lambda) N_j(l), \tag{4}$$

*Available on the Internet at `https://earth.esa.int/documents/10174/384988/GOMOS_ATBD_V3.pdf`

†Available on the Internet at `http://www.atmos-chem-phys.net/special_issue153.html`

where $N_j(l) = \int_l \rho_j(s)\,\mathrm{d}s$ is the so-called line density of the constituent j and $\alpha_j^{\mathrm{eff}}(\lambda)$ is the constituent and wavelength dependent effective cross-section. When the problem has been discretized, we can use the geometry kernel matrix K to obtain the line densities N_j from the actual local density profiles ρ_j. This can be written as

$$N_j = K\rho_j. \tag{5}$$

Using the vectorized notation, the full transmission model can be written as

$$T_{\mathrm{ext}}^{\mathrm{mod}}(\rho) = \exp(-(K \otimes A)\rho), \tag{6}$$

where $K \otimes A$ is the Kronecker product between the kernel operator K and the cross-sections A. If three gases (O_3, NO_2 and NO_3), three aerosol parameters and 100 transmission spectrums at 1 500 wavelengths are considered, the sizes of the elements of the model (6) are the following. The length of the transmission model output $T_{\mathrm{ext}}^{\mathrm{mod}}$ is $1\,500 \times 100 = 150\,000$ and the length of the input ρ is $6 \times 100 = 600$. Hence, the linear operator $K \otimes A$ has to be a $150\,000 \times 600$ matrix. Here the altitudes of the transmission model input and output are assumed to be the same. In addition to O_3, NO_2, NO_3 and aerosol parameters, neutral density and some minor trace gases like OClO and BrO can be included in the model, too.

The aim of the GOMOS retrieval algorithms is to find the atmospheric profiles ρ_j that model the measured transmission in statistical sense given the measurement error and some *a priori* knowledge. Next, we present an outline of the operative approach.

2.1 Operative retrieval algorithm

The operative GOMOS retrieval algorithm is based on a two-step approach, where the *spectra*l and the *vertical* inversion are conducted separately. This procedure can be iterated for improving accuracy. The use of the reduced model with the effective cross-sections enables the splitting.

In the spectral inversion, the line densities N are obtained by minimizing the following cost function

$$J(N) = \frac{1}{2}(T_{\mathrm{ext}}^{\mathrm{mod}}(N) - T_{\mathrm{ext}}^{\mathrm{meas}})^{\mathrm{T}} C^{-1} (T_{\mathrm{ext}}^{\mathrm{mod}}(N) - T_{\mathrm{ext}}^{\mathrm{meas}}). \tag{7}$$

using the Levenberg-Marquardt (LM) optimizing algorithm. The covariance matrix C includes the measurement noise and possible modeling uncertainties. Note that this part of the retrieval uses uninformative flat prior and only the likelihood part of the Bayesian cost function is considered.

When the spectral inversion part has been done and the different line densities N_j obtained, the vertical inversion can be started. In the vertical inversion, the aim is to find the vertical profiles ρ_j given the line densities N_j separately for every constituent. In principle, this could be done solving the linear Eq. (5). However, because of the low signal-to-noise ratio and other factors, the solution is based on Tikhonov regularization. Using the matrix calculus, the solution can be written as

$$\rho_{\gamma,j} = (K^{\mathrm{T}}K + \gamma L^{\mathrm{T}}L)^{-1} K^{\mathrm{T}} N_j, \tag{8}$$

where γ is the Tikhonov regularization parameter. In the retrieval, the smoothness requirements are considered. Hence, L is selected to be the tridiagonal matrix that approximate the second derivative. It can be written as

$$L = \frac{1}{h^2} \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \\ 0 & \cdots & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{9}$$

where h is the layer thickness (in practice different for every layer). In the operative algorithm, the regularization parameter γ has been selected so that the so-called target resolution—defined as a spread of the averaging kernel

computed using (5) and (8), see, e.g., Ref. 7 for details—for ozone is 2 km below 30 km and 3 km above 40 km. For NO_2, NO_3 and aerosol parameters the selected target resolution is 4 km at all altitudes.

