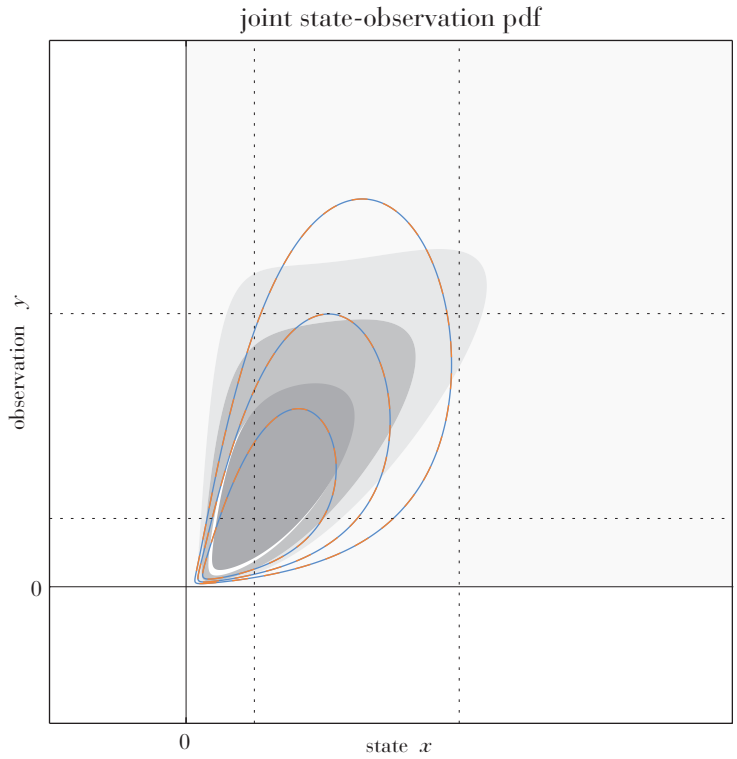
Max-Planck-Institut
für Meteorologie

MAX-PLANCK-GESELLSCHAFT

International Max Planck Research School
on Earth System Modelling

Analysis and application of the ensemble Kalman filter for the estimation of bounded quantities



Gernot Geppert

Hamburg 2015

Berichte zur Erdsystemforschung
Reports on Earth System Science

168
2015

Hinweis

Die Berichte zur Erdsystemforschung werden vom Max-Planck-Institut für Meteorologie in Hamburg in unregelmäßiger Abfolge herausgegeben.

Sie enthalten wissenschaftliche und technische Beiträge, inklusive Dissertationen.

Die Beiträge geben nicht notwendigerweise die Auffassung des Instituts wieder.

Die "Berichte zur Erdsystemforschung" führen die vorherigen Reihen "Reports" und "Examensarbeiten" weiter.

Notice

The Reports on Earth System Science are published by the Max Planck Institute for Meteorology in Hamburg. They appear in irregular intervals.

They contain scientific and technical contributions, including Ph. D. theses.

The Reports do not necessarily reflect the opinion of the Institute.

The "Reports on Earth System Science" continue the former "Reports" and "Examensarbeiten" of the Max Planck Institute.

Anschrift / Address

Max-Planck-Institut für Meteorologie
Bundesstrasse 53
20146 Hamburg
Deutschland

Tel./Phone: +49 (0)40 4 11 73 - 0
Fax: +49 (0)40 4 11 73 - 298

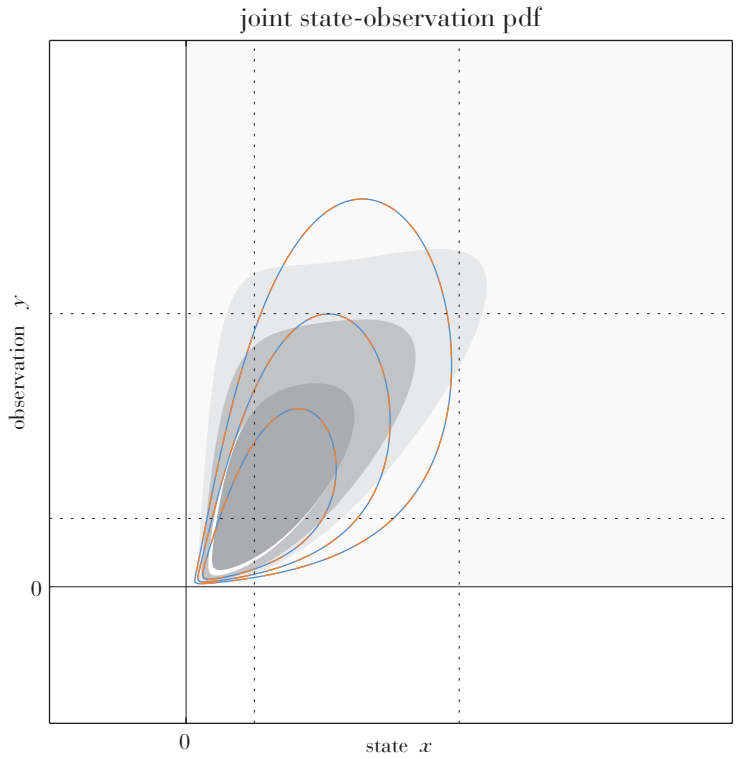
name.surname@mpimet.mpg.de
www.mpimet.mpg.de

Layout

Bettina Diallo and Norbert P. Noreiks
Communication

Copyright

Analysis and application of the ensemble Kalman filter for the estimation of bounded quantities



Gernot Geppert

Hamburg 2015

Gernot Geppert
aus Gera

Max-Planck-Institut für Meteorologie
Bundesstrasse 53
20146 Hamburg

Als Dissertation angenommen
vom Fachbereich Geowissenschaften der Universität Hamburg

auf Grund der Gutachten von
Prof. Dr. Felix Ament
und
Dr. Alexander Löw

Hamburg, den 17. Juni 2014
Professor Dr. Christian Betzler
Leiter des Departments Geowissenschaften

Berichte zur Erdsystemforschung / Max-Planck-Institut für Meteorologie 168
Reports on Earth System Science / Max Planck Institute for Meteorology 2015

ISSN 1614-1199

Abstract

The Kalman filter and its Monte Carlo approximation, the ensemble Kalman filter (EnKF), are best suited to problems involving unbiased, Gaussian errors. Non-Gaussian error distributions induced by bounded quantities make the EnKF sub-optimal and cause biased estimates. Further, EnKF estimates of bounded quantities may violate physical bounds and lead to a failure of the involved model. Extending the EnKF with a nonlinear variable transformation technique can mitigate the first and solve the second problem.

Motivated by a parameter estimation problem from land surface modelling, we analyse the effects of non-Gaussian distributions and non-zero mean errors on EnKF estimates theoretically and experimentally. For the first time, we use a linear regression framework to qualitatively examine and explain errors in the EnKF estimates and we analyse their behaviour with and without variable transformations. From theoretical considerations, we derive a covariance scaling approach for the estimation of the transformed observation error covariance that ensures a constant transformed observation error covariance, independent of the observed value.

Comparing estimates derived with the new covariance scaling approach, with two other transformation-based approaches, and with the EnKF without variable transformation, we find that covariance scaling is superior to the other methods with respect to the quality of the estimates (for all other methods) and with respect to its computational cost (for all methods except the EnKF without anamorphosis).

We verify these findings in a series of data assimilation experiments using synthetic land surface albedo observations and a newly implemented data assimilation framework based on the dynamic global vegetation model JSBACH and the Data Assimilation Research Testbed.

Contents

Chapter 1

Introduction

1.1 Constrained data assimilation with the ensemble Kalman filter

Our knowledge about the state of the Earth system originates from models and observations. Both are uncertain due to various sources of error but they often contain complementary information. Data assimilation combines this complementary information to reduce the uncertainty in the combined estimate of the system's state. The ensemble Kalman filter (EnKF; Evensen, 1994) is a data assimilation method which is simple to implement and which has become ubiquitous in geophysical research (cf. references in Evensen, 2009a). Despite its apparent simplicity, the EnKF is a powerful tool for the successive combination of observational data with a numerical model.

The EnKF is linked to Bayesian estimation (van Leeuwen and Evensen, 1996) as well as minimum variance (Gelb, 1974) and least squares techniques (Duncan and Horn, 1972). But no matter how we derive and interpret the EnKF, it is contingent on strong assumptions. And the quality of the EnKF estimates is contingent on the compliance of the model and the observations with these assumptions. The two most restricting assumptions concern, on the one hand, the character of the uncertainty of the model state and of the observations and, on the other hand, how the observations are related to the model state. For the description of the uncertainty, the EnKF requires Gaussian distributions and for the link between states and observations, the EnKF requires a linear observation operator.

The EnKF is a statistical estimator that builds both on the Gaussian and the linear assumption. And owing to this purely statistical nature, EnKF estimates do not automatically satisfy physical constraints like boundedness. For any estimate to be useful, however, such constraints have to be met and various modifications of the EnKF have been proposed. The efforts to constrain the EnKF to bounded domains can be broadly categorised into three types:

1. ad-hoc approaches that replace unphysical values with compliant ones,
2. constrained optimisation approaches, and
3. variable transformation approaches.

Constrained optimisation approaches (Pan and Wood, 2006; Yilmaz et al., 2011; Janjić et al., 2014) consider the EnKF from the viewpoint of mathematical optimisation theory. In this sense, the EnKF minimises the misfit between the estimate and the observations as well as between the estimate and the prior data, that is model forecasts. Adding constraints to the otherwise unconstrained optimisation problem ensures physically consistent estimates in this approach.

Variable transformation approaches map the quantities in the state vector from the model's physical space to an unbounded domain for the estimation and then back to the physical, bounded domain afterwards (Bertino et al., 2002; Nielsen-Gammon et al., 2010; Schirber et al., 2013). Taking into account the effect of the variable transformation on the state variable distributions and on the observation error distributions, this approach preserves the Bayesian character of the EnKF and leads to the Gaussian anamorphosis technique (Bertino et al., 2002, 2003; Simon and Bertino, 2009). Gaussian anamorphosis refers to a variable transformation that renders the distribution of transformed state variable and of the transformed observation errors Gaussian. The reasoning behind Gaussian anamorphosis is that the transformed variables will be more compliant with the EnKF assumptions than the model space variables. Consequently, the application of the EnKF to the transformed variables yields better estimates.

1.2 Land data assimilation systems and the assimilation of albedo observations

The use of the EnKF for the assimilation of observations into land surface and vegetation models is dominated by hydrological applications (Reichle et al., 2002, 2007; Moradkhani et al., 2005; Hendricks Franssen and Kinzelbach, 2008; Schöniger et al., 2012) and carbon cycle data assimilation systems (Williams et al., 2005; Chatterjee and Michalak, 2013). Other applications include the assimilation of observations of the fraction of absorbed photosynthetically active radiation and leaf area index to estimate vegetation parameters of a phenology model (Stöckli et al., 2008, 2011).

None of the aforementioned studies used albedo observations. And neither does any of the prevalent variational assimilation frameworks such as the Earth Observation Land Data Assimilation System (EO-LDAS; Lewis et al., 2012) or the Carbon Cycle Data Assimilation System (CCDAS; Rayner et al., 2005; Kaminski et al., 2013). The only studies known to us that used albedo observations in a data assimilation system are related to snow and snow albedo (Durand and Margulis, 2007; Dumont et al., 2012; Malik et al., 2012).

To ensure physically consistent estimates in land data assimilation systems, ad-hoc approaches, constrained optimisation, and variable transformation techniques are used. In the ad-hoc methods, the unphysical estimates are shifted to the physical domain (Stöckli et al., 2011; Lewis et al., 2012). In the constrained optimisation approach of the variational

EO-LDAS, the allowed values of estimates are confined to a bounded domain. And in the variable transformation techniques that are currently being explored for CCDAS, bounded parameters are transformed to unbounded ones (Kemp et al., 2014). None of these approaches, however, uses transformed observations. Gaussian anamorphosis does exactly that but, to our knowledge, the applications of Gaussian anamorphosis for land surface models are limited to hydrological parameters of the soil (Zhou et al., 2011; Schöniger et al., 2012).

1.3 Error sources in the ensemble Kalman filter

Bounded quantities like albedo follow non-Gaussian distributions and can introduce nonlinearities in the relation between the model state and the observations. The EnKF becomes a sub-optimal estimator in such cases (Bertino et al., 2003) and a variety of modifications have been proposed to overcome these limitations. For example, nonlinear observations are commonly handled by state augmentation (Evensen, 2003) and different approaches modify the EnKF for non-Gaussian distributions (Lauvernet et al., 2009; Anderson, 2010; Lei and Bickel, 2011). Further, the effects of non-Gaussian state distributions on the updated ensembles in different versions of the EnKF have been previously explored (Lawson and Hansen, 2004; Lei et al., 2010).

While there are numerous suggestions how to mitigate the adverse effects of non-Gaussianity and nonlinearity, these effects themselves, that is the estimation errors, have not yet been explored rigorously. In particular, the effects of state-dependent observation error distributions have not yet been explored. Pires et al. (2010) state that heteroscedastic observation errors cause non-Gaussianity and Lei and Bickel (2009, 2011) implicitly include state-dependent observation errors in their theories. But an explanation of how such deviations from the standard EnKF assumptions impact the EnKF estimates has not yet been given.

1.4 Research questions and contributions

The goal of this thesis is to explain the effects of deviating from the standard EnKF assumptions and the resulting estimation errors. In particular, we explore the case of state-dependent observation error distributions. The insights from this analysis lead us to the development of a new way to estimate the transformed observation error covariance when using Gaussian anamorphosis.

Our research is motivated by the analysis of a new data set of radiative transfer parameters for vegetation canopies in chapter 2. These parameters describe the albedo of vegetation canopies and are constrained to the interval [0, 1]. The emerging question is:

- Can we retrieve a climatology of canopy albedo parameters from observations of land surface albedo with the ensemble Kalman filter and Gaussian anamorphosis?

In chapter 3, we present the theory of Kalman filtering and Gaussian anamorphosis that is necessary to answer this question. To analyse the error sources in the EnKF and their impact on the EnKF estimates, we use the framework of linear regression (section 3.6). Linear regression has been related to the Kalman filter before (Duncan and Horn, 1972), but it has not been used to understand the effects of nonlinearity, non-Gaussianity and state-dependent observation errors in the EnKF. We provide an explanation of these effects using linear regression theory.

Further, we derive a statistical framework for the characterisation of the errors of transformed observations when Gaussian anamorphosis is used with the EnKF. We use this framework to justify and modify an existing method for the transformation of observation error covariances. We then suggest a new a method for the transformed observation error covariances that overcomes statistical and computational drawbacks of the previous method (section 3.7.5).

Finally, we compare our new method and the modified method with a direct method that does not require transformed observation error covariances and with the EnKF without Gaussian anamorphosis. This comparison provides an answer to the question:

- What is the best method (out of these four) to assimilate albedo observations with the ensemble Kalman filter from a theoretical point of view?

In chapter 4, we apply the four methods in data assimilation experiments using the EnKF for the assimilation of synthetic observations of land surface albedo into a comprehensive land surface model. The results of these experiments verify our theoretical findings.

Chapter 5 provides a summary of our results and our conclusions. We give recommendations for the assimilation of real observations as well as for applications of our findings to other quantities than canopy albedo. Lastly, we close with an outlook suggesting further developments.

Chapter 2

Canopy albedo in Earth system models and observations

2.1 Land surface albedo in the Max Planck Institute Earth System Model

Surface albedo is the most influential parameter on the surface energy budget because it largely determines the amount of available energy for latent and sensible heat fluxes. These fluxes affect the circulation and the climate locally as well as globally (Charney et al., 1977; Sud and Fennessy, 1982; Sellers, 1997). Hence, Earth system models require an accurate description of surface albedo. Since land covers approximately 30% of the Earth's surface, land surface albedo is an essential part of this description and Sellers et al. (1995) suggest an absolute accuracy of ± 0.02 for the surface albedo in land surface models.

The surface albedo of vegetation-covered areas depends on the vegetation layer and the background below. An accurate description of the albedo of vegetated surfaces requires calculations of the radiative transfer through the vegetation canopy. Approximate solutions of this problem are available and form one approach to simulate land surface albedo in Earth system models (Sellers, 1985; Yuan et al., 2014). The land component of the Max Planck Institute Earth System Model (MPI-ESM; Giorgetta et al., 2013) uses a different approach that avoids radiative transfer calculations. Instead, the dynamic global vegetation model JSBACH (Raddatz et al., 2007; Reick et al., 2013), which is the land component of the MPI-ESM, partitions the land surface in vegetation canopy and background to calculate the surface albedo (Rechid et al., 2009; Vamborg et al., 2011).

For snow-free surfaces, JSBACH calculates the surface albedo α of a homogeneously covered part of a model grid box as a weighted average of background albedo α_{bg} and canopy albedo α_{c}, that is,

$$\alpha = f_{\mathrm{c}}\, \alpha_{\mathrm{c}} + (1 - f_{\mathrm{c}})\, \alpha_{\mathrm{bg}}.$$

The canopy fraction f_{c} is calculated from the prognostic leaf area index (LAI) and from

the fraction $V_{\max}$ of the grid box that is covered by vegetation according to

$$f_c = V_{\max} \left(1 - \exp\left(-\frac{LAI}{2}\right)\right).$$

Within a model grid box, different cover types may be present. JSBACH uses plant functional types (PFTs) to represent different cover types and assigns a fraction of the grid box, called a tile, to each PFT. The surface albedo of the whole grid box is then calculated as a weighted average of these tiles with the weights given by the cover fractions.