Two features about the algorithm can be noted. First, the (smoothness) priors have only effect in the vertical inversion part and are given for each constituent separately. This means that the prior given, e.g., to aerosol profiles does not affect the retrievals of other species. Second, the averaging kernel of the Tikhonov solution (8) does not depend of the measurement noise, and it depends of the actual measurements only via the kernel matrix K. This allows the setting of the target resolution, since the measurement geometry is relatively similar from one occultation to another and the layer thickness is included in (9).

3. ONE-STEP RETRIEVAL ALGORITHM

The basic idea of the GOMOS one-step retrieval algorithm is to conduct the spectral and the vertical inversion of the operative GOMOS algorithm simultaneously using the full model (6). From the methodological point of view this makes sense, since the assumption of the effective cross-sections, see formula (4), can be dropped. In addition, the splitting of the inversion in two parts is somewhat artificial. Historically, the use of the two-step strategy in the operative algorithm is mainly related to the enormous size of the problem. The possibility of the one-step inversion was studied already in the early days of the GOMOS mission.[8,9] Here we present our own version of the GOMOS one-step retrieval algorithm. The algorithm is written in MATLAB programing language.

The one-step retrieval of the GOMOS measurements can be seen as a minimization of the following cost function consisting of the likelihood and the prior:

$$J(\rho) = \frac{1}{2}(T_{\text{ext}}^{\text{mod}}(\rho) - T_{\text{ext}}^{\text{meas}})^{\text{T}} C^{-1} (T_{\text{ext}}^{\text{mod}}(\rho) + T_{\text{ext}}^{\text{meas}}) + \frac{1}{2}(\rho_{\text{prior}} - H\rho) C_{\text{prior}}^{-1} (\rho_{\text{prior}} - H\rho), \tag{10}$$

where H is a linear operator based on how the prior given. If, e.g., the smoothness prior is considered, the matrix H is selected like the matrix L in (9) and ρ_{prior} is set to zero. This cost function could, in principle, be minimized using the non-linear optimizing algorithms like the LM algorithm used in the spectral inversion part of the operative algorithm. However, in practice, we solve this problem using generalized linear model (GLIM) approach with the iterative re-weighted least squares algorithm (IRLS), where the special exponential structure of the problem can be exploited for numerical efficiency.[10] The IRLS algorithm is a special case of a more general approach to optimization called the Newton-Raphson algorithm.

Let us next show how the GLIM approach works in practice. Let us consider the following model

$$y = \exp(Ax) + \varepsilon, \tag{11}$$

where y are the measurements, A is the linear operator, $\varepsilon \sim N(0, \sigma I)$ is the measurement noise and x is the unknown state vector. Readers familiar with the GLIM vocabulary, can note that the *link function* between the measurements and the linear operator is logarithmic and—as the measurement error is assumed Gaussian—the *variance function* is simply identity.

Using the IRLS algorithm, starting form the initial guess of x, the unknown x can be solved iteratively. In the iteration steps, the auxiliary problem based on the linear equation

$$z = Ax, \tag{12}$$

is solved using the standard-deviation weights w, where

$$\eta = Ax, \tag{13}$$
$$\mu = \exp(\eta), \tag{14}$$
$$z = \eta + (y - \mu)/\mu, \tag{15}$$
$$w = \sigma/\mu. \tag{16}$$

For more information, see Ref. 10 for details. The linear step is operating on the original unknowns x and if we assign prior regularization to each step then the final solution uses the same prior constraints.

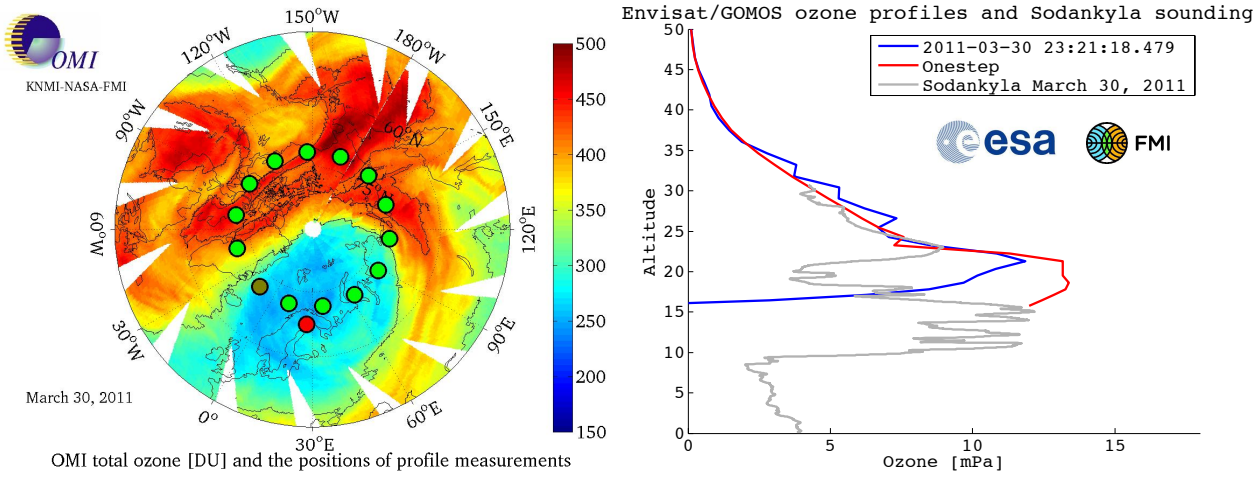


Figure 1. An example of the GOMOS ozone profiles retrieved using the one-step and the operative retrieval algorithms. In addition, the Sodankylä sounding profile is illustrated. In the left panel, GOMOS measurement locations and the total ozone map from OMI instrument are illustrated. See an animation at `http://dl.dropboxusercontent.com/u/35228286/gomos_video20110330.gif`

When prior information x_{prior} is considered, the iterative solution x can be seen as the argument that minimizes the following quadratic cost function

$$J(x) = \frac{1}{2}(Ax - z)^{\mathrm{T}} C_w^{-1}(Ax - z) + \frac{1}{2}(x_{\text{prior}} - Hx)^{\mathrm{T}} C_{\text{prior}}^{-1}(x_{\text{prior}} - Hx), \tag{17}$$

where C_w is the diagonal covariance matrix induced by the weights w. As the operators A and H in the cost function (17) are linear, we can obtain the solution x using basic linear algebra

$$x = (A^{\mathrm{T}} C_w^{-1} A + H^{\mathrm{T}} C_{\text{prior}}^{-1} H)^{-1}(A^{\mathrm{T}} C_w^{-1} z + H^{\mathrm{T}} C_{\text{prior}}^{-1} x_{\text{prior}}). \tag{18}$$

In practice, we solve this linear equation using Cholesky factorization of $(A^{\mathrm{T}} C_w^{-1} A + H^{\mathrm{T}} C_{\text{prior}}^{-1} H)$ and MATLAB's backslash[‡] operator, although other options exist too. We note that in full case, the presented GLIM approach is computationally much lighter than the computationally rather expensive LM algorithm.

Prior information. In the one-step algorithm, there are currently three ways of giving the prior information: one can give the prior in absolute units or give the prior for the first or the second derivative of the profiles. Naturally, all three types of priors can be considered at the same time. The non-trivial part in the one-step algorithm is the selection of the (diagonal) prior covariance matrix C_{prior} for the derivatives. From the statistical point of view, the covariance matrix should reflect our prior knowledge, e.g., from the theory, about how smooth the profiles are or how big steps can the profiles make between the altitudes. See Ref. 9 and 11 for details. From the practical point of view, the prior covariance matrix can be seen as a similar regularization tuning handle of the inversion as the Tikhonov regularization parameter γ is in the operative algorithm.

4. CASE STUDY: ARCTIC OZONE

In this section, we show the operative and the one-step GOMOS ozone-profile retrievals during the 2011 Arctic ozone-loss period.[12] Our aim is to highlight the visual differences of the ozone profiles retrieved with the different algorithms.

The GOMOS measurements are obtained using star number 100 (Gamma Corvi) with visual magnitude of 2.580 V and effective temperature of 13 100 K. The star is characterized as hot and dim[6] and the signal-to-noise