The canopy albedo α_c is a PFT-specific parameter and the background albedo α_{bg} of a grid box is given by a global map of background albedos projected onto the model grid. JSBACH simulates the land surface albedo in the visible (0.4 – 0.7 µm) and the near-infrared (0.7 – 4.0 µm) domain. Therefore, the canopy albedo and background albedo parameters are also differentiated for these two spectral domains. The currently used values for canopy and background albedo were derived from a linear regression of albedo observations on observations of the fraction of photosynthetically active radiation (fAPAR) from the Moderate Resolution Imaging Spectroradiometer (MODIS) as described in Rechid et al. (2009).

2.2 Seasonal behaviour of canopy albedo

The canopy albedo parameters in JSBACH are constant in time. Without changes in the PFT distribution, changes in snow-free surface albedo are only due to changes in the simulated LAI. Observational studies, however, find changing canopy albedos during the growing season which also affect the total surface albedo. These changes in the observed canopy albedo are attributed either to changing nitrogen levels in the canopy (Ollinger et al., 2008; Hollinger et al., 2010) or, objecting to the nitrogen hypothesis, to structural changes within the canopy (Knyazikhin et al., 2013). Both suggestions are based on correlations between in-situ or remote sensing observations of surface albedo over dense canopies with either the nitrogen content of the canopy or structural variables such as the broad-leaf fraction.

The products from the Joint Research Centre Two-stream Inversion Package (JRC-TIP; Pinty et al., 2011a,b) offer another possibility to examine the seasonality of canopy albedo. JRC-TIP uses a variational approach to retrieve the effective parameters of a two-stream radiative transfer model (Pinty et al., 2006) from white-sky albedo values derived from MODIS observations. These parameters include effective visible and near-infrared canopy single scattering albedo (SSA) at a spatial resolution of 0.01° for the years 2001 to 2010. For our analysis we used only values for which the posterior standard deviation was at least 75% smaller than the prior standard deviation used in the variational scheme and we resampled the results to a spatial resolution of 0.25°. Figure 2.1 shows the mean seasonal amplitude of the effective canopy SSA in the visible and the near infrared domain and an exemplary mean seasonal cycle of these two quantities at the location of the Hainich forest

in Germany. Relating the seasonal variations to the absolute magnitude of the visible and near-infrared effective SSA, we conclude that these quantities exhibit a seasonal behaviour that differs with location.

2.3 Seasonal canopy albedo parameters for JSBACH

The effective canopy parameters in the two-stream model of JRC-TIP are not quantitatively comparable to the canopy albedo parameters of JSBACH. But both describe the radiative properties of the canopy. Thus, qualitative insights from the analysis of the effective single scattering albedos in the JRC-TIP data set can be related to the canopy albedo parameters of JSBACH. This qualitative argument suggests that the canopy albedo parameters of JSBACH should possibly also follow a seasonal cycle.

The implications of seasonal canopy albedo parameters in JSBACH depend on the amplitude of the seasonal cycles of the visible and near-infrared parameters. The JRC-TIP products and physiological considerations (Gitelson and Merzlyak, 1996) indicate that these cycles will be opposed to each other, that is, decreasing visible canopy albedo and increasing near-infrared canopy albedo during summer time. When assuming similar amplitudes for both spectral domains as done in chapter 4 (Figure 4.3), the seasonal effects annihilate each other and the total upward shortwave flux remains nearly unchanged (Figure 2.2). The dependence of changes in the radiative balance on the seasonal cycles of the parameters underlines the importance of a climatology of canopy albedo parameters. Such a climatology would allow reliable statements about changes in the seasonal upward shortwave fluxes due to seasonal changes in canopy albedo.

Insights into the seasonal behaviour of the JSBACH canopy albedo parameters require a time series of the parameter values. But the canopy albedo parameters are effective model parameters without an observable equivalent. Neither the albedo of single leaves nor the apparent albedo of a closed canopy, that could both be measured, would be adequate to characterise the JSBACH canopy albedo. We therefore require a model inversion similar to JRC-TIP to retrieve a time series of JSBACH canopy albedo parameters from observations of JSBACH model states.

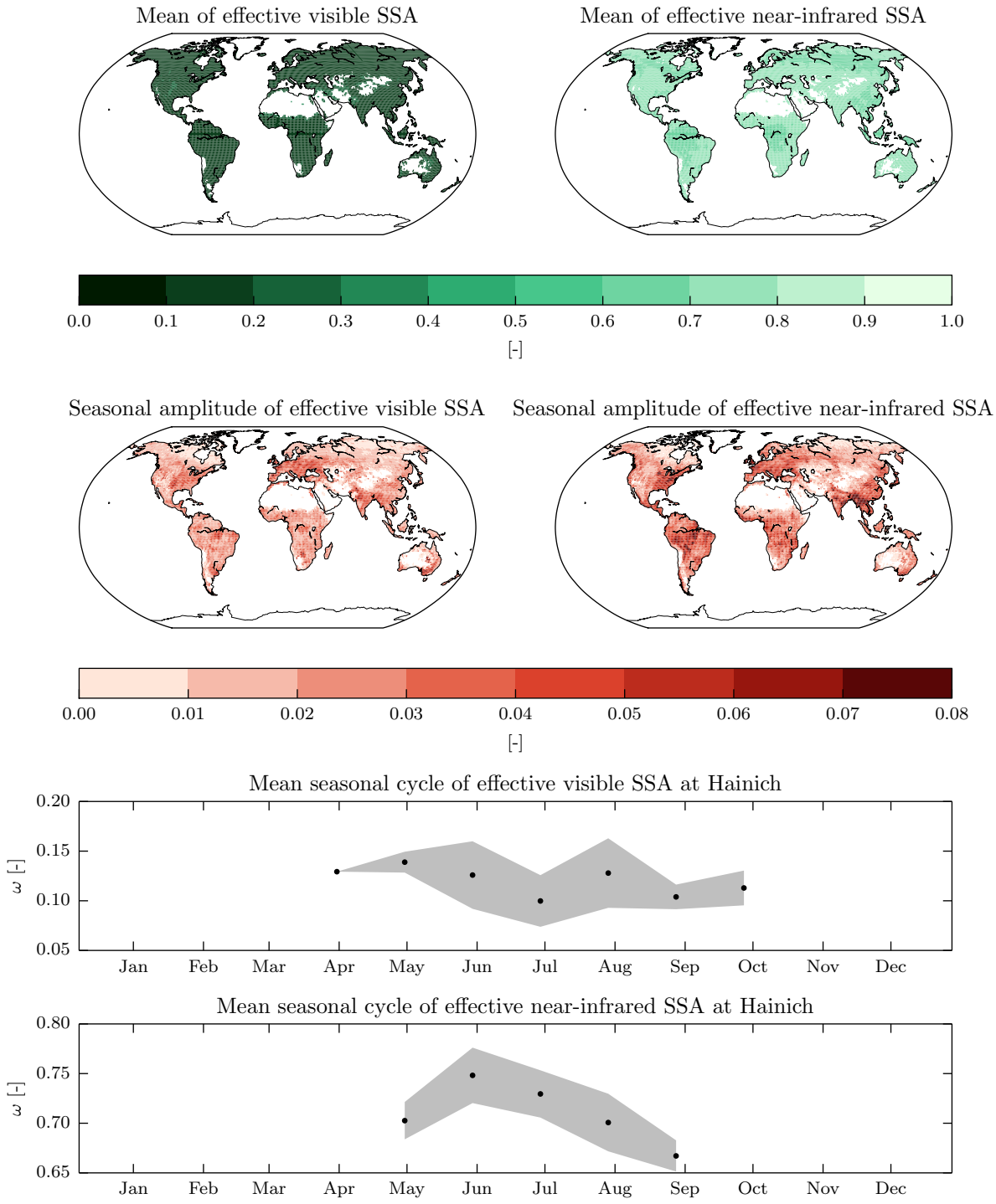


Figure 2.1: Amplitude of mean seasonal cycle of effective visible (upper left) and effective near-infrared (upper right) canopy single scattering albedo derived from JRC-TIP data from 2001-2010 and mean seasonal cycle for the location of the Hainich forest (51.09° N, 10.44° E) including multi-year standard deviations (white areas indicate no successful retrieval).

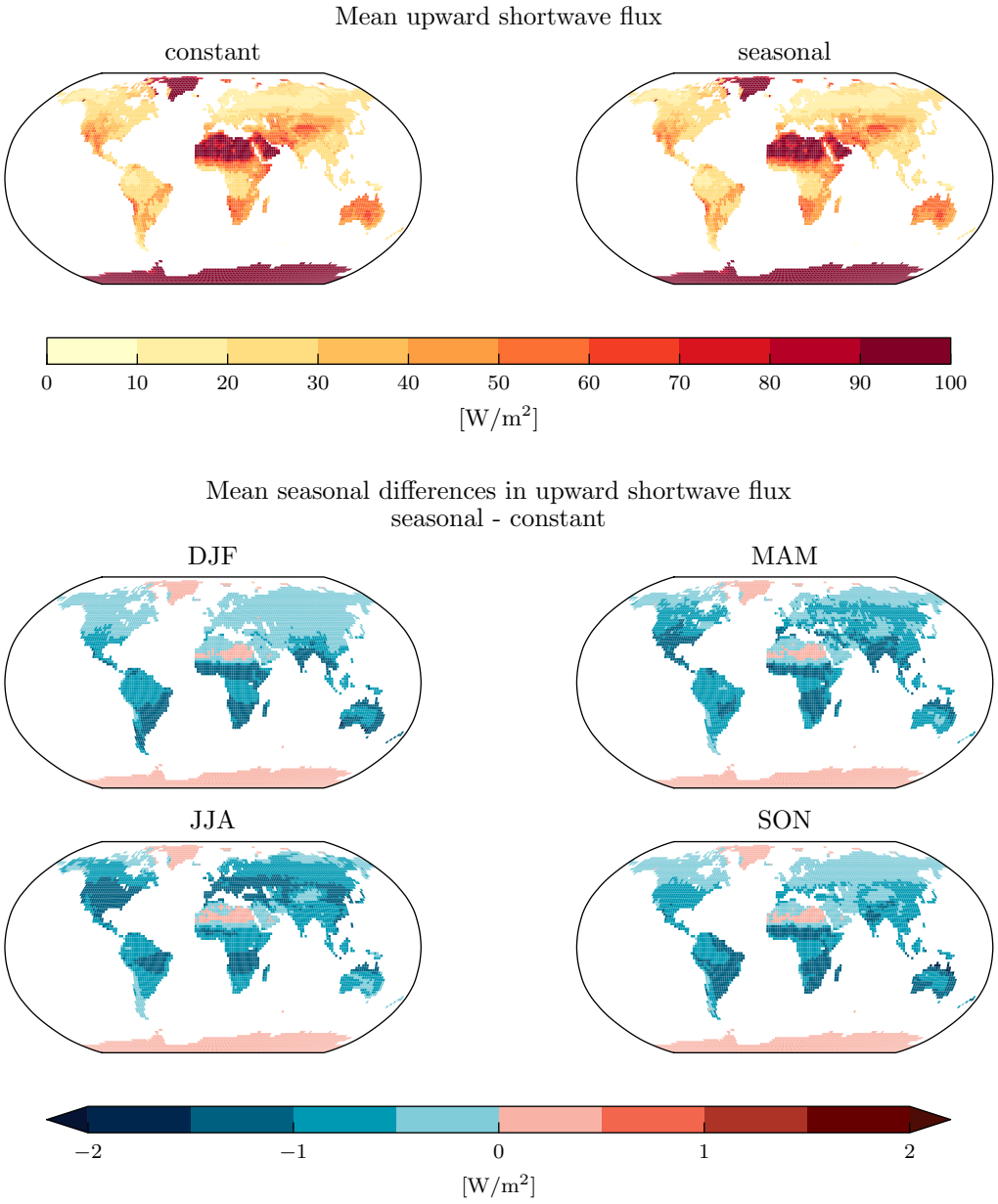


Figure 2.2: Mean values of upward shortwave flux and differences in seasonal upward shortwave flux between simulations with seasonal and constant canopy albedo parameters as prescribed in chapter 4 (Figure 4.3).

Chapter 3

Sequential data assimilation with the ensemble Kalman filter

3.1 Representation of uncertainty in models and observations

Numerical models and observations provide information about the state of a physical system but both are subject to errors which limit the credibility of this information. This lack of certainty in the output of a numerical model and in the output of a measurement device is called uncertainty. To go beyond qualitative statements and to quantify uncertainty, we need to derive the relevant errors and we need to specify statistical models which describe the available knowledge about these errors.

We follow Cohn (1997) and let the vector $\mathbf{s}(t_k) \in B$, where B is some appropriate function space, describe the system's true state at a given time t_k. The components of $\mathbf{s}(t_k)$ are functions of space and time and fully describe the individual variables of the system. Further let g describe the propagation of a state $\mathbf{s}(t_{k-1})$ over a fixed time interval from t_{k-1} to t_k as

$$\mathbf{s}(t_k) = g(\mathbf{s}(t_{k-1})). \tag{3.1}$$

A numerical model employs discretisations of the components of $\mathbf{s}(t_k)$ that form the true, discretised state vector $\mathbf{x}_k \in \mathbb{R}^n$. The mapping from $\mathbf{s}(t_k)$ to $\mathbf{x}_k$ is given by a projection operator $\mathbf{\Pi} : B \to \mathbb{R}^n$ as

$$\mathbf{x}_k = \mathbf{\Pi}(\mathbf{s}(t_k)). \tag{3.2}$$

The true, discretised state evolves as

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \boldsymbol{\eta}_k, \tag{3.3}$$

where f is the numerical model that propagates $\mathbf{x}_k$ and

$$\begin{aligned}\boldsymbol{\eta}_k &= \mathbf{x}_k - f(\mathbf{x}_{k-1}) \\ &= \boldsymbol{\Pi}(\mathbf{s}(t_k)) - f(\mathbf{x}_{k-1}) \\ &= \boldsymbol{\Pi}(g(\mathbf{s}(t_{k-1}))) - f(\mathbf{x}_{k-1}) \\ &= \boldsymbol{\Pi}(g(\mathbf{s}(t_{k-1}))) - f(\boldsymbol{\Pi}(\mathbf{s}(t_{k-1})))\end{aligned} \tag{3.4}$$

is the *model error* term. It describes the model's inability to predict the true, discretised future state and originates from errors in the model's formulation and forcing, on the one hand, and from errors due to numerical approximations, discretisation and round-off errors, on the other hand.

For the characterisation of the observation error, let the observations $\mathbf{y}_k \in \mathbb{R}^m$ at time t_k be given by

$$\mathbf{y}_k = m(\mathbf{s}(t_k)) + \boldsymbol{\varepsilon}_k^m, \tag{3.5}$$

where $m : B \to \mathbb{R}^m$ is the observation operator that maps the full state $\mathbf{s}(t_k)$ to discrete observations $\mathbf{y}_k$ and where $\boldsymbol{\varepsilon}_k^m$ is the measurement error of any involved instruments and devices (Cohn, 1997). The discretised state $\mathbf{x}_k$ is related to $\mathbf{y}_k$ through the discrete observation operator $h : \mathbb{R}^n \to \mathbb{R}^m$ as

$$\mathbf{y}_k = h(\mathbf{x}_k) + \boldsymbol{\varepsilon}_k. \tag{3.6}$$

Using (3.5) and inserting $-m(\boldsymbol{\Pi}(\mathbf{s}(t_k)) + m(\boldsymbol{\Pi}(\mathbf{s}(t_k))$ yields

$$\begin{aligned}\boldsymbol{\varepsilon}_k &= \mathbf{y}_k - h(\mathbf{x}_k) \\ &= m(\mathbf{s}(t_k)) + \boldsymbol{\varepsilon}_k^m - h(\mathbf{x}_k) \\ &= m(\mathbf{s}(t_k)) + \boldsymbol{\varepsilon}_k^m - h(\boldsymbol{\Pi}(\mathbf{s}(t_k))) \\ &= \boldsymbol{\varepsilon}_k^m + \underbrace{m(\mathbf{s}(t_k)) - m(\boldsymbol{\Pi}(\mathbf{s}(t_k))}_{\boldsymbol{\varepsilon}_k^r} + \underbrace{m(\boldsymbol{\Pi}(\mathbf{s}(t_k))) - h(\boldsymbol{\Pi}(\mathbf{s}(t_k)))}_{\boldsymbol{\varepsilon}_k^a}\end{aligned} \tag{3.7}$$

and shows that the discrete *observation error* $\boldsymbol{\varepsilon}_k$ consists of the measurement error $\boldsymbol{\varepsilon}_k^m$, the error due to unresolved scales or representativeness error $\boldsymbol{\varepsilon}_k^r$ (Lorenc, 1986), and the error $\boldsymbol{\varepsilon}_k^a$ from the approximation of m with h.

The system's state $\mathbf{s}(t_k)$ and frequently also the propagator g are unknown. And the errors $\boldsymbol{\eta}_k$ and $\boldsymbol{\varepsilon}_k$ together with the error of any initial discretised state are the sources of uncertainty about the discrete representations $\mathbf{x}_k$ and $\mathbf{y}_k$ of $\mathbf{s}(t_k)$. The error terms are just as unattainable as the true state - notably they depend on $\mathbf{s}(t_k)$. But once they have been identified as the sources of uncertainty, information about their characteristics can be obtained from controlled and repeated experiments. Subsequently, the available knowledge can be cast into statistical models which allow to quantify the uncertainty by

means of probabilities and probability density functions (pdfs).