[‡]See `http://www.mathworks.se/help/matlab/ref/mldivide.html`

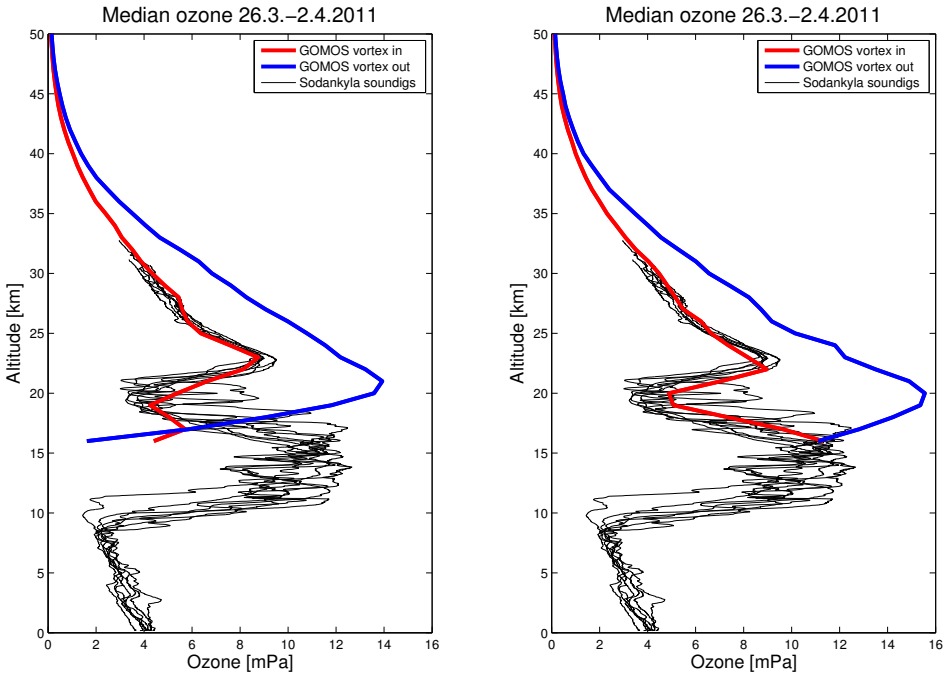


Figure 2. The median GOMOS ozone profiles retrieved using the operative (left panel) and the one-step (right panel) retrieval algorithms inside and outside the Arctic polar vortex. Sodankylä sounding profiles are illustrated too. Clear differences between the retrieval algorithms can be observed.

ratio is low. The measurements are obtained in challenging twilight illumination conditions and are from a very narrow latitude band at 71° N−72° N.

In the one-step retrieval algorithm the prior covariance matrix C_{prior} was visually "tuned" to a single Sodankylä ozone-sounding profile observed 26 March 2011. The covariance matrix was kept fixed for all other retrievals. Aerosols are known to be the main cause of systematic uncertainty in the GOMOS ozone retrievals in the lower atmosphere below 25 km.[6] Thus, very strong priors for aerosol smoothness were set.

An example of the operative and the one-step retrieval results are illustrated in Fig. 1. The profiles are obtained in March 30, 2011. In addition to GOMOS profiles, the Sodankylä sounding profile and the Ozone Measurement Instrument (OMI)[13] total ozone maps are illustrated too. In the total ozone map, low (blue) values are inside and high (red) values are outside the polar vortex. The full one-day animation of the GOMOS profiles are available online and the web-link is given in the caption of Fig. 1. From the results presented in the animation, it can be noted that the GOMOS one-step retrieval algorithm produced more realistic retrieval results than operative algorithm. In particular, outside the polar vortex below 20 km the operative profiles show an unexpected shape producing lower values than expected. In practice, this can be visually seen in all profiles outside the polar vortex. Surprisingly, the operative algorithm performs better inside than outside polar vortex and the inversion is valid at lower altitudes. In one-step retrievals, the unexpected shape cannot be seen. Nevertheless, in some cases, the one-step retrieval algorithm produces too high values inside the polar vortex.

The median ozone profiles inside and outside the polar vortex using the operative (left panel) and the one-step (right panel) retrieval algorithms are illustrated in Fig. 2. The medians are calculated from the profiles that are observed from 26.3.2011 to 2.4.2011. In addition, all Sodankylä sounding profiles made during the period are illustrated. Again, it can be noted that the one-step algorithm produces more realistic results than the operative algorithm. The median of the operative profiles starts to produce too low values at 21 km outside and at 17 km

inside the polar vortex. The medians of the one-step profiles are as expected. The median of the profiles inside the polar vortex is visually close to the Sodankylä sounding profiles.