The uncertain elements $\boldsymbol{\eta}_k$, $\boldsymbol{\varepsilon}_k$, and the uncertain initial condition $\mathbf{x}_0$ are now considered to be random variables that follow *known* probability distributions. The choice of these distributions is crucial for all further statements about uncertainty and the results of any data assimilation experiment. This choice is governed by the information that is available about the system of interest and about the observation process before the experiment starts and it is governed by statistical considerations such as the maximum entropy principle (Jaynes, 2007). With $\mathbf{x}_0$, $\boldsymbol{\eta}_k$, and $\boldsymbol{\varepsilon}_k$ being random variables, $\mathbf{x}_k$ and $\mathbf{y}_k$ also become random variables as they are now functions of at least one random variable. Within this probabilistic framework, the most comprehensive description of $\mathbf{x}_k$ based on observations $\mathbf{y}_1, \ldots, \mathbf{y}_k$ is the conditional pdf $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k)$. Finding $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k)$ is called the *filtering problem* and its solution is commonly found making three basic assumptions about $\boldsymbol{\eta}_k$ and $\boldsymbol{\varepsilon}_k$:

- the pdfs of $\boldsymbol{\eta}_k$ are known for all t_k and $\boldsymbol{\eta}_k$ is white in time, that is, $\boldsymbol{\eta}_j$ is independent of $\boldsymbol{\eta}_k$ for all time steps $t_j \neq t_k$ and $\boldsymbol{\eta}_k$ has mean zero and finite variance,
- the pdfs of $\boldsymbol{\varepsilon}_k$ are known for all t_k and $\boldsymbol{\varepsilon}_k$ is white in time,
- $\boldsymbol{\eta}_k$ is independent of $\boldsymbol{\varepsilon}_j$ for all time steps t_j and t_k (Cohn, 1997).

Under these assumptions, the state equation (3.3) and the observation equation (3.6),

$$\begin{aligned}\mathbf{x}_k &= f(\mathbf{x}_{k-1}) + \boldsymbol{\eta}_k,\\ \mathbf{y}_k &= h(\mathbf{x}_k) + \boldsymbol{\varepsilon}_k,\end{aligned}$$

form a hidden Markov model (Marin and Robert, 2007).

The general solution of the filtering problem is given by Bayes' Theorem (Jazwinski, 1970),

$$p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k) = \frac{p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})\, p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})}. \tag{3.8}$$

From the definition of conditional pdfs and marginal pdfs the denominator is

$$\begin{aligned}p(\mathbf{y}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}) &= \int p(\mathbf{y}_k, \mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})\mathrm{d}\mathbf{x}_k\\ &= \int p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})\, p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1})\mathrm{d}\mathbf{x}_k,\end{aligned}$$

which is the integral of the product in the numerator and thus only normalises this product such that the right hand side of (3.8) is a pdf (Jazwinski, 1970). Bayes' Theorem states that the conditional or *posterior* pdf $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k)$ is proportional to the product of the *prior* pdf $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})$ and of the *likelihood* $p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1})$. The expression for the likelihood simplifies to $p(\mathbf{y}_k|\mathbf{x}_k)$ due to the independence of observation and model errors. With this simplification, the likelihood is given by the pdf of the observation error

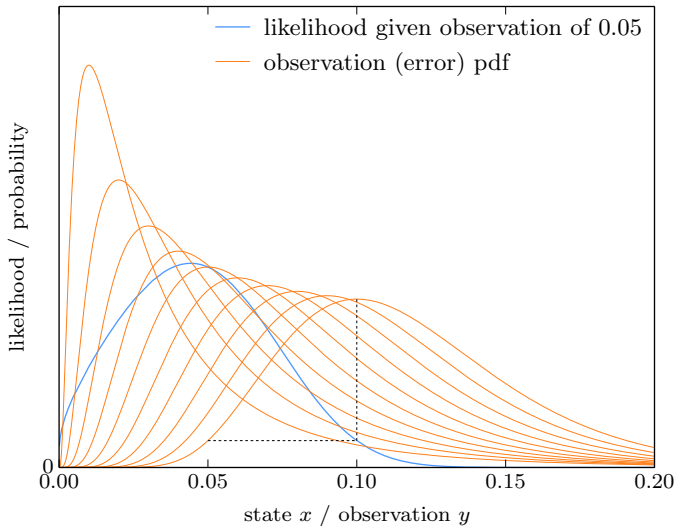


Figure 3.1: Construction of likelihood from observation error pdfs for an identity observation $y = h(x) = x$ at $y = 0.05$. The observation error pdfs $p_{\varepsilon(x)}$ are shifted to the state $h^{-1}(y) = x$ for all x and evaluated at 0.05. These values are then assigned to the likelihood at x. For example, the value of the observation error pdf given a state value of 0.1 at the observation 0.05 is the likelihood of the state 0.1 given the observation 0.05 (the horizontal axis represents the state space as well as the observation space because of the identity observation operator).

$p_{\boldsymbol{\varepsilon}_k}(\boldsymbol{\varepsilon}_k)$ and by a change of variable from $\boldsymbol{\varepsilon}_k$ to $\mathbf{y}_k - h(\mathbf{x}_k)$ according to (3.6). The likelihood then reads (Jazwinski, 1970)

$$p(\mathbf{y}_k|\mathbf{x}_k) = p_{\boldsymbol{\varepsilon}_k}(\mathbf{y}_k - h(\mathbf{x}_k)). \tag{3.9}$$

We note that the likelihood is a function of $\mathbf{x}_k$ and not $\mathbf{y}_k$ because the observations $\mathbf{y}_k$ are fixed parameters in the filtering problem. The likelihood is therefore not necessarily a pdf (Jaynes, 2007) and may also not be interpreted as a probability as in the concept of the chance of a future event. It should rather be understood as a measure of how likely any state $\mathbf{x}_k$ has caused the given observation $\mathbf{y}_k$. In terms of pdfs, this means that the likelihood of a state $\mathbf{x}_k$ is given by the conditional pdf $p(\mathbf{y}_k|\mathbf{x}_k)$ evaluated at the given observation $\mathbf{y}_k$ (Figure 3.1). Since $p(\mathbf{y}_k|\mathbf{x}_k)$ is the pdf of the observation error $\boldsymbol{\varepsilon}_k$ whose distribution may depend on the observed state $\mathbf{x}_k$, not only the location of $p(\mathbf{y}_k|\mathbf{x}_k)$ but also its shape may depend on $\mathbf{x}_k$. Consequently the likelihood of $\mathbf{x}_k$ given $\mathbf{y}_k$ may have little resemblance with the observation error distributions although they are closely related to each other.

In contrast to this retrospective information about $\mathbf{x}_k$, the prior pdf $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})$ describes the probability of $\mathbf{x}_k$ in a prognostic sense given the available information at time t_{k-1}. Again using the independence of the errors, the prior pdf is given by the

Chapman-Kolmogorov equation as (Jazwinski, 1970)

$$p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1}|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}) \mathrm{d}\mathbf{x}_{k-1}. \tag{3.10}$$

The so called transition density $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is derived with the same change of variable argument as for the likelihood from (3.3) as

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p_{\boldsymbol{\eta}_k}(\mathbf{x}_k - f(\mathbf{x}_{k-1})), \tag{3.11}$$

where $p_{\boldsymbol{\eta}_k}$ is the pdf of the model error $\boldsymbol{\eta}_k$ introduced in (3.3).

Accepting the assumptions on the errors $\boldsymbol{\eta}_k$ and $\boldsymbol{\varepsilon}_k$, the posterior pdf $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k)$ can be found using a recursive algorithm that proceeds sequentially in time (Gordon et al., 1993). And in conclusion, the Chapman-Kolmogorov equation and Bayes' Theorem form a recurrence relation that allows a recursive solution of the filtering problem as in

Algorithm 1

1. **initialise** the forecast distribution $p(\mathbf{x}_0)$,
2. **for** i **from** 1 **to** k**:**
   a) **forecast** the prior pdf $p(\mathbf{x}_i|\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$,
   b) **update** the forecast with the observation $\mathbf{y}_i$ to find the posterior pdf $p(\mathbf{x}_i|\mathbf{y}_1, \ldots, \mathbf{y}_i)$.

3.2 The Kalman filter and the ensemble Kalman filter

The Kalman filter (KF) is a special case of Algorithm 1 for linear models and Gaussian errors (Kalman, 1960; Kalman and Bucy, 1961; see also Cohn, 1997). It was derived minimising expected squared errors. But next to this minimum variance interpretation, the KF solution corresponds to the maximum likelihood solution as well as to the recursive, weighted least squares estimate of a state given past observations (Jazwinski, 1970). In the context of the filtering problem, Ho and Lee (1964) and van Leeuwen and Evensen (1996) noted its recursive Bayesian character for linear, Gaussian problems. Because Gaussian distributions are fully characterised by their mean and covariance, only solutions for mean and covariance – as provided by the KF – are required to solve the filtering problem. The ensemble Kalman filter (EnKF) uses a Monte Carlo method to approximate the KF with less computational effort and to extend its applicability to nonlinear models (Evensen, 1994).

3.2.1 The Kalman filter

The KF builds on the two facts that linear transformations of Gaussian random variables again yield Gaussian random variables and that the product of Gaussian pdfs is again Gaussian.

Consider a linear forecast model $\mathbf{M} \in \mathbb{R}^{n\times n}$ and a linear observation operator $\mathbf{H} \in \mathbb{R}^{m\times n}$,

$$\mathbf{x}_k = \mathbf{M}\mathbf{x}_{k-1} + \boldsymbol{\eta}_k, \tag{3.12}$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \boldsymbol{\varepsilon}_k. \tag{3.13}$$

And assume that the errors $\boldsymbol{\eta}_k$ and $\boldsymbol{\varepsilon}_k$ follow Gaussian distributions with mean zero and known, constant covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ and assume further that the mean and the covariance of the initial state $\mathbf{x}_0$ are $\mathbf{x}_0^a$ and $\mathbf{P}_0^a$,

$$\boldsymbol{\eta}_k \sim N(0,\, \mathbf{Q}), \tag{3.14}$$

$$\boldsymbol{\varepsilon}_k \sim N(0,\, \mathbf{R}), \tag{3.15}$$

$$\mathbf{x}_0 \sim N(\mathbf{x}_0^a,\, \mathbf{P}_0^a). \tag{3.16}$$

Then, the mean $\mathbf{x}_{k-1}^a$ and the covariance matrix $\mathbf{P}_{k-1}^a$ of $p(\mathbf{x}_{k-1}|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})$ evolve as

$$\mathbf{x}_k^f = \mathbf{M}\mathbf{x}_{k-1}^a, \tag{3.17}$$

$$\mathbf{P}_k^f = \mathbf{M}\mathbf{P}_{k-1}^a\mathbf{M}^T + \mathbf{Q} \tag{3.18}$$

and $\mathbf{x}_k^f$ and $\mathbf{P}_k^f$ are the mean and the covariance matrix of the forecast distribution $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})$. Thus, (3.17) and (3.18) solve the Chapman-Kolmogorov equation (3.10) for linear, Gaussian models and constitute the forecast step of Algorithm 1. This follows from the linearity of the expected value operator and the fact that linear transformations of Gaussian random variables are again Gaussian (see also Jazwinski (1970) and Gardiner (2004) for a rigorous derivation of this result from the differential form of the Chapman-Kolmogorov equation with Gaussian errors, called the Fokker-Planck equation).

The update step of Algorithm 1 follows directly from Bayes' Theorem because the product in the numerator of (3.8) can be algebraically calculated for Gaussian distributions. The result of this calculation is

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k) &= c_1 \exp\left(-\frac{1}{2}(\mathbf{x}_k - \mathbf{x}_k^f)^T(\mathbf{P}_k^f)^{-1}(\mathbf{x}_k - \mathbf{x}_k^f)\right) \\ &\quad \times \exp\left(-\frac{1}{2}(\mathbf{y}_k - \mathbf{H}\mathbf{x}_k)^T\mathbf{R}^{-1}(\mathbf{y}_k - \mathbf{H}\mathbf{x}_k)\right) \\ &= c_1 \exp\Big(-\frac{1}{2}\Big[(\mathbf{x}_k - \mathbf{x}_k^f)^T(\mathbf{P}_k^f)^{-1}(\mathbf{x}_k - \mathbf{x}_k^f) \\ &\qquad + (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k)^T\mathbf{R}^{-1}(\mathbf{y}_k - \mathbf{H}\mathbf{x}_k)\Big]\Big) \qquad (3.19) \\ &= c_2 \exp\left(-\frac{1}{2}(\mathbf{x}_k - \mathbf{x}_k^a)^T(\mathbf{P}_k^a)^{-1}(\mathbf{x}_k - \mathbf{x}_k^a)\right) \qquad (3.20) \end{aligned}$$

with

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}\left(\mathbf{y}_k - \mathbf{H}\mathbf{x}_k^f\right), \tag{3.21}$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}_k^f, \tag{3.22}$$

$$\mathbf{K} = \mathbf{P}_k^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_k^f\mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{3.23}$$

and with c_1 and c_2 being normalising constants (Cohn, 1997). The right hand side of (3.20) shows that the posterior pdf is again Gaussian with mean $\mathbf{x}_k^a$ and covariance matrix $\mathbf{P}_k^a$. The result $\mathbf{x}_k^a$ is also called the analysis and $\mathbf{P}_k^a$ is also called the analysis covariance matrix. The updated value of the predicted observation is

$$\mathbf{y}_k^a = \mathbf{y}_k^f + \mathbf{H}\mathbf{P}_k^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_k^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}_k - \mathbf{y}_k^f\right).$$

The KF consists of the recursive application of (3.17) – (3.18) and (3.21) – (3.23). Originally, these equations were derived by minimising the expected squared error of the estimate $\mathbf{x}_k^a$ for $\mathbf{x}_k$ given observations $\mathbf{y}_1, \ldots, \mathbf{y}_k$ (Kalman, 1960). As noted already by Kalman (1960), the minimising solution is the mean of the posterior pdf $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k)$ given in (3.21) and the minimised expected squared error is the trace of the posterior covariance matrix, $\mathrm{tr}(\mathbf{P}_k^a)$. Therefore, $\mathbf{x}_k^a$ is often called the minimum variance solution. Maximising the posterior probability $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k)$ also leads to $\mathbf{x}_k^a$ as given in (3.21) and $\mathbf{x}_k^a$ is then the most likely value, called the posterior mode or maximum likelihood solution (Jazwinski, 1970). The equivalence of the posterior mode and posterior mean also follows intuitively from the symmetric, unimodal shape of the Gaussian posterior pdf.

Relaxing the assumptions on the errors such that they have only zero mean and known, constant covariances but are not required to follow any specific distribution leads to an interpretation of the KF result as the solution of the weighted, linear least squares problem, where the updated state $\mathbf{x}_k^a$ is fit to a background state $\mathbf{x}_k^f$ and observations $\mathbf{y}_k$. Together with the forecast equations, this interpretation leads to the equivalence of the KF with the solution of the recursive, weighted, linear least squares problem,

$$\mathbf{x}_k^a = \min_x \left\|(\mathbf{P}_k^f)^{-\frac{1}{2}}\left(\mathbf{x}_k - \mathbf{x}_k^f\right)\right\|_2^2 + \left\|\mathbf{R}^{-\frac{1}{2}}\left(\mathbf{y}_k - \mathbf{H}\mathbf{x}_k\right)\right\|_2^2 \tag{3.24}$$

$$= \min_x \left\| \begin{array}{c} (\mathbf{P}_k^f)^{-\frac{1}{2}}\left(\mathbf{x}_k - \mathbf{x}_k^f\right) \\ \mathbf{R}^{-\frac{1}{2}}\left(\mathbf{y}_k - \mathbf{H}\mathbf{x}_k\right) \end{array} \right\|_2^2, \tag{3.25}$$

as can be seen from equation (3.19) (Duncan and Horn, 1972). For zero-mean errors, the Gauß-Markov Theorem states that the solution of this least squares problem, that is, the KF estimate $\mathbf{x}_k^a$, is still the best *linear* unbiased estimate (BLUE) of $\mathbf{x}_k$ while for Gaussian errors with zero mean, the KF was optimal among all possible estimators (Jazwinski, 1970; optimal in the sense of minimising the expected squared error). Last, we note that the forecast or background term $\left\|\left(\mathbf{P}_k^f\right)^{-\frac{1}{2}}\left(\mathbf{x}_k - \mathbf{x}_k^f\right)\right\|_2^2$ can be interpreted

as a Tikhonov regularisation term to the ill-posed problem of retrieving the state from the observations (Freitag and Potthast, 2013).

3.2.2 The ensemble Kalman filter

The size of the matrices in the KF forecast and update equations increases quadratically with the dimension of the state space. With current Earth system models' or numerical weather prediction models' state space dimensions of the order of 10^7 and above, the storage and the computational requirements of the KF would quickly exceed practical bounds (see for example Talagrand, 1997; Houtekamer et al., 2013). The more important limitation of the KF, however, is the restriction to linear models which allow the explicit evolution of the state's mean and covariance to solve the Chapman-Kolmogorov equation (3.10).

The EnKF uses Monte Carlo methods to overcome both limitations. Instead of using the mean and the covariance to describe the state vector distribution, the EnKF uses an ensemble of N states which represents a sample from the state vector distribution. To solve the Chapman-Kolmogorov equation for the evolution of state vector pdf, each ensemble member evolves independently according to the, possibly nonlinear, model equations. The resulting ensemble will then be a sample from the prior distribution at the next time step (Gordon et al., 1993; Kitagawa, 1996). Model error terms can be included in the evolution or can be accounted for in an intermediate step (section 3.3.1). The error in the estimates of the statistical moments of the involved distributions decreases proportional to $\frac{1}{\sqrt{N}}$ (Evensen, 1994; Doucet et al., 2001).

The update step of the EnKF uses the sample estimate $\hat{\mathbf{P}}_k^f$ of the true covariance $\mathbf{P}_k^f$ to construct a sample estimate $\hat{\mathbf{K}}$ of $\mathbf{K}$ and to update each ensemble member such that the mean of the updated ensemble $\hat{\mathbf{x}}_k^a$ and the covariance of the updated ensemble $\hat{\mathbf{P}}_k^a$ follow the KF equations (3.21) – (3.23),

$$\hat{\mathbf{x}}_k^a = \hat{\mathbf{x}}_k^f + \hat{\mathbf{K}}\left(\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_k^f\right), \tag{3.26}$$

$$\hat{\mathbf{P}}_k^a = \left(\mathbf{I} - \hat{\mathbf{K}}\mathbf{H}\right)\hat{\mathbf{P}}_k^f, \tag{3.27}$$

$$\hat{\mathbf{K}} = \hat{\mathbf{P}}_k^f\mathbf{H}^T\left(\mathbf{H}\hat{\mathbf{P}}_k^f\mathbf{H}^T + \mathbf{R}\right)^{-1}. \tag{3.28}$$

The ensemble update can be calculated in several ways. Replacing $\hat{\mathbf{x}}_k^a$ and $\hat{\mathbf{x}}_k^f$ above with $\mathbf{x}_k^{a,i}$ and $\mathbf{x}_k^{f,i}$, where $i = 1, \ldots, N$ indexes the ensemble members, leads to the perturbed observations EnKF. The KF equations are thus applied directly to each ensemble member with $\hat{\mathbf{K}}$ and $\hat{\mathbf{P}}_k^f$ estimated from the forecast ensemble. This method requires the use of perturbed observations

$$\mathbf{y}_k^i = \mathbf{y}_k + \boldsymbol{\varepsilon}^i, \quad i = 1, \ldots, N \tag{3.29}$$

in place of $\mathbf{y}_k$ to achieve the correct posterior covariance, where $\boldsymbol{\varepsilon}^i$ is sampled from the

observation error distribution (Burgers et al., 1998; Houtekamer and Mitchell, 1998).

Opposed to this stochastic version of the EnKF, deterministic versions such as the ensemble transform Kalman filter (ETKF; Bishop et al., 2001) and the ensemble adjustment Kalman filter (EAKF; Anderson, 2001) use matrix square roots of the analysis ensemble covariance matrix $\hat{\mathbf{P}}_k^a$ to derive the analysis ensemble (section 3.2.3). For linear models, linear observation operators, and Gaussian error distributions, the estimates of stochastic and deterministic EnKFs will converge to the KF estimates with increasing ensemble size. This follows by construction for the deterministic EnKF versions and was shown by Mandel et al. (2011) for the perturbed observations EnKF (Burgers et al., 1998 showed the same but made implicit assumptions that are not required for the proof by Mandel et al., 2011).

Using an ensemble to evolve and to update the state vector distribution lowers the computational and storage requirements because the state vector covariance matrices $\mathbf{P}_k^{a/f}$ do neither need to be computed nor stored. Their information is inherent in the ensemble and efficient implementations allow to update the ensemble without explicit use of these matrices (Evensen, 2003; Anderson and Collins, 2007; Houtekamer et al., 2013; Nerger and Hiller, 2013).

If the observation errors of individual observations $y_k^j,\ 1 \leq j \leq m$, or of different sets of observations $\{y_k^j,\ j \in I \subset \{1, \ldots, m\}\}$ are uncorrelated with each other, the required amount of computation and storage can be further reduced. In this case, single observations or sets of observations can be used one after another in the assimilation. This reduces the size of the matrix $\mathbf{H}\hat{\mathbf{P}}_k^f\mathbf{H}^T + \mathbf{R}$ which has to be inverted, but yields the same result as if they were assimilated all at once (Houtekamer and Mitchell, 2001).

3.2.3 Square root filters and the ensemble adjustment Kalman filter

The idea of square root filters originated from the "poor numerical properties" (Paige and Saunders, 1977) of the KF covariance update in (3.22). The numerical solution of this equation led to covariance matrices that were no longer positive-semidefinite, which is a theoretical requirement for every covariance matrix (Jazwinski, 1970). Instead of calculating $\mathbf{P}^a$ (we drop the time index k for this section), square root filters solve for a matrix $\mathbf{X}^a$ such that

$$\mathbf{P}^a = \mathbf{X}^a\mathbf{X}^{aT}, \tag{3.30}$$

where $\mathbf{X}^a$ is a so called matrix square root of $\mathbf{P}^a$ (Kaminski et al., 1971). The matrix product $\mathbf{X}^a\mathbf{X}^{aT}$ is then ensured to be positive semidefinite.

The fact that sample estimate of $\mathbf{P}^a$ is

$$\hat{\mathbf{P}}^a = \frac{1}{N-1}\left((\mathbf{x}^{a,1} - \mathbf{x}^a) \quad \ldots \quad (\mathbf{x}^{a,N} - \mathbf{x}^a)\right)\left((\mathbf{x}^{a,1} - \mathbf{x}^a) \quad \ldots \quad (\mathbf{x}^{a,N} - \mathbf{x}^a)\right)^T, \tag{3.31}$$

where $\mathbf{x}^a$ is the updated ensemble mean, shows that the matrix $\mathbf{X}^a$ in (3.30) is the matrix

of scaled analysis ensemble perturbations,

$$\mathbf{X}^a = \frac{1}{\sqrt{N-1}} \left((\mathbf{x}^{a,1} - \mathbf{x}^a) \quad \dots \quad (\mathbf{x}^{a,N} - \mathbf{x}^a) \right). \tag{3.32}$$

Consequently, updating the ensemble mean according to (3.26) and solving for the square root of

$$\begin{aligned} \mathbf{X}^a \mathbf{X}^{aT} \;=\; \hat{\mathbf{P}}^a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\, \hat{\mathbf{P}}^f \\ &= \hat{\mathbf{P}}^f - \hat{\mathbf{P}}^f \mathbf{H}^T \left(\mathbf{H} \hat{\mathbf{P}}^f \mathbf{H}^T + \mathbf{R} \right)^{-1} \mathbf{H} \hat{\mathbf{P}}^f \\ &= \mathbf{X}^f \mathbf{X}^{fT} - \mathbf{X}^f \mathbf{X}^{fT} \mathbf{H}^T \left(\mathbf{H} \mathbf{X}^f \mathbf{X}^{fT} \mathbf{H}^T + \mathbf{R} \right)^{-1} \mathbf{H} \mathbf{X}^f \mathbf{X}^{fT} \\ &= \mathbf{X}^f \left(\mathbf{I} - \mathbf{X}^{fT} \mathbf{H}^T \left(\mathbf{H} \mathbf{X}^f \mathbf{X}^{fT} \mathbf{H}^T + \mathbf{R} \right)^{-1} \mathbf{H} \mathbf{X}^f \right) \mathbf{X}^{fT} \end{aligned} \tag{3.33}$$

yields an updated ensemble (by adding the rescaled columns of $\mathbf{X}^a$ to $\mathbf{x}^a$) with the exact updated mean and covariance as given by the KF equations (Whitaker and Hamill, 2002; Tippett et al., 2003). In particular, this avoids sampling the observation error distribution as in the stochastic EnKF, which introduces additional sampling error into the ensemble (section 3.3.1). In this respect, ensemble square root filters are superior to the perturbed observation EnKF. The solution of (3.33) is not unique as explained by Tippett et al. (2003) and, for example, the ETKF and the EAKF are two square root filters that solve (3.33) differently.

Because we will later derive a new method related to non-Gaussian distributions that is motivated by the EAKF approach, we explain the EAKF here in detail. The reasoning behind the EAKF is to retain as much of the prior ensemble structure, that is, as much of the higher statistical moments of the ensemble, in the analysis as possible. To this end, the EAKF transforms the prior ensemble into a coordinate system in which the covariance matrix of the prior ensemble $\mathbf{P}^f$ and the scaled inverse observational covariance matrix $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ become diagonal matrices. Further, the transformed state covariance matrix is scaled such that all diagonal elements are one and the same scaling is applied to the transformed inverse observational covariance matrix. The EAKF also calculates the updated mean in this transformed space. Finally, because all covariance matrices are diagonal, the updated ensemble can be derived by shifting the transformed ensemble to the updated mean and contracting it along the transformed coordinate axes according to the transformed inverse observational covariances.The posterior ensemble in the original state space is then obtained by applying the respective inverse transformations to the updated ensemble in the transformed coordinates (Anderson, 2001, 2009a).

Using single, sequential observations y, the EAKF can be understood more directly within the local least squares framework of Anderson (2003). The EAKF can then be summarised in three steps:

1. evolve the ensemble and predict an ensemble of observations,

2. update the ensemble of observations,

3. transfer the observational update to a state update using linear regression.

The EAKF applies the observation operator $\mathbf{H}$, which now maps $\mathbf{x}$ to a scalar observation y, to each member of the prior ensemble to generate a prior ensemble of observations $y^{f,i} = \mathbf{H}\mathbf{x}^{f,i}$. Using the ensemble mean y^f and the ensemble covariance $(\sigma^f)^2$ of the prior observation ensemble as well as the observed value y and its prescribed observation error covariance $(\sigma^o)^2$, the updated observation ensemble covariance is

$$\begin{aligned}(\sigma^a)^2 &= \frac{(\sigma^f)^2\,(\sigma^o)^2}{(\sigma^f)^2 + (\sigma^o)^2} \\ &= \frac{1}{\frac{1}{(\sigma^f)^2} + \frac{1}{(\sigma^o)^2}}.\end{aligned} \tag{3.34}$$

And the updated observation ensemble mean is

$$y^a = \frac{(\sigma^o)^2\,y^f + (\sigma^f)^2\,y}{(\sigma^f)^2 + (\sigma^o)^2} \tag{3.35}$$

$$= (\sigma^a)^2 \left(\frac{y^f}{(\sigma^f)^2} + \frac{y}{(\sigma^o)^2}\right). \tag{3.36}$$

They result from the product in (3.19) (for Gaussian distributions) or from the Gauß-Markov Theorem (for zero-mean errors). The observation ensemble is then shifted and scaled such that the updated ensemble has mean y^a and covariance $(\sigma^a)^2$. The differences between the prior observation ensemble and the updated observation ensemble define the observation increments $\Delta y^i = y^{a,i} - y^{f,i}$, $i = 1, \ldots, N$. These increments are scaled to state increments for the j-th element of the state vector, $j = 1, \ldots, n$, according to

$$\Delta x^{a,i,j} = \frac{\sigma_{x^j y}}{(\sigma^f)^2}\Delta y^i, \tag{3.37}$$

where $\sigma_{x^j y}$ is the covariance of the j-th element of the state vector and the observation as estimated from the prior ensemble of states and predicted observations. The term $\frac{\sigma_{x^j y}}{(\sigma^f)^2}$ corresponds to the estimated slope of a linear regression line fitted between the prior state and observation ensembles.

This approach is easily extended to nonlinear observation operators by replacing $\mathbf{H}$ with $h(\mathbf{x})$ and following the same steps. Given that $\mathbf{H}$ is linear, the points $(\mathbf{x}^{f,i}, y^i) = (\mathbf{x}^{f,i}, \mathbf{H}\mathbf{x}^{f,i})$ all lie on a straight line defined by the observation operator. If $h(\mathbf{x})$ is nonlinear, the line defined by the regression slope $\frac{\sigma_{x^j y}}{(\sigma^f)^2}$ will be the best linear fit to the nonlinear observation operator estimated from the state-observation pairs of the prior ensemble. Because this fit changes with the location and the spread of the prior ensemble and is not equivalent to a global least squares fit, Anderson (2003) calls it a "local least squares fit".

3.2.4 State augmentation for nonlinear observations and parameter estimation

The handling of nonlinear observations described above can be generalised for all EnKF types by augmenting the state vector with the results of the nonlinear observation operator $h(\mathbf{x})$ (Evensen, 2003). The vector of predicted observations $h(\mathbf{x}^{f,i})$ is appended to the state vector $\mathbf{x}^{f,i}$ for all ensemble members. Writing

$$\mathbf{y}_k = h(\mathbf{x}_k) + \boldsymbol{\varepsilon}_k \tag{3.38}$$

$$= \begin{pmatrix} \mathbf{0}_n & \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \mathbf{x}_k \\ h(\mathbf{x}_k) \end{pmatrix} + \boldsymbol{\varepsilon}_k \tag{3.39}$$

$$= \mathbf{H} \begin{pmatrix} \mathbf{x}_k \\ h(\mathbf{x}_k) \end{pmatrix} + \boldsymbol{\varepsilon}_k \tag{3.40}$$

shows that the now linear observation operator is given by $\mathbf{H} = \begin{pmatrix} \mathbf{0}_n & \mathbf{I}_m \end{pmatrix}$, where $\mathbf{0}_n$ and $\mathbf{I}_m$ are zero and identity matrices of dimension $n \times n$ and $m \times m$, respectively.

State augmentation also enables the estimation of parameters with the EnKF. For this purpose, parameters are treated like state variables and appended to the state vector. This approach has been suggested to estimate correlation and bias parameters for the error terms (Jazwinski, 1970; Dee and Da Silva, 1998) but is easily transferred to model parameters (Moradkhani et al., 2005; Evensen, 2009b). Besides the combined state-parameter estimation, where the complete updated state-parameter vector is used for the next forecast step, a pure parameter estimation approach is also possible. This method uses only the updated parameters to replace the ones from the forecast. The next forecast cycle is then started with the updated parameters but with the unchanged state from the previous forecast (Nowak, 2009; Schöniger et al., 2012).

3.3 Sources of error in the ensemble Kalman filter

The KF is a statistical algorithm that is driven by assumptions about the uncertainty of the initial state, about the evolution of this initial uncertainty, and about the uncertainty of the observations. These assumptions are cast into a statistical model as described in section 3.1. A flawed or incomplete specification of this statistical model will lead to errors in the estimates obtained by the KF. And the ensemble representation of the state's probability distribution in the EnKF will incur additional errors due to the finite sample size. Any combination of these types of errors can lead to what is called filter divergence.

The filter diverges if the state vector estimate follows an incorrect trajectory with ever decreasing estimated covariance, that is, with ever increasing certainty in the – wrong – estimate. From the covariance update in (3.22), we see that the state vector covariance decreases in every update step because $\mathbf{P}_k^f$ and $\mathbf{KHP}_k^f$ on the right-hand side and $\mathbf{P}_k^a$ on the left hand side of (3.22) are all covariance matrices and therefore positive semidefinite.

Consequently, $\mathrm{tr}(\mathbf{P}_k^a)$ must be smaller than $\mathrm{tr}(\mathbf{P}_k^f)$. But a continuously decreasing covariance of the state vector makes the assimilation of additional observations increasingly irrelevant because the observations will not be given any influential weight anymore and the filter will not move away from its locked-in trajectory.

3.3.1 Forecast model error and sampling error

Consider the evolution of the state vector covariance matrix given by

$$\mathbf{P}_k^f = \mathbf{M}\mathbf{P}_{k-1}^a\mathbf{M}^T + \mathbf{Q}.$$

The model error covariance matrix $\mathbf{Q}$ contributes to the error covariance of the forecast. If $\mathbf{Q}$ is neglected or chosen too small, the estimate of $\mathbf{P}_k^f$ will be too small. Moreover, the EnKF systematically underestimates the analysis covariance matrix $\mathbf{P}_k^a$ which further reduces $\mathbf{P}_k^f$ in the next forecast ("inbreeding"; Houtekamer and Mitchell, 1998; van Leeuwen, 1999; Sacher and Bartello, 2008). In the update step, a too small estimate of $\mathbf{P}_k^f$ leads to an erroneously high weight for the predicted state (given by its inverse covariance matrix, cf. (3.24)) compared to the weight given to the observations and, eventually, to a loss of impact of the observations on the state vector estimate and thus to filter divergence.

The representation of the forecast distribution by an ensemble introduces additional errors in the update step of the EnKF because it uses sample estimates of the covariance matrices instead of the exact values. This causes spurious correlations where, due to the limited ensemble size, the estimated value of entries in $\mathbf{P}_k^f$ is not zero although the true value is zero. These estimation errors lead to errors in the updated state and in the updated covariance because observations and states will be erroneously linked to each other by the spurious correlations.

Both effects are well known error sources in the EnKF and different techniques have been developed to handle them (Anderson, 2012; Whitaker and Hamill, 2012). To ensure a sufficient spread of the forecast ensemble, various model error representations are currently used. Multiplicative inflation multiplies the forecast covariances or the updated covariances by an inflation factor. Implemented into an ensemble filter, this corresponds to scaling the ensemble perturbations to increase the sample covariance (Anderson and Anderson, 1999). Additive inflation follows directly from the evolution of the covariance matrix. This method adds random perturbations sampled from the model error distribution to the ensemble members (Mitchell and Houtekamer, 2000).

Spurious correlations are reduced by localisation of the covariances, that is, by constraining non-zero covariances to the physical vicinity of a state variable and by tapering the covariances with increasing distance between states (Hamill et al., 2001; Houtekamer and Mitchell, 2001).