In this case study, the effect of the algorithm selection is drastic. This may not always be the case. The main difference comes from the priors used in the inversion algorithm. In particular, in this case, the strong smoothened requirements for aerosol profiles make the ozone perform well.

5. DISCUSSION

Although the operative two-step and the presented one-step algorithms consist of exactly the same elements, they are still fundamentally different: the one-step being statistically more sound, since some of the approximation can be avoided and the prior and the measurement errors are correctly treated together. Some of the algorithmic differences are discussed next.

The main difference of the retrieval algorithms comes from the use of the prior information. In the one-step algorithm, the prior given to one constituent affects the other constituent too. This can be clearly seen from the formulation of the cost function (10). The opposite is true for the operative algorithm, where the prior takes place only in the vertical inversion and is given for every constituent separately.

On the other hand, the resolution of the operative algorithm depends very little of the actual occultations. In particular, it is independent of the measurement noise that varies strongly from one star type to another. This allows the setting of the so-called target resolution, which makes the operative dataset user-friendly and easy to use in, e.g., time-series analysis and validation studies.[5]

In the one-step algorithm, similar target resolution cannot be set, since the resolution depends on the occultation depended measurement noise. In particular, the target resolution cannot be set, if the prior covariance matrix is kept fixed for all star types. In practice, this means that the smoothness regularization has to be selected for each star type separately, which is a challenging task.

From the case study, it can be noted that prior covariance matrix of the one-step retrieval algorithm can be "tuned" to obtain realistic results, where the signal-to-noise ratio is very low. In particular, it should be noted that the results obtained using one-step algorithm cannot be reproduced using the operative algorithm with any smoothness priors, since in the operative algorithm, the smoothness priors set to the aerosols do not affect the ozone profiles.

Motivated by the positive results of this case study, the future research is aimed at studying more in detail the capabilities of the one-step algorithm in the UTLS region, where the satellite retrievals are challenging. The computer code of the one-step algorithm is flexible for different kind of prior information and, e.g., the tropopause hight can be easily included.

6. CONCLUSION

In this paper, GOMOS one-step retrieval algorithm was presented, experimented and discussed. The algorithm proved it usefulness in the Arctic region and produced more realistic ozone-profile results than the operative algorithm at the lower stratosphere at 15–20 km altitude. It can be concluded that, because of the prototype nature of the retrieval algorithm, the one-step approach will suit best for the targeted case studies and will not replace the operative two-step algorithm.

ACKNOWLEDGMENTS

The authors would like to acknowledge the funding from Academy of Finland's MIDAT and LASBAY projects and European Space Agency's AERGOM project. The authors would like to thank Heikki Haario for commenting the manuscript.