3.3.2 State-dependent and non-zero mean errors

The derivation of both, the recursive Bayesian estimation and the recursive best linear unbiased estimation, assume a prior state vector distribution with mean $\mathbf{x}_k^f$ and zero-mean observation errors (section 3.2.1) with a constant observation error covariance matrix $\mathbf{R}$. The zero-mean observation error assumption together with the assumption of a constant observation error covariance means that the observation error distribution is independent of the observed state. As a consequence, the likelihood of $\mathbf{x}_k$ given $\mathbf{y}_k$,

$$p(\mathbf{y}_k|\mathbf{x}_k) = p_{\boldsymbol{\varepsilon}}(\mathbf{y}_k - h(\mathbf{x}_k)),$$

will have the same shape as the observation error distribution $p_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon})$ for all observations $\mathbf{y}_k$ and will only be shifted along the $\mathbf{x}_k$-coordinate axes. For a Gaussian observation error distribution, for example, the likelihood will then also be a Gaussian function. If, however, $\mathbf{R}$ is not constant but a function of the unknown state $\mathbf{x}_k$, that is, $\mathbf{R} = \mathbf{R}(\mathbf{x}_k)$, the likelihood will in general be non-Gaussian even if all observation error distributions are Gaussian. This is because to construct the likelihood, a different pdf $p_{\boldsymbol{\varepsilon}_k}(\mathbf{y}_k - h(\mathbf{x}_k))$ has to be evaluated for every state $\mathbf{x}_k$. Further, the shape of the likelihood may be different for every observation $\mathbf{y}_k$. Given that the prior state vector estimate and the observations have zero mean errors, Zehnwirth (1988) extended the KF to accommodate a state-dependent observation error covariance matrix. The equation for the Kalman gain then changes to

$$\mathbf{K} = \mathbf{P}_k^f \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_k^f \mathbf{H}^T + E\big(\mathbf{R}(\mathbf{x})\big) \right)^{-1}. \tag{3.41}$$

A modification of the EnKF to accommodate $E\big(\mathbf{R}(\mathbf{x})\big)$ is discussed in section 3.6.3. The estimates of the modified KF are no longer the conditional mean and the conditional covariance of $\mathbf{x}_k$ given $\mathbf{y}_1, \ldots, \mathbf{y}_k$ because, due to the non-Gaussian likelihood, the right hand side of

$$p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k) = \frac{p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})\, p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})}$$

will no longer be a Gaussian pdf. The only remaining interpretation of the KF state vector and covariance matrix estimates in this case are the BLUE and its estimated error covariance.

Non-zero mean errors, that is biases, can be included in the KF framework using state augmentation and can in principle be estimated online, given prior estimates of the biases (Dee and Da Silva, 1998). It is, however, assumed that either the observations or the prior state vector estimate are unbiased which is an inappropriate assumption for certain non-Gaussian observation error distributions (section 3.7). If non-zero mean errors, possibly also with a state-dependent mean, are neglected, a statistical interpretation of the KF and the EnKF results becomes difficult. Such an interpretation will be given in section 3.6.

3.4 Physical consistency of updated states

The nature of the KF's update step causes a blindness for physical constraints and nonlinearities. The only link between the state vector and the observations in the KF update step is the cross-covariance matrix $\text{cov}(\mathbf{x}, \mathbf{y})$ between states and observations. For linear observation operators, $\text{cov}(\mathbf{x}, \mathbf{y})$ (we drop the time index for this section) is given by

$$\begin{aligned}
\text{cov}(\mathbf{x}, \mathbf{y}) &= E\left((\mathbf{x} - E(\mathbf{x}))\ (\mathbf{y} - E(\mathbf{y}))^T\right) \\
&= E\left((\mathbf{x} - E(\mathbf{x}))\ (\mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon} - E(\mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}))^T\right) \\
&= E\left((\mathbf{x} - \mathbf{x}^f)\ (\mathbf{x} - \mathbf{x}^f)^T\mathbf{H}^T\right) + E\left((\mathbf{x} - \mathbf{x}^f)\boldsymbol{\varepsilon}^T\right) \\
&= E\left((\mathbf{x} - \mathbf{x}^f)\ (\mathbf{x} - \mathbf{x}^f)^T\mathbf{H}^T\right) \\
&= \mathbf{P}^f\mathbf{H}^T,
\end{aligned}$$

which appears in the definition of the Kalman gain $\mathbf{K}$. This term transfers the update from observation space into state space. To see this, consider the right-hand side of

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_k^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right),$$

where $\left(\mathbf{H}\mathbf{P}_k^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}_k - \mathbf{H}\mathbf{x}_k^f\right)$ is a vector in $\mathbb{R}^m$ that corresponds to the observation increments scaled with the inverse prior covariance of $\mathbf{y}$ (cf. (3.45)). The term $\mathbf{P}^f\mathbf{H}^T$ maps these analysis increments in observation space to analysis increments in state space and corresponds to the cross-covariance matrix of $\mathbf{x}$ and $\mathbf{y}$.

For nonlinear observation operators, $\text{cov}(\mathbf{x}, \mathbf{y})$ and the scaling factor for the observation increments are estimated from the ensemble of augmented state vectors and used for the update of $\mathbf{x}^f$ (section 3.2.4). Consequently, the KF assumes a linear relationship between states and observations that is a statistical, linear approximation of the nonlinear observation operator around the ensemble mean (section 3.6.2). This linear approximation is not limited to any bounded domain because the linear relationship given by $\left(\text{cov}(\mathbf{x}, \mathbf{y})\ \text{cov}(\mathbf{y}, \mathbf{y})^{-1}\right)$ can be arbitrarily applied to any $\mathbf{x}$ and $\mathbf{y}$. Due to observation errors, sampling errors of the ensemble, and nonlinear observation operators, the mapping of an observation $\mathbf{y}$ to an updated state $\mathbf{x}^a$ will lead to physically invalid results for $\mathbf{x}^a$ if $\mathbf{y}$ falls outside a certain range $[\mathbf{y}_{min}, \mathbf{y}_{max}]$ (Figure 3.2).

Various approaches have been developed to constrain updated states (and parameters) to physically valid ranges. These approaches can be broadly categorised into variable transformation techniques (Bertino et al., 2002; Nielsen-Gammon et al., 2010; Schirber et al., 2013) and constrained optimisation approaches. The latter solve the minimisation problem (3.24) under appropriate constraints on the solution $\mathbf{x}_k^a$ (Pan and Wood, 2006; Janjić et al., 2014) or add penalty terms to the objective function (Yilmaz et al., 2011; this corresponds to the weak-constraint four-dimensional variational formulation by Gauthier and Thépaut, 2001). Because constrained optimisation changes the problem formulation

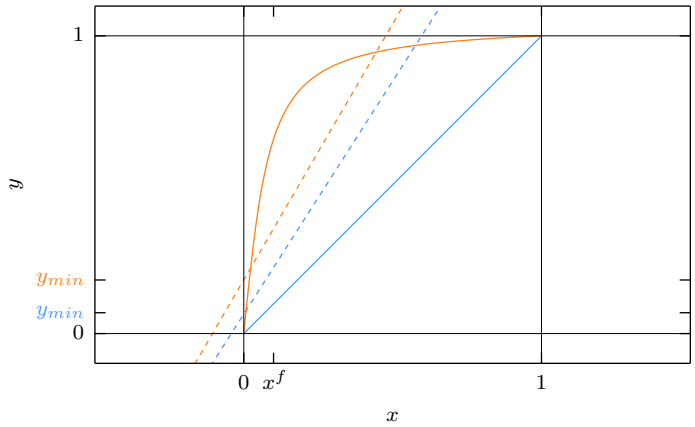


Figure 3.2: Linear (blue) and nonlinear (orange) observation operators (solid lines) on a bounded interval and the linear relation given by $\left(\mathrm{cov}(\mathbf{x}, \mathbf{y})\ \mathrm{cov}(\mathbf{y}, \mathbf{y})^{-1}\right)$ (dashed lines) that is used to update $\mathbf{x}^f$ from an observation $\mathbf{y}$. If the physically valid range for x is [0, 1], only observations larger than y_{min} will yield physically consistent values in the KF update. Prior and observation error covariances in this example are 0.5 and 0.25, respectively.

that was used to derive the recursive solution of the filtering problem, these approaches prohibit an immediate statistical interpretation of the estimates in terms of conditional distributions or best linear unbiased estimates. Variable transformation techniques, in contrast, retain the Bayesian character of the EnKF. These are discussed in section 3.7.

3.5 Nonlinearity and non-Gaussianity

Nonlinearity and non-Gaussianity are closely linked to each other. The reason is that any nonlinear transformation of a Gaussian random variable, for example a model forecast or an observation operator, will generally transform the variables distribution from a Gaussian distribution to a non-Gaussian distribution. And while multivariate Gaussian distributions are fully described by their means and covariances – which correspond to linear relationships –, a multivariate non-Gaussian distribution has higher non-zero moments and requires nonlinear functions to characterise the relationships between individual random variables.

The development of the EnKF was motivated by computational and memory limitations when handling the KF covariance matrices, on the one hand, and by the limited applicability of the KF to nonlinear forecast models, on the other hand (Evensen, 1994). In fact, the forecast step of the EnKF poses no constraints on the linearity of the model, given that the ensemble is large enough (section 3.2.2; for a discussion of strongly nonlinear models in conjunction with the EnKF see Sakov et al., 2012). The update step, however, is still based on covariances and thus on linear relationships. Therefore, the update step also requires Gaussian distributions in order to be a Bayesian method that yields the correct posterior pdf.

The KF was introduced in section 3.2.1 as a special case of recursive Bayesian estimation (Algorithm 1 in section 3.1) for linear models and Gaussian error distributions. In this case, the update step of the KF yields the mean and the covariance of the conditional pdf $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_k)$ which is also Gaussian and fully described by the KF estimates of its mean and covariance. We also noted that the KF reduces to the BLUE and its error covariance estimate in case of non-Gaussian distributions with zero mean. The BLUE is sub-optimal with respect to the expected squared estimation error. And more important, the BLUE is hard to interpret because it does not allow to draw any conclusions on the probability of the estimated state to be the true state. After all, the BLUE could lie in a low probability region of a multimodal or long-tailed pdf and could be very unlikely to be the true state (note that being unbiased here refers to the expectation taken over $\mathbf{x}$ only, without any considerations on the available observations).

The EnKF does not alleviate this issue because the EnKF is only a Monte Carlo approximation of the KF and as such only a Monte Carlo approximation of the BLUE in the non-Gaussian case. Starting from the same initial ensembles, different versions of the EnKF (for example the perturbed observations EnKF and the EAKF) will lead to different updated ensembles that only agree in their ensemble means and ensemble covariance matrices. Only in the Gaussian case will the two ensembles be samples from the same, Gaussian posterior distribution. Comparisons of stochastic and deterministic filters under non-Gaussianity show that stochastic filters like the perturbed observations EnKF are more resilient to outliers in the ensemble. This means that the spread is actually generated by randomly differing states instead of being generated by only one member far away from a nearly collapsed ensemble or by two nearly collapsed groups of ensemble members (Lawson and Hansen, 2004; Lei et al., 2010).

Nonlinear observation operators or non-Gaussian observation errors also invalidate the Bayesian interpretation of the KF and the EnKF. This is because the conditional mean of the state $\mathbf{x}$ given the observations $\mathbf{y} = h(\mathbf{x}) + \boldsymbol{\varepsilon}$ is, in general, a nonlinear function of the observations $\mathbf{y}$. But the KF update is linear in $\mathbf{y}$ and can consequently only be an approximation of the conditional mean. The same holds for the conditional covariance matrix. The BLUE interpretation of the EnKF, however, also holds for nonlinear observation operators and non-Gaussian observation errors, provided that they have zero mean (section 3.6).

3.6 The Kalman filter as linear regression

The connection between the KF and linear regression was noted by Duncan and Horn (1972) and the KF has been described as the "evolution of a series of regression functions" by Meinhold and Singpurwalla (1983). In this section, we explain how the effects of nonlinearity, non-Gaussianity, and state-dependent, non-zero mean errors in the KF and the EnKF can be understood using the linear regression framework.

Consider the joint pdf $p(\mathbf{x}, \mathbf{y})$. Then the conditional mean of $\mathbf{x}$ given $\mathbf{y}$ is a function of

$\mathbf{y}$ given by

$$E(\mathbf{x}|\mathbf{y}) = \int_{\mathbf{x}} \mathbf{x}\, p(\mathbf{x}|\mathbf{y})\, \mathrm{d}\mathbf{x} \\ = f(\mathbf{y}) \tag{3.42}$$

The function $f : \mathbb{R}^m \to \mathbb{R}^n$ is called the regression of $\mathbf{x}$ on $\mathbf{y}$ (Papoulis, 1991). As noted in section 3.2.1, the conditional mean $f(\mathbf{y})$ is the globally optimal estimate of $\mathbf{x}$ because it minimises the expected squared error

$$E((\mathbf{x} - g(\mathbf{y}))^T(\mathbf{x} - g(\mathbf{y}))$$

over all possible estimators $g(\mathbf{y})$.

In general, the conditional mean $f(\mathbf{y})$ is a nonlinear function of $\mathbf{y}$ that is not attainable because it requires the knowledge of the conditional pdf $p(\mathbf{x}|\mathbf{y})$ and the solution of the multidimensional integral in (3.42) (also note, that $p(\mathbf{x}|\mathbf{y})$ would be the solution of the filtering problem which we are trying to find). Instead, we seek the best linear approximation of $f(\mathbf{y})$, that means we seek a linear function

$$\hat{f}(\mathbf{y}) = \mathbf{A}\mathbf{y} + \mathbf{b}$$

that minimises the expected squared error

$$\mathrm{tr}\left[E\left((\mathbf{x} - \hat{f}(\mathbf{y}))^T\ (\mathbf{x} - \hat{f}(\mathbf{y}))\right)\right].$$

This is called linear regression and $\mathbf{A}$ and $\mathbf{b}$ are given by (Pfeiffer, 1990; chapter 16)

$$\mathbf{A} = \mathrm{cov}(\mathbf{x},\, \mathbf{y})\ \mathrm{cov}(\mathbf{y},\, \mathbf{y})^{-1}, \\ \mathbf{b} = E\left(\mathbf{x}\right) - \mathrm{cov}(\mathbf{x},\, \mathbf{y})\ \mathrm{cov}(\mathbf{y},\, \mathbf{y})^{-1} E\left(\mathbf{y}\right).$$

The linear regression estimate is unbiased,

$$E(\hat{f}(\mathbf{y})) = E\left((E(\mathbf{x}) - \mathrm{cov}(\mathbf{x},\, \mathbf{y})\ \mathrm{cov}(\mathbf{y},\, \mathbf{y})^{-1}(\mathbf{y} - E(\mathbf{y}))\right) \\ = E(\mathbf{x}),$$

and the error covariance matrix of the linear regression estimate is

$$E\Big((\mathbf{x} - \hat{f}(\mathbf{y}))(\mathbf{x} - \hat{f}(\mathbf{y}))^T\Big) \tag{3.43}$$

where the expectation is taken over $\mathbf{x}$ and $\mathbf{y}$.

Using now the KF assumptions that $E(\mathbf{x}) = \mathbf{x}^f$, that $\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}$ with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, and

that $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ are independent, we get

$$\begin{aligned} E(\mathbf{y}) &= E(\mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}) \\ &= \mathbf{H}E(\mathbf{x}) + E(\boldsymbol{\varepsilon}) \\ &= \mathbf{H}\mathbf{x}^f, \end{aligned} \tag{3.44}$$

and

$$\begin{aligned} \operatorname{cov}(\mathbf{y}, \mathbf{y}) &= E\left((\mathbf{y} - E(\mathbf{y}))\,(\mathbf{y} - E(\mathbf{y}))^T\right) \\ &= E\left((\mathbf{y} - \mathbf{H}\mathbf{x}^f)\,(\mathbf{y} - \mathbf{H}\mathbf{x}^f)^T\right) \\ &= E\left((\mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon} - \mathbf{H}\mathbf{x}^f)\,(\mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon} - \mathbf{H}\mathbf{x}^f)^T\right) \\ &= \mathbf{H}E\left((\mathbf{x} - \mathbf{x}^f)\,(\mathbf{x} - \mathbf{x}^f)^T\right)\mathbf{H}^T + E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right) \\ &= \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}, \end{aligned} \tag{3.45}$$

as well as

$$\begin{aligned} \operatorname{cov}(\mathbf{x}, \mathbf{y}) &= E\left((\mathbf{x} - E(\mathbf{x}))\,(\mathbf{y} - E(\mathbf{y}))^T\right) \\ &= E\left((\mathbf{x} - \mathbf{x}^f)\,(\mathbf{y} - \mathbf{H}\mathbf{x}^f)^T\right) \\ &= E\left((\mathbf{x} - \mathbf{x}^f)\,(\mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon} - \mathbf{H}\mathbf{x}^f)^T\right) \\ &= E\left((\mathbf{x} - \mathbf{x}^f)\,(\mathbf{x} - \mathbf{x}^f)^T\mathbf{H}^T\right) + E\left((\mathbf{x} - \mathbf{x}^f)\boldsymbol{\varepsilon}^T\right) \\ &= E\left((\mathbf{x} - \mathbf{x}^f)\,(\mathbf{x} - \mathbf{x}^f)^T\right)\mathbf{H}^T \\ &= \mathbf{P}^f\mathbf{H}^T. \end{aligned}$$

The linear regression estimate of $\mathbf{x}$ given $\mathbf{y}$ now reads

$$\begin{aligned} \hat{f}(\mathbf{y}) &= \mathbf{x}^f + \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \\ &= \mathbf{x}^a, \end{aligned}$$

which is the KF estimate of $\mathbf{x}$ given $\mathbf{y}$. Likewise, the estimate of the error covariance of $\hat{f}(\mathbf{y})$ is equal to the KF estimate $\mathbf{P}^a$. This shows that the KF performs a linear regression of the state $\mathbf{x}$ on the observation $\mathbf{y}$.

In the special case of Gaussian distributions for $\mathbf{x}$, $\boldsymbol{\varepsilon}$, and thus also for $\mathbf{y}$, the joint pdf $p(\mathbf{x}, \mathbf{y})$ is Gaussian. We can write the joint pdf of $\mathbf{x}$ and $\mathbf{y}$ as

$$p(\mathbf{x}, \mathbf{y}) = c_1 \exp\left(-\frac{1}{2}\begin{pmatrix}\mathbf{x} - \mathbf{x}^f \\ \mathbf{y} - \mathbf{H}\mathbf{x}^f\end{pmatrix}^T \begin{pmatrix}\mathbf{P}^f & \mathbf{P}^f\mathbf{H}^T \\ \mathbf{H}^T\mathbf{P}^f & \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\end{pmatrix}^{-1} \begin{pmatrix}\mathbf{x} - \mathbf{x}^f \\ \mathbf{y} - \mathbf{H}\mathbf{x}^f\end{pmatrix}\right) \tag{3.46}$$

and then the conditional pdf of $\mathbf{x}$ given $\mathbf{y}$ is also Gaussian and given by

$$p(\mathbf{x}|\mathbf{y}) = c_2 \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}^a)^T(\mathbf{P}^a)^{-1}(\mathbf{x}-\mathbf{x}^a)\right).$$

The linear approximation $\hat{f}(\mathbf{y})$ to the conditional mean yields the conditional mean itself in this case, that is $\hat{f}(\mathbf{y}) = f(\mathbf{y})$, because for joint Gaussian pdfs, the conditional mean $f(\mathbf{y})$ is only a linear function of $\mathbf{y}$. Further, the error covariance matrix of the linear regression estimate in (3.43) coincides with the conditional covariance matrix,

$$\begin{aligned}
\mathrm{cov}(\mathbf{x}|\mathbf{y},\, \mathbf{x}|\mathbf{y}) &= E\left((\mathbf{x}-f(\mathbf{y}))\,(\mathbf{x}-f(\mathbf{y}))^T|\mathbf{y}\right) \\
&= E\left((\mathbf{x}-f(\mathbf{y}))\,(\mathbf{x}-f(\mathbf{y}))^T\right) \\
&= \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\mathbf{P}^f \\
&= \mathbf{P}^a,
\end{aligned}$$

because for joint Gaussian pdfs the conditional covariance is independent of $\mathbf{y}$ (Jazwinski, 1970). Therefore, if the involved distributions are Gaussian and the observation operator is linear with zero-mean errors, the linear regression estimate and its error covariance are the conditional mean and the conditional covariance matrix of $\mathbf{x}$ given $\mathbf{y}$.

Instead of a single linear regression, the EnKF performs N linear regressions

$$\mathbf{x}^{a,\,i} = \mathbf{x}^{f,\,i} + \hat{\mathbf{P}}^f\mathbf{H}^T(\mathbf{H}\hat{\mathbf{P}}^f\mathbf{H}^T+\mathbf{R})^{-1}(\mathbf{y}^i - \mathbf{H}\mathbf{x}^{f,\,i}), \quad i = 1,\,\ldots,\,N,$$

where $\hat{\mathbf{P}}\mathbf{H}^T$ and $(\mathbf{H}\hat{\mathbf{P}}^f\mathbf{H}^T+\mathbf{R})$ are sample estimates of $\mathrm{cov}(\mathbf{x},\,\mathbf{y})$ and $\mathrm{cov}(\mathbf{y},\,\mathbf{y})$ from the forecast ensemble. Burgers et al. (1998) showed that the use of perturbed observations $\mathbf{y}^i$ as defined in (3.29) is necessary to yield an updated ensemble whose sample covariance matrix is an estimate of the linear regression error covariance matrix and thus of the conditional covariance matrix in the Gaussian case. Ensemble square root filters such as the EAKF implicitly construct perturbed observations such that the updated ensemble covariance matrix matches the linear regression error covariance matrix (section 3.2.3). For increasing ensemble size, the Monte Carlo estimates of the linear regression will converge to the true linear regression of $\mathbf{x}$ on $\mathbf{y}$.

We now argue, that the KF estimates $\mathbf{x}^a$ and $\mathbf{P}^a$ can be understood as approximations of the conditional mean and the conditional covariance matrix also for non-Gaussian distributions and nonlinear observation operators. Further, we regard the normal distribution with mean $\mathbf{x}^a$ and $\mathbf{P}^a$ as an approximation of the conditional distribution of $p(\mathbf{x}|\mathbf{y})$. By reverting the arguments that led to the equivalence of the KF estimates with the conditional mean and the conditional covariance in the linear, Gaussian case, we find that this normal approximation of $p(\mathbf{x}|\mathbf{y})$ corresponds to an approximation of the joint pdf $p(\mathbf{x},\,\mathbf{y})$ with a normal joint pdf with mean $\begin{pmatrix}\mathbf{x}^f \\ \mathbf{H}\mathbf{x}^f\end{pmatrix}$ and covariance matrix $\begin{pmatrix}\mathbf{P}^f & \mathbf{P}^f\mathbf{H}^T \\ \mathbf{H}\mathbf{P}^f & \mathbf{H}\mathbf{P}^f\mathbf{H}^T+\mathbf{R}\end{pmatrix}$. This normal approximation of the true joint pdf corresponds to an exclusion of information about the higher moments of $\mathbf{x}$ and $\mathbf{y}$ as it is done in the KF. In this way, we can derive

the implicit approximations of observation error pdfs and likelihoods that we make when we apply the KF to non-Gaussian, nonlinear problems. The same holds asymptotically for the EnKF with $N \to \infty$ and we refer to KF and EnKF interchangeably in the next sections. In the sense of approximating the Bayesian Algorithm 1, the KF becomes

Algorithm 2

1. **initialise** the forecast distribution $p(\mathbf{x}_0)$,
2. **for** i **from** 1 **to** k**:**
   a) **forecast** the prior pdf $p(\mathbf{x}_i|\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$,
   b) **estimate** the conditional mean and covariance using linear regression,
   c) **approximate** the conditional pdf $p(\mathbf{x}_i|\mathbf{y}_1, \ldots, \mathbf{y}_i)$ with a normal pdf with the estimated conditional mean and covariance.

3.6.1 The Gaussian case with linear observations

For Gaussian distributions and a linear observation operator, the joint pdf is Gaussian and there are no approximations. Figure 3.3 visualises the KF estimation of the conditional mean in this case. The KF approximations of the observation error pdf, the likelihood, and the posterior pdf agree with the true pdfs and likelihoods. Further, the linear regression agrees with the conditional mean. Thus, all four regression curves in panels a) and e), that is the conditional mean of $\mathbf{x}$ given $\mathbf{y}$ (labelled "mean of $p(x|y)$"), the true linear regression of $\mathbf{x}$ on $\mathbf{y}$ and the two KF approximations of the linear regression, are identical. The constant observation error covariance, visualised by the grey, filled contours in panels e) and f), is the reason for the congruence of the two KF approximations of the linear regression and the congruence of the approximating joint normal pdfs. Having a observation error covariance means that the normal approximation of $p(\mathbf{y})$, which is the prior pdf of $\mathbf{y}$, is independent of the observation that is used to construct it. This independence is a natural requirement because $p(\mathbf{y})$ does not depend on realisations of $\mathbf{y}$.

3.6.2 The Gaussian case with nonlinear observations

For nonlinear observation operators, the linear regression reads

$$\mathbf{x}^a = \mathbf{x}^f + \mathrm{cov}(\mathbf{x}, \mathbf{y})\, \mathrm{cov}(\mathbf{y}, \mathbf{y})^{-1}(\mathbf{y} - E(h(\mathbf{x}))). \tag{3.47}$$

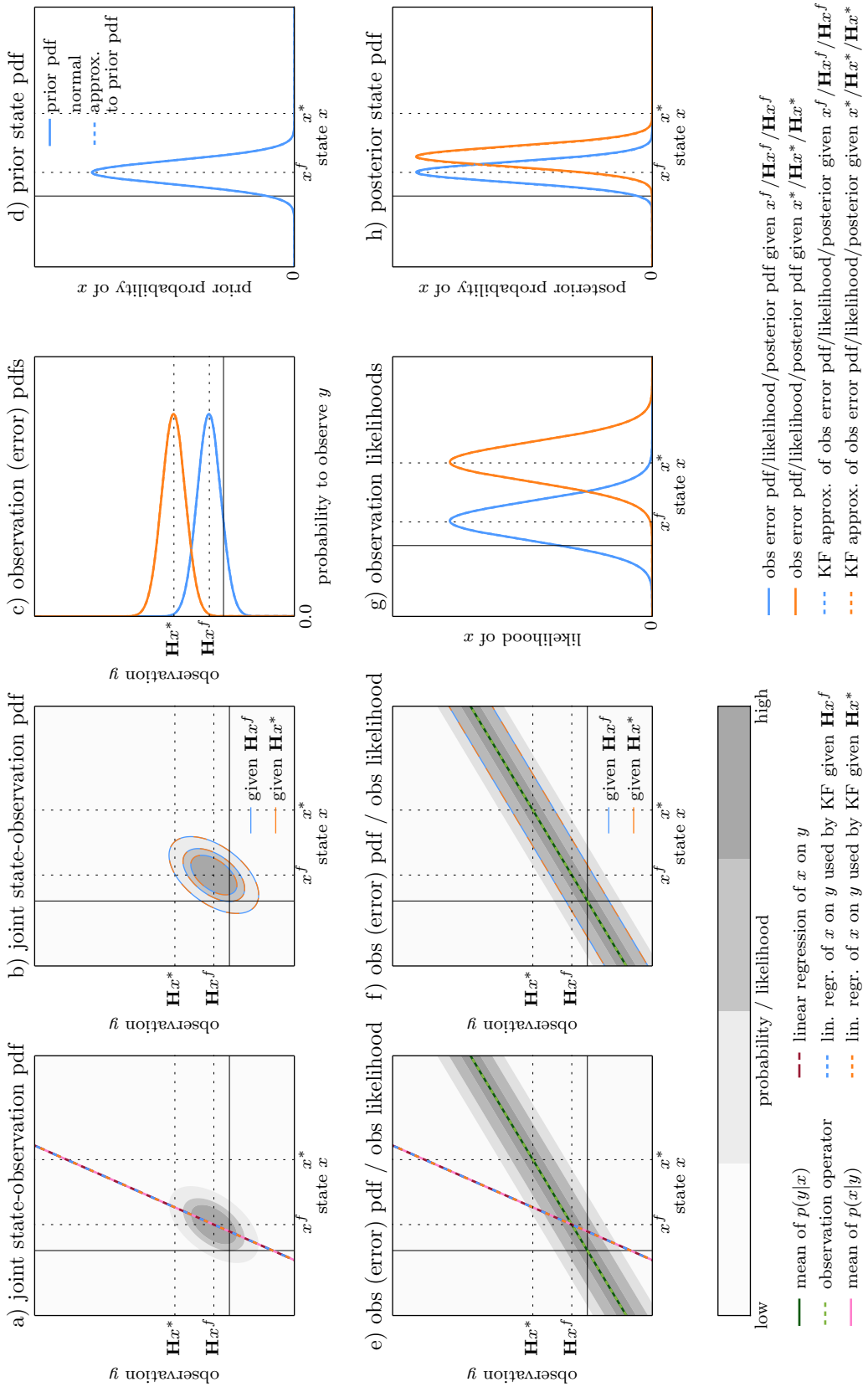


Figure 3.3: KF estimation with linear observation operator and Gaussian distributions (continued on next page).

Figure 3.3: The KF estimate of the linear regression for two different realisations of the same observation (different random observation error) is shown as orange and blue dashed lines in panels a) and e). The true linear regression line is shown as purple dashed line and the curve of the conditional mean of x given y is shown as pink solid line. The true joint pdf is shown as filled grey contours in panel a) and b) and the implicit approximations of the KF are shown as orange and blue contours. The true observation pdf (function of y) and the true observation likelihood (function of x) are shown as filled grey contours in panels e) and f). The observation operator is shown as light green, dashed line and the mean of the observation pdf is shown as dark green, solid line. The implicit approximations of the KF to the true observation pdf/true likelihood are shown as orange and blue contours in panels e) and f). The true observation error pdf and the KF approximation shifted to the perfect observation $\mathbf{H}x$ for two different true states corresponding to the realisations of the observation are shown in panel c). The prior pdf and the KF approximation are shown in panel d). The true likelihood of x given two different realisations of the observation and the KF approximations are shown in panel g). And the true posterior pdf and the KF approximation for two different observations are shown in panel f).

The terms $\mathrm{cov}(\mathbf{x}, \mathbf{y})$, $\mathrm{cov}(\mathbf{y}, \mathbf{y})$ and $E(h(\mathbf{y}))$ cannot easily be determined. But we can write (without loss of generality assuming $E(\boldsymbol{\varepsilon}) = \mathbf{0}$)

$$\begin{aligned}
\mathrm{cov}(\mathbf{y}, \mathbf{y}) &= E\left((\mathbf{y} - E(\mathbf{y}))\,(\mathbf{y} - E(\mathbf{y}))^T\right) && (3.48)\\
&= E\left((\mathbf{y} - E(h(\mathbf{x})))\,(\mathbf{y} - E(h(\mathbf{x})))^T\right)\\
&= E\left((h(\mathbf{x}) + \boldsymbol{\varepsilon} - E(h(\mathbf{x})))\,(h(\mathbf{x}) + \boldsymbol{\varepsilon} - E(h(\mathbf{x})))^T\right)\\
&= E\left((h(\mathbf{x}) - E(h(\mathbf{x})))\,(h(\mathbf{x}) - E(h(\mathbf{x})))^T\right) + E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right)\\
&= \mathrm{cov}(h(\mathbf{x}), h(\mathbf{x})) + \mathbf{R} && (3.49)
\end{aligned}$$

where the cross terms between $(h(\mathbf{x}) - E(h(\mathbf{x})))$ and $\boldsymbol{\varepsilon}$ are zero because the observation error is independent of the state. Using the independence assumption again we get

$$\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \mathrm{cov}(\mathbf{x}, h(\mathbf{x})) \tag{3.50}$$

and

$$\mathrm{cov}(h(\mathbf{x}), \mathbf{y}) = \mathrm{cov}(h(\mathbf{x}), h(\mathbf{x})). \tag{3.51}$$

Now we use the ensemble of predicted observations

$$\mathbf{y}^{f,i} = h(\mathbf{x}^{f,i})$$

and formulate the ensemble of linear regressions with the augmented state vector (section 3.2.4)

$$\begin{aligned}\begin{pmatrix}\mathbf{x}^{a,i}\\ \mathbf{y}^{a,i}\end{pmatrix} &= \begin{pmatrix}\mathbf{x}^{f,i}\\ \mathbf{y}^{f,i}\end{pmatrix}\begin{pmatrix}\mathrm{cov}(\mathbf{x},\,\mathbf{y})\\ \mathrm{cov}(h(\mathbf{x}),\,\mathbf{y})\end{pmatrix}(\mathrm{cov}(\mathbf{y},\,\mathbf{y})+\mathbf{R})^{-1}\,(\mathbf{y}^{i}-\mathbf{y}^{f,i})\\ &= \begin{pmatrix}\mathbf{x}^{f,i}\\ \mathbf{y}^{f,i}\end{pmatrix}\begin{pmatrix}\mathrm{cov}(\mathbf{x},\,h(\mathbf{x}))\\ \mathrm{cov}(h(\mathbf{x}),\,h(\mathbf{x}))\end{pmatrix}(\mathrm{cov}(h(\mathbf{x}),\,h(\mathbf{x}))+\mathbf{R})^{-1}\,(\mathbf{y}^{i}-\mathbf{y}^{f,i}),\end{aligned}\tag{3.52}$$

where all covariance matrices can be estimated from the ensemble. The updated ensemble members $\mathbf{x}^{a,i}$, $i = 1, \ldots, N$, can be derived directly from the perturbed observations $\mathbf{y}^i = \mathbf{y} + \boldsymbol{\varepsilon}^i$ as in the perturbed observations EnKF. Alternatively, the state increments $\mathbf{x}^{a,i} - \mathbf{x}^{f,i}$ can be derived from the observation increments $\mathbf{y}^{a,i} - \mathbf{y}^{f,i}$ using the linear approximation to the observation operator given by

$$\mathbf{x}^{a,i} - \mathbf{x}^{f,i} = \mathrm{cov}(\mathbf{x},\,h(\mathbf{x}))\,\mathrm{cov}(h(\mathbf{x}),\,h(\mathbf{x}))^{-1}\,(\mathbf{y}^{a,i} - \mathbf{y}^{f,i}).\tag{3.53}$$

The EAKF updates the ensemble in this way without perturbations of the actual observation $\mathbf{y}$. The equivalence of the EAKF to the direct update of $\mathbf{x}^{f,i}$ from perturbed observations $\mathbf{y}^i$ is seen by deriving the implicitly used perturbed observations of the EAKF from (3.48) – (3.51) and by using the expression for $\mathbf{y}^{a,i}$ from the second line of (3.52) in (3.53). With increasing ensemble size, the mean of the updated ensemble in (3.52) then converges to $\mathbf{x}^a$ as defined in (3.47). The same holds for the ensemble covariance matrix and the linear regression error covariance matrix.

Non-Gaussian prior distributions, non-Gaussian observation errors, or nonlinear observation operators cause the joint pdf $p(\mathbf{x},\,\mathbf{y})$ to be non-Gaussian. Consequently, the conditional mean will be a nonlinear function $f(\mathbf{y})$ and the conditional covariance matrix will depend on $\mathbf{y}$. As argued before, the KF uses linear regression to approximate the conditional mean and the linear regression error covariance matrix to approximate the conditional covariance matrix and these approximations define a normal pdf which is an approximation of the true conditional pdf.

In the update step, the KF approximates the nonlinear relationship between the state vector and the observations with the linear relationship given by the cross-covariances $\mathrm{cov}(\mathbf{x},\,h(\mathbf{x})$. Together with the Gaussian distributions, this leads to an approximating joint normal pdf as shown in Figure 3.4. As in section 3.6.1, due to the constant observation error covariance, the true linear regression and the KF approximations agree. But because the true joint pdf is non-Gaussian, the linear regressions are only an approximation to the nonlinear conditional mean $f(\mathbf{y})$. Further, the linear approximation of the observation operator leads to a shift in the observation pdf that corresponds to the vertical difference between the nonlinear observation operator and its linear approximation in panel f).