REFERENCES

1. J. Bak, X. Liu, J. C. Wei, L. L. Pan, K. Chance, and J. H. Kim, "Improvement of OMI ozone profile retrievals in the upper troposphere and lower stratosphere by the use of a tropopause-based ozone profile climatology," *Atmos. Meas. Tech. Discuss.* **6**(3), pp. 4333–4369, 2013.
2. J. C. Wei, L. L. Pan, E. Maddy, J. V. Pittman, M. Divarkarla, X. Xiong, and C. Barnet, "Ozone profile retrieval from an advanced infrared sounder: Experiments with tropopause-based climatology and optimal estimation approach," *J. Atmos. Oceanic Technol.* **27**, pp. 1123–1139, 2013/07/18 2010.
3. J. L. Bertaux, E. Kyrölä, D. Fussen, A. Hauchecorne, F. Dalaudier, V. Sofieva, J. Tamminen, F. Vanhellemont, O. Fanton d'Andon, G. Barrot, A. Mangin, L. Blanot, J. C. Lebrun, K. Pérot, T. Fehr, L. Saavedra, G. W. Leppelmeier, and R. Fraisse, "Global ozone monitoring by occultation of stars: an overview of GOMOS measurements on ENVISAT," *Atmos. Chem. Phys.* **10**(24), pp. 12091–12148, 2010.
4. E. Kyrölä, J. Tamminen, V. Sofieva, J. L. Bertaux, A. Hauchecorne, F. Dalaudier, D. Fussen, F. Vanhellemont, O. Fanton D'Andon, G. Barrot, M. Guirlet, T. Fehr, and L. Saavedra de Miguel, "GOMOS O_3, NO_2, and NO_3 observations in 2002-2008," *Atmos. Chem. Phys.* **10**(16), pp. 7723–7738, 2010.
5. E. Kyrölä, J. Tamminen, V. Sofieva, J. L. Bertaux, A. Hauchecorne, F. Dalaudier, D. Fussen, F. Vanhellemont, O. Fanton d'Andon, G. Barrot, M. Guirlet, A. Mangin, L. Blanot, T. Fehr, L. Saavedra de Miguel, and R. Fraisse, "Retrieval of atmospheric parameters from GOMOS data," *Atmos. Chem. Phys.* **10**(23), pp. 11881–11903, 2010.
6. J. Tamminen, E. Kyrölä, V. F. Sofieva, M. Laine, J.-L. Bertaux, A. Hauchecorne, F. Dalaudier, D. Fussen, F. Vanhellemont, O. Fanton-d'Andon, G. Barrot, A. Mangin, M. Guirlet, L. Blanot, T. Fehr, L. Saavedra de Miguel, and R. Fraisse, "GOMOS data characterisation and error estimation," *Atmos. Chem. Phys.* **10**(19), pp. 9505–9519, 2010.
7. C. D. Rodgers, *Inverse methods for atmospheric sounding: theory and practice*, World Scientific, 2000.
8. F. Vanhellemont, D. Fussen, and C. Bingen, "Global one-step inversion of satellite occultation measurements: A practical method," *J. Geophys. Res.* **109**(D9), 2004.
9. H. Haario, M. Laine, M. Lehtinen, E. Saksman, and J. Tamminen, "Markov chain Monte Carlo methods for high dimensional inversion in remote sensing," *J. R. Stat. Soc. (B)* **66**(3), pp. 591–607, 2004.
10. P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman & Hall, London, second ed., 1989.
11. V. F. Sofieva, J. Tamminen, H. Haario, E. Kyrölä, and M. Lehtinen, "Ozone profile smoothness as a priori information in the inversion of limb measurements," *Ann. Geophys.* **22**(10), pp. 3411–3420, 2004.
12. G. L. Manney, M. L. Santee, M. Rex, N. J. Livesey, M. C. Pitts, P. Veefkind, E. R. Nash, I. Wohltmann, R. Lehmann, L. Froidevaux, L. R. Poole, M. R. Schoeberl, D. P. Haffner, J. Davies, V. Dorokhov, H. Gernandt, B. Johnson, R. Kivi, E. Kyrö, N. Larsen, P. F. Levelt, A. Makshtas, C. T. McElroy, H. Nakajima, M. Concepcion Parrondo, D. W. Tarasick, P. von der Gathen, K. A. Walker, and N. S. Zinoviev, "Unprecedented Arctic ozone loss in 2011," *NATURE* **478**, pp. 469–U65, OCT 27 2011.
13. P. Levelt, E. Hilsenrath, G. Leppelmeier, G. H. J. Van den Oord, P. Bhartia, J. Tamminen, J. de Haan, and J. Veefkind, "Science objectives of the ozone monitoring instrument," *IEEE Trans. Geos. Remote Sensing* **44**(5), pp. 1199–1208, 2006.

ACTA UNIVERSITATIS LAPPEENRANTAENSIS

504. IKONEN, MIKA. Power cycling lifetime estimation of IGBT power modules based on chip temperature modeling. 2012. Diss.

505. LEIVO, TIMO. Pricing anomalies in the Finnish stock market. 2012. Diss.

506. NISKANEN, ANTTI. Landfill gas management as engineered landfills – Estimation and mitigation of environmental aspects. 2012. Diss.

507. QIU, FENG. Surface transformation hardening of carbon steel with high power fiber laser. 2012. Diss.

508. SMIRNOV, ALEXANDER. AMB system for high-speed motors using automatic commissioning. 2012. Diss.

509. ESKELINEN, HARRI, ed. Advanced approaches to analytical and systematic DFMA analysis. 2013.

510. RYYNÄNEN, HARRI. From network pictures to network insight in solution business – the role of internal communication. 2013. Diss.

511. JÄRVI, KATI. Ecosystem architecture design: endogenous and exogenous structural properties. 2013. Diss.

512. PIILI, HEIDI. Characterisation of laser beam and paper material interaction. 2013. Diss.

513. MONTO, SARI. Towards inter-organizational working capital management. 2013. Diss.

514. PIRINEN, MARKKU. The effects of welding heat input usability of high strength steels in welded structures. 2013. Diss.