The difference in the location of the true likelihoods and their approximations corresponds to the horizontal distance between the nonlinear observation operator and its linear

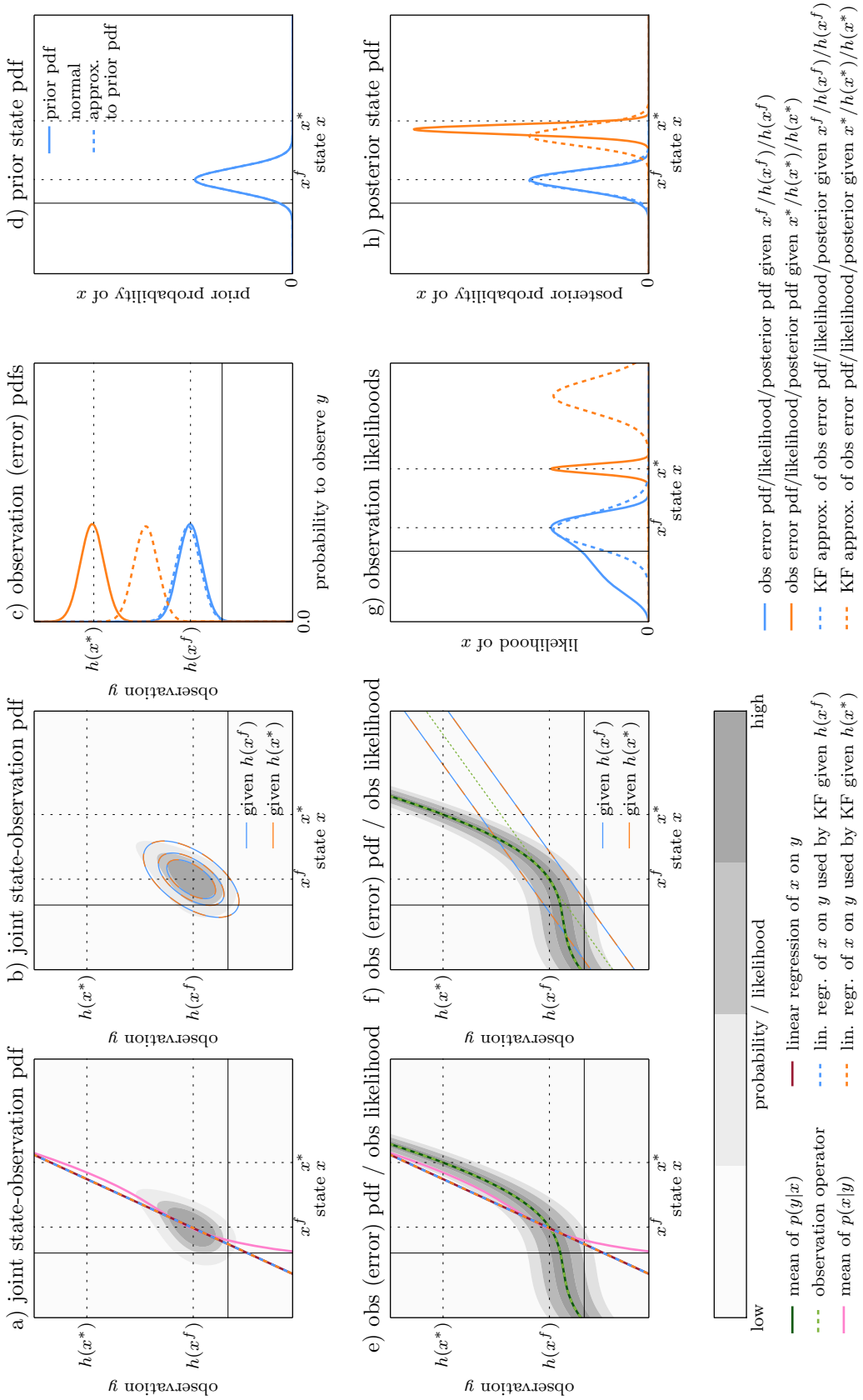


Figure 3.4: KF estimation with nonlinear observation operator and Gaussian distributions. The linear approximation of the observation operator is shown as thin, dashed, light green line in panel f) (for detailed explanation see Figure 3.3).

approximation at the level of the realisation of the observation. The non-Gaussian shape of the true likelihoods arises from the nonlinearity of the observation operator, which causes the observation error pdf to be shifted differently for the same change $\Delta\mathbf{x}$ depending on $\mathbf{x}$. The Gaussian shape of the approximations is due to linear approximation of the observation operator which leads to a Gaussian joint pdf.

The differences between the locations of the true posterior pdfs and the KF approximations in panel f) mirror the horizontal distance between the linear regression lines and the curve of the conditional mean at the level of the realisation of the observation in panels a) and e). The differences in the shape and width of the true posterior pdfs and the KF approximations originate from the approximation of the true joint pdf by a normal joint pdf. Figure 3.4 visualises the fact, the for nonlinear observation operators, the KF estimates will only be good approximations when the assimilated observations are in the vicinity of $h(\mathbf{x}^f)$, that means the perfect observation that would have originated from the mean of the prior state vector.

In case of a linear observation operator and non-Gaussian prior distribution or non-Gaussian observation errors, the KF approximation would also agree with the true linear regression. The differences between the linear regressions and $f(\mathbf{y})$ would be qualitatively similar. The approximating likelihoods would not be shifted (this is an effect of the linear approximation of the nonlinear observation operator) but their shape would not match the true likelihood, either, due to the normal approximation. The difference between the KF approximations of the posterior pdf and the true posterior pdf would also be qualitatively similar.

3.6.3 The Gaussian case with state-dependent observation error covariance

For the derivation of $\mathrm{cov}(\mathbf{x}, \mathbf{y})$ and $\mathrm{cov}(\mathbf{y}, \mathbf{y})$, we have so far assumed that the observation error covariance $\mathbf{R}$ is constant and independent of the observed state $\mathbf{x}$ (homoscedastic errors). Consider now observations

$$\mathbf{y} = h(\mathbf{x}) + \boldsymbol{\varepsilon}(\mathbf{x})$$

with zero-mean but state-dependent observation error $\boldsymbol{\varepsilon}(\mathbf{x})$ and thus with a state-dependent observation error covariance (heteroscedastic errors).

The distribution of $\mathbf{y}$ now depends twofold on $\mathbf{x}$, through the location given by $h(\mathbf{x})$ and through the shape of the distribution of $\boldsymbol{\varepsilon}(\mathbf{x})$. Consequently, the joint pdf will be non-Gaussian even if all error distributions $p_{\boldsymbol{\varepsilon}(\mathbf{x})}(\mathbf{y} - h(\mathbf{x}))$ are Gaussian. We may still assume that the observation errors and the state are uncorrelated. In this case, the covariance

matrix of $\mathbf{y}$ is

$$\begin{aligned}\operatorname{cov}(\mathbf{y}, \mathbf{y}) &= E\left((\mathbf{y} - E(h(\mathbf{x})))\,(\mathbf{y} - E(h(\mathbf{x})))^T\right) \\ &= E\left((h(\mathbf{x}) + \boldsymbol{\varepsilon}(\mathbf{x}) - E(h(\mathbf{x})))\,(h(\mathbf{x}) + \boldsymbol{\varepsilon}(\mathbf{x}) - E(h(\mathbf{x})))^T\right) \\ &= E\left((h(\mathbf{x}) - E(h(\mathbf{x})))\,(h(\mathbf{x}) - E(h(\mathbf{x})))^T\right) + E\left(\boldsymbol{\varepsilon}(\mathbf{x})\boldsymbol{\varepsilon}(\mathbf{x})^T\right) \\ &= \operatorname{cov}(h(\mathbf{x}), h(\mathbf{x})) + E\left(\mathbf{R}(\mathbf{x})\right)\end{aligned}$$

while the other covariance matrices remain unchanged.

Analogous to the modification of the KF for state-dependent observation error covariances (section 3.3.2), the perturbed observations EnKF can be extended to accommodate $E(\mathbf{R}(\mathbf{x}))$. First note that

$$\begin{aligned}E(\mathbf{R}(\mathbf{x})) &= E(\boldsymbol{\varepsilon}(\mathbf{x})\boldsymbol{\varepsilon}(\mathbf{x})^T) \\ &= E\left(E(\boldsymbol{\varepsilon}(\mathbf{x})\boldsymbol{\varepsilon}(\mathbf{x})^T|\mathbf{x})\right)\end{aligned}$$

where the inner expectation is taken over $\boldsymbol{\varepsilon}$ and the outer expectation is taken over $\mathbf{x}$. Thus, we can obtain a sample estimate of $\mathbf{R}(\mathbf{x})$ by sampling the state-dependent observation error $\boldsymbol{\varepsilon}(\mathbf{x})$ for every member $\mathbf{x}^{f,i}$. Ensemble square root filters, on the other hand, require the analytical calculation of $E(\mathbf{R}(\mathbf{x}))$ because they avoid the additional sampling of the observation error and use a prescribed observation error covariance matrix.

An ad-hoc approach for assigning state-dependent observations errors is to use the observation error covariance which results from inverting the observation operator, that is,

$$\hat{\mathbf{R}} = \mathbf{R}(h^{-1}(\mathbf{y})).$$

The consequences of this approach are illustrated in Figure 3.5. Two different realisations of the same observation lead to two different approximating joint normal pdfs. This is statistically inconsistent because the joint pdf describes the probability of the realisations of observations and the approximation used in the KF should be independent of the realisations of observations. The two different approximations to the joint pdf explain also the two different KF regression lines which differ both from the true linear regression because both KF approximations use an incorrect observation error covariance.

The observation error approximations are correct by construction. The KF approximations to the likelihoods are acceptable in the proximity of the prior mode which results in good approximations of the posterior covariance (width of the posterior pdf). But due to the error in the estimated regression curves, the locations of the KF approximations of the posterior pdf are wrong. The blue posterior pdf is by chance closer to the true posterior pdf because the observation error covariance used in this case is close to the true observation error covariance which makes the blue KF regression line by chance a good approximation of the true linear regression.

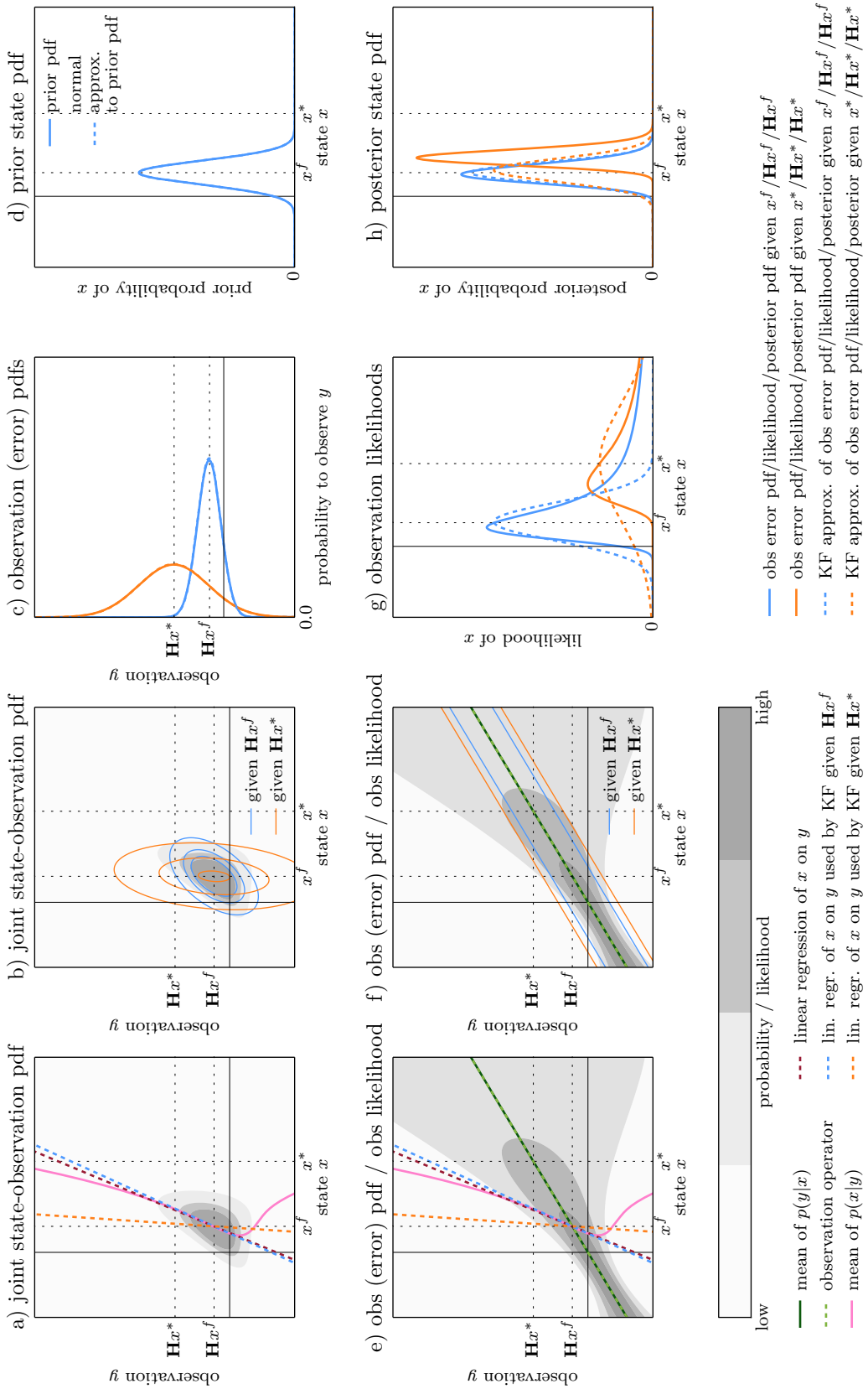


Figure 3.5: KF estimation with linear observation operator and Gaussian distributions with state-dependent observation error covariance (for detailed explanation see Figure 3.3).

3.6.4 Non-zero mean observation errors and state-correlated observation errors

If the observation errors do not have zero mean, the expectation of $\mathbf{y}$ will no longer be the expectation of the observation operator applied to $\mathbf{x}$,

$$\begin{aligned} E(\mathbf{y}) &= E(h(\mathbf{x})) + E(\boldsymbol{\varepsilon}) \\ &\neq E(h(\mathbf{x})). \end{aligned}$$

Therefore, the linear regression estimate of $\mathbf{x}$ given $\mathbf{y}$ reads

$$\begin{aligned} \hat{f}(\mathbf{y}) &= \mathbf{x}^f + \mathrm{cov}(\mathbf{x}, \mathbf{y})(\mathrm{cov}(\mathbf{y}, \mathbf{y}))^{-1}(\mathbf{y} - E(h(\mathbf{x})) - E(\boldsymbol{\varepsilon})) \\ &= \mathbf{x}^a - \mathrm{cov}(\mathbf{x}, \mathbf{y})(\mathrm{cov}(\mathbf{y}, \mathbf{y}))^{-1}E(\boldsymbol{\varepsilon}). \end{aligned}$$

Similar to the state-dependent observation errors, the term $E(\boldsymbol{\varepsilon})$ could be included in the sampling of perturbed observations in the stochastic EnKF. If not accounted for, non-zero mean observation errors will lead to a shift of the regression line estimated by the KF along the $\mathbf{x}$-coordinates. This shift results in a bias of the estimated conditional mean.

Even for the state-dependent observation errors, we hitherto assumed that they are uncorrelated with the state. Otherwise, the cross-covariance terms

$$E((\mathbf{x} - E(\mathbf{x}))(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon}))^T)$$

and

$$E((h(\mathbf{x}) - E(h(\mathbf{x}))(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon}))^T)$$

in the derivations of $\mathrm{cov}(\mathbf{y}, \mathbf{y})$ and $\mathrm{cov}(\mathbf{x}, \mathbf{y})$ will not be zero. This assumption, however, ensures that the ensemble estimates of $\mathrm{cov}(\mathbf{y}, \mathbf{y})$ and $\mathrm{cov}(\mathbf{x}, \mathbf{y})$ are unbiased. Errors in the estimates of $\mathrm{cov}(\mathbf{y}, \mathbf{y})$ and $\mathrm{cov}(\mathbf{x}, \mathbf{y})$ cause errors in the slope of the estimated regression line as can be seen in Figure 3.6. In this example, the prior pdf is defined on the bounded interval $(0, 1)$. The observation error pdf is chosen such that its mode is at the perfect observation and that the observation is within $(0, 1)$. Due to the boundedness, the prior pdf and the observation error pdf are non-Gaussian and the joint pdf of $\mathbf{x}$ and $\mathbf{y}$ is therefore also non-Gaussian and defined on $(0, 1) \times (0, 1)$.

The conditional mean of $\mathbf{x}$ given $\mathbf{y}$ is a nonlinear function that maps every observation $\mathbf{y}$ to a value in $(0, 1)$. The mean of the observation pdf does not coincide with the observation operator which implies $E(\mathbf{y}) \neq E(h(\mathbf{x}))$ or equivalently $E(\boldsymbol{\varepsilon}) \neq 0$. Moreover the observation error pdf is state dependent, as shown by the bent grey contours in panel e) and f) which are not parallel to the observation operator.