515. SARKKINEN, MINNA. Strategic innovation management based on three dimensions diagnosing innovation development needs in a peripheral region. 2013. Diss.

516. MAGLYAS, ANDREY. Overcoming the complexity of software product management. 2013. Diss.

517. MOISIO, SAMI. A soft contact collision method for real-time simulation of triangularized geometries in multibody dynamics. 2013. Diss.

518. IMMONEN, PAULA. Energy efficiency of a diesel-electric mobile working machine. 2013. Diss.

519. ELORANTA, LEENA. Innovation in a non-formal adult education organisation – multi-case study in four education centres. 2013. Diss.

520. ZAKHARCHUK, IVAN. Manifestation of the pairing symmetry in the vortex core structure in iron-based superconductors. 2013. Diss.

521. KÄÄRIÄINEN, MARJA-LEENA. Atomic layer deposited titanium and zinc oxides; structure and doping effects on their photoactivity, photocatalytic activity and bioactivity. 2013. Diss.

522. KURONEN, JUHANI. Jatkuvan äänitehojakautuman algoritmi pitkien käytävien äänikenttien mallintamiseen. 2013. Diss.

523. HÄMÄLÄINEN, HENRY. Identification of some additional loss components in high-power low-voltage permanent magnet generators. 2013. Diss.

524. SÄRKKÄ, HEIKKI. Electro-oxidation treatment of pulp and paper mill circulating waters and wastewaters. 2013. Diss.

525. HEIKKINEN, JANI. Virtual technology and haptic interface solutions for design and control of mobile working machines. 2013. Diss.

526. SOININEN, JUHA. Entrepreneurial orientation in small and medium-sized enterprises during economic crisis. 2013. Diss.

527. JÄPPINEN, EERO. The effects of location, feedstock availability, and supply-chain logistics on the greenhouse gas emissions of forest-biomass energy utilization in Finland. 2013. Diss.

528. SÖDERHOLM, KRISTIINA. Licensing model development for small modular reactors (SMRs) – focusing on the Finnish regulatory framework. 2013. Diss.

529. LAISI, MILLA. Deregulation's impact on the railway freight transport sector's future in the Baltic Sea region. 2013. Diss.

530. VORONIN, SERGEY. Price spike forecasting in a competitive day-ahead energy market. 2013. Diss.

531. PONOMAREV, PAVEL. Tooth-coil permanent magnet synchronous machine design for special applications. 2013. Diss.

532. HIETANEN, TOMI. Magnesium hydroxide-based peroxide bleaching of high-brightness mechanical pulps. 2013. Diss.

533. TYKKÄLÄ, TOMMI M. Real-time image-based RGB-D camera motion tracking and environment mapping. 2013. Diss.

534. PEKKOLA, SANNA. Performance measurement and management in a collaborative network. 2013. Diss.

535. PANOREL, IRIS CHERRY. Pulsed corona discharge as an advanced oxidation process for the degradation of organic compounds in water. 2013. Diss.

536. TORKKELI, LASSE. The influence of network competence of internationalization of SMEs. 2013. Diss.

537. MOLANDER, SOLE. Productivity and services – safety telephone services for the elderly. 2013. Diss.

538. SITARZ, ROBERT. Identification of research trends in the field of separation processes. Application of epidemiological model, citation analysis, text mining, and technical analysis of the financial markets. 2013. Diss.

539. KATTEDEN, KAMIEV. Design and testing of an armature-reaction-compensated permanent magnet synchronous generator for island operation. 2013. Diss.

540. HÄMÄLÄINEN, HARRI. Integration of learning supportive applications to development of e-portfolio construction process. 2013. Diss.

541. RATCHANANUSORN, WARIN. Development of a process for the direct synthesis of hydrogen peroxide in a novel microstructured reactor. 2013. Diss.

542. PERFILEV, DANIIL. Methodology for wind turbine blade geometry optimization. 2013. Diss.

543. STROKINA, NATALIYA. Machine vision methods for process measurements in pulping. 2013. Diss.

544. MARTTONEN, SALLA. Modelling flexible asset management in industrial maintenance companies and networks. 2013. Diss.