Approximating the prior distribution and the observation error distributions with normal distributions, leads to an approximating joint normal distribution that has non-zero

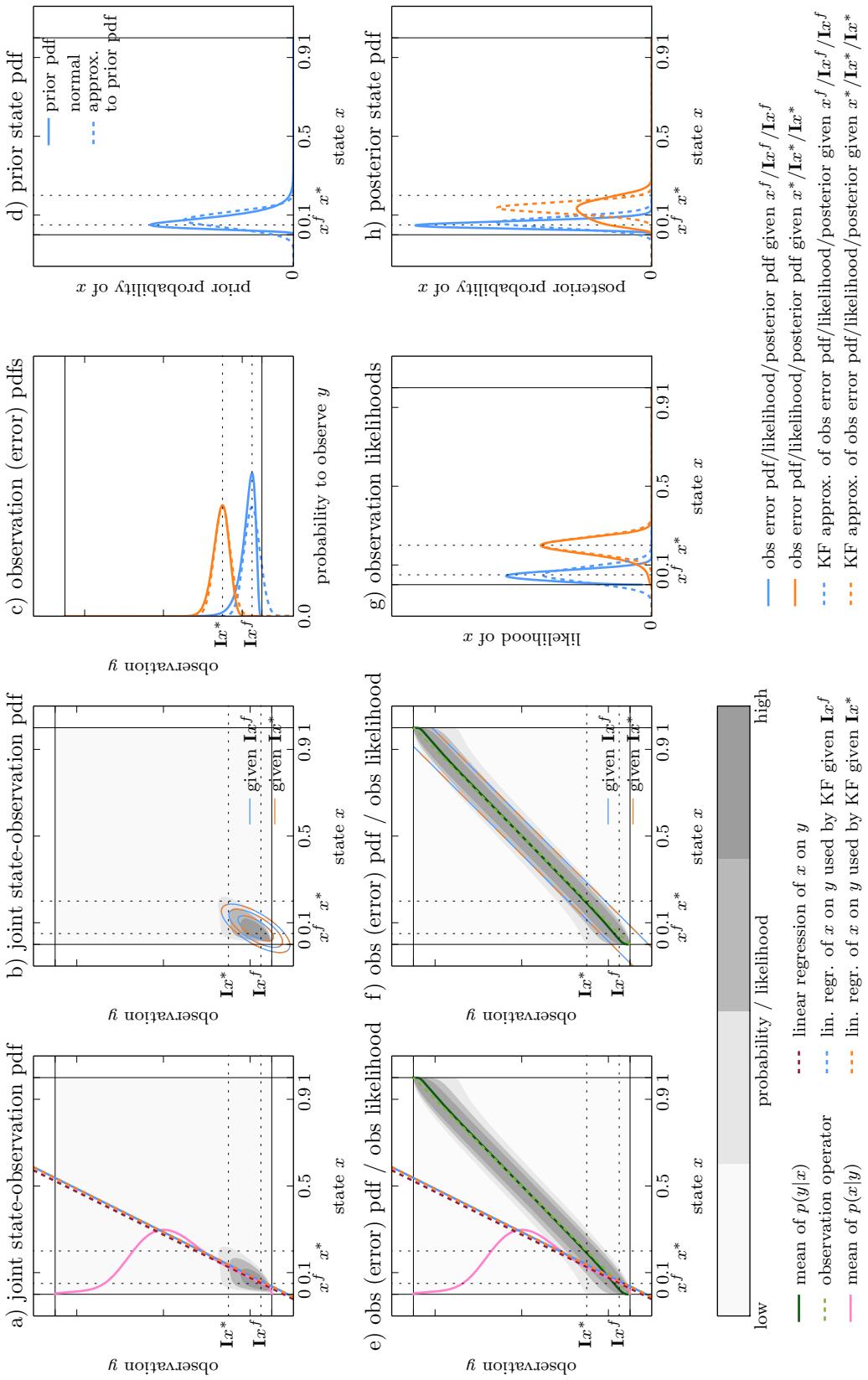


Figure 3.6: KF estimation with linear observation operator with state-dependent observation error distribution, non-zero mean observation error and non-Gaussian prior (for detailed explanation see Figure 3.3).

probabilities outside the bounded domain of the state and the observation. And likewise, the linear regression of $\mathbf{x}$ on $\mathbf{y}$ approximates the nonlinear conditional mean of $\mathbf{x}$ given $\mathbf{y}$ but is not restricted to (0, 1). The KF approximations of the linear regression coincide for both realisations of the observation because the observation error covariance is constant. They are shifted relative to the true linear regression along the $\mathbf{x}$-axis by $\text{cov}(\mathbf{x}, \mathbf{y})(\text{cov}(\mathbf{y}, \mathbf{y}))^{-1}E(\boldsymbol{\varepsilon})$ because of the non-zero mean observation errors. And they are not parallel to the true linear regression line, that is, they have a different slope, because the observation errors are correlated with the state (for small $\mathbf{x}$, positive $\boldsymbol{\varepsilon}$ are more likely than negative $\boldsymbol{\varepsilon}$ and with increasing $\mathbf{x}$ negative $\boldsymbol{\varepsilon}$ become more likely compared to positive $\boldsymbol{\varepsilon}$ which results in a slight negative correlation). For a discussion on the effects of the bounded domain see section 3.7.

As previously the KF approximations of the posterior pdf are horizontally shifted due to the horizontal distance between the conditional mean curve and the KF approximations of the linear regression. Further, the joint pdf and the posterior pdfs of $\mathbf{x}$ given $\mathbf{y}$ are significantly non-Gaussian and thus the error covariance of the linear regression is not a good approximation of the conditional covariance. Consequently the width of the KF approximations of the posterior pdf disagree strongly with the width of the true posterior pdfs. Lastly, the KF approximations assign positive probabilities to states outside (0, 1) which is inconsistent with the true prior pdf and the bounded domain of the state (see also section 3.4).

3.6.5 Summary of errors in estimated conditional means and conditional covariance matrices

The KF corresponds to a linear regression of the state vector $\mathbf{x}$ on the observations $\mathbf{y}$ and the result of this regression is an approximation of the conditional mean and of the conditional covariance matrix of $\mathbf{x}$ given $\mathbf{y}$. But, depending on which KF assumptions are or are not satisfied, the KF only approximates the estimates that would result from a linear regression. The estimated conditional mean and covariance matrix can be used to approximate the posterior pdf with a normal pdf. In this interpretation the KF is an approximate Bayesian computation algorithm (ABC algorithm; Nott et al., 2011). The total error of this approximation depends on three sources of error:

1. the error due to the approximation of the (nonlinear) function $f(\mathbf{y})$ that defines the conditional mean with a linear function $\hat{f}(\mathbf{y})$,
2. the error in the estimate of the linear function $\hat{f}(\mathbf{y})$,
3. the error of approximating a non-Gaussian posterior pdf with a Gaussian pdf.

The first type of error arises from non-Gaussian prior distributions, non-Gaussian observation errors, nonlinear observation operators, or state-dependent observation errors. These all induce a non-Gaussian joint pdf and a nonlinear dependency of the conditional

mean on the observation. Consequently, the linear regression estimate and its error covariance matrix are only approximations of the conditional mean and the conditional covariance matrix instead of agreeing with the true values.

The second type of error arises from using the KF equations to estimate the linear regression of $\mathbf{x}$ on $\mathbf{y}$ and not accounting for non-zero mean observation errors, state-dependent observation errors and correlations of the observation error with the state. These errors cause the estimated regression line be shifted along the x-axis and they cause errors in the covariance matrices cov($\mathbf{x}$, $\mathbf{y}$ and cov($\mathbf{y}$, $\mathbf{y}$) which lead to an incorrect slope of the estimated regression line. For either of the first two types of errors, the estimated conditional mean of $\mathbf{x}$ given $\mathbf{y}$ will be shifted from its true value. And the errors in the estimated conditional covariance matrices will increase with the mismatch between the true joint pdf $p(\mathbf{x}, \mathbf{y})$ and the normal joint pdf which is implicitly used to approximate the true one.

The third type of error arises if the posterior pdf is non-Gaussian due to any of a non-Gaussian prior distribution, non-Gaussian observation errors, a nonlinear observation operator, or state-dependent observation errors. This type of error becomes relevant when other properties than the mean or the covariance of the posterior pdf, for example its mode, are sought after (section 3.7).

3.7 Gaussian anamorphosis for the assimilation of bounded quantities

The quality of the KF estimates of the conditional mean and the conditional covariance matrix depends on the joint pdf of $\mathbf{x}$ and $\mathbf{y}$ being approximately Gaussian. The quality of the estimates deteriorates as the joint pdf becomes less Gaussian while, at the same time, the normal approximation of the conditional pdf becomes less useful because the conditional pdf will also depart from Gaussianity. And a non-Gaussian conditional pdf can in general not be adequately described by only its mean and covariance matrix, even if these estimates are correct. We may, however, restrict the conditional pdf to be of such a type that it can be characterised by its first two moments, even if it is non-Gaussian, and we may then try to improve the estimated values of the first two moments. This is the idea of Gaussian anamorphosis applied in conjunction with the EnKF.

Gaussian anamorphosis (Chilès and Delfiner, 1999) transforms a random variable x such that the transformed variable $\tilde{x}$ follows a Gaussian distribution. The estimates of the mean and the covariance of $\tilde{x}$ fully describe the distribution of $\tilde{x}$ and thus also the distribution of x, even if this distribution is non-Gaussian. This approach is also called normal score transform (Krzysztofowicz, 1997; and references therein). In conjunction with the EnKF, Bertino et al. (2002, 2003) suggested to use Gaussian anamorphosis to transform the state vector and the observations such that their distributions are Gaussian or close to Gaussian, which improves the quality of the KF estimates. Gaussian anamorphosis is

particularly useful for the assimilation of bounded quantities because it can transform bounded variables into unbounded variables which are used in the assimilation process. The inverse transformation ensures estimates that are then consistent with the variable's bounds.

3.7.1 Assimilation of bounded quantities

Section 3.4 explains how physically inconsistent updated states appear due to the purely statistical nature of the KF and its use of linear relationships between state vector and observations. Gaussian anamorphosis uses a variable transformation (section 3.7.3) to improve the Gaussianity of the variables in the transformed space that are used in the assimilation. If this transformation maps the bounded quantities in state space to an unbounded domain in the transformed space, then the inverse transformation of the estimates from the transformed space to the physical space ensures physically consistent estimates. The same can be achieved with any such transformation only to ensure physically consistent estimates, without considering the effects of the variable transformation on the distribution of the variable in the transformed space (Nielsen-Gammon et al., 2010).

Section 3.6 explains how nonlinearity and non-Gaussianity cause errors in the estimated conditional mean and the estimated conditional covariance. These estimation errors cause biases in the estimates which are derived from the KF approximation of the conditional pdf because this pdf is shifted. The adverse effects are particularly strong if states and observations close to the bounds of the interval are considered. Panel e) of Figure 3.6 shows the linear regression used by the KF to estimate the conditional mean. Compared to the true curve of the conditional mean of $\mathbf{x}$ given $\mathbf{y}$, these estimates have a bias towards the centre of the interval (the curves would be symmetric about $x = 0.5$ for values close to 1). Even for observations very close to zero, the estimated conditional mean is very different from zero and any approximating pdf that uses this estimate will be shifted to the centre of the interval. Consequently, any estimates derived from this approximate pdf, such as its mean, will be biased.

Lastly, the use of approximate conditional pdfs that are Gaussian, and thus assign non-zero probabilities to values outside the physical domain of the state variables, complicates the interpretation of these pdfs and any estimates derived from them.

3.7.2 Estimation of conditional mode

For practical applications such as the use of an estimated parameter in a model, the approximate conditional pdf has to be reduced to one value, the state or parameter estimate. The two intuitive estimates are the conditional mean, which minimises the expected squared error of the estimate and which can be regarded as the "average estimate", and the conditional mode. The conditional mode has the highest probability to be the true value and is also referred to as the maximum a posteriori estimate (MAP estimate). For normal distributions, both estimates coincide and no decision has to be made. In general,

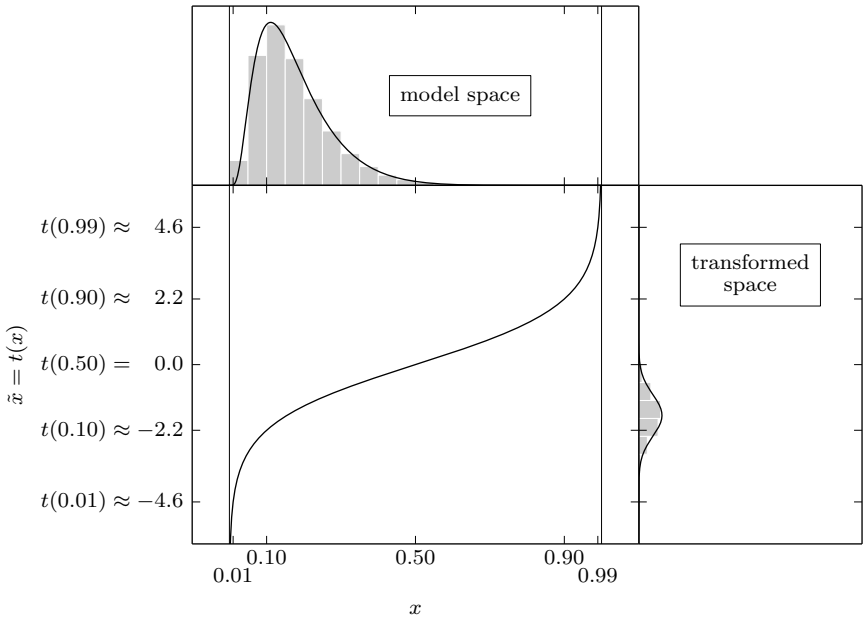


Figure 3.7: Transformation of a double-bounded random variable from (0, 1) to $(-\infty, \infty)$ using the logit function described in section 3.7.4.

however, they will differ. We argue that the conditional mode is the intuitively more appealing estimate because the conditional mean could, after all, be a very unlikely state or parameter value, for example if the conditional pdf has a sharp peak and a long, flat tail or if the conditional pdf is strongly bimodal.

3.7.3 Transformation of states and observations

The transformation of a random variable x to a new variable $\tilde{x}$ changes the distribution of x. The distribution of the transformed variable can be derived from the cumulative distribution function, which is the anti-derivative of the pdf, and the rules for the change of variables in integration (Figure 3.7). Given a bijective transformation

$$\tilde{x} = t(x), \tag{3.54}$$

the pdf of $\tilde{x}$ is

$$\begin{aligned} p_{\tilde{x}}(\tilde{x}) &= p_x(t^{-1}(\tilde{x})) \, \frac{\mathrm{d}t^{-1}(\tilde{x})}{\mathrm{d}\tilde{x}} \\ &= p_x(x) \frac{\mathrm{d}x}{\mathrm{d}\tilde{x}}. \end{aligned} \tag{3.55}$$

For random vectors, the transformation reads

$$\begin{aligned} p_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}}) &= p_{\mathbf{x}}(t^{-1}(\tilde{\mathbf{x}})) \left| \frac{\mathrm{d}t^{-1}(\tilde{\mathbf{x}})}{\mathrm{d}\tilde{\mathbf{x}}} \right| \\ &= p_{\mathbf{x}}(\mathbf{x}) \left| \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tilde{\mathbf{x}}} \right|, \end{aligned}$$

where $\left|\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tilde{\mathbf{x}}}\right|$ is the determinant of the Jacobian of the inverse transformation t^{-1} evaluated at $\tilde{\mathbf{x}}$ (Papoulis, 1991).

These transformations are commonly applied to single variables, changing the univariate marginal distributions $p(x^j)$, where x^j is the j-th element of the state vector, because this simplifies the application of Gaussian anamorphosis significantly (Bocquet et al., 2010). Applying the transformations in this univariate fashion, however, does not ensure multivariate Gaussian distributions in the transformed space that would be required for the KF estimates to be optimal. Hence to ensure optimal updates of the transformed state vector, at least bi-Gaussianity of the transformed state-observation pairs has to be checked (Brankart et al., 2012).

The EnKF uses an ensemble of states to represent the non-Gaussian prior pdf and the transformation of this pdf corresponds to the transformation of all ensemble members according to a transformation $t_{\mathbf{x}}$. The transformed ensemble then represents the transformed prior pdf and is a true sample of a normal distribution. The anamorphosis can be applied to the whole state vector and the observations as well as to parts of the state vector or the observations only. Also, different transformations for the state vector and the observations are possible. In general, the anamorphosis will change the relation between the state vector and the observations. Consider a transformation $t_{\mathbf{x}}$ for the state vector (note that individual components of $t_{\mathbf{x}}$ may be the identity operator such that the variable is not changed) and a transformation $t_{\mathbf{y}}$ for the observations. The transformed observation operator is then given by (Bertino et al., 2002)

$$\tilde{h}(\tilde{\mathbf{x}}) = t_{\mathbf{y}} \circ h \circ t_{\mathbf{x}}^{-1}(\tilde{\mathbf{x}}),$$

where $\circ$ means the composition of two functions. The choices of $t_{\mathbf{x}}$ and $t_{\mathbf{y}}$ may improve or deteriorate the linearity between states and observations. For our application, which is the estimation of canopy albedo parameters from surface albedo observations, we assume identity observations of the first part of the state vector in model space and we use the same transformation t for states and observations such that

$$\begin{aligned} \tilde{\mathbf{y}} &= \tilde{h}(\tilde{\mathbf{x}}) \\ &= t \circ \begin{pmatrix} \mathbf{I}_m & \mathbf{0}_{n-m} \end{pmatrix} \circ t^{-1}(\tilde{\mathbf{x}}) \\ &= \begin{pmatrix} \mathbf{I}_m & \mathbf{0}_{n-m} \end{pmatrix} \tilde{\mathbf{x}}. \end{aligned}$$

Such an observation operator corresponds to identity observations of m model states while

the state vector is augmented with $n - m$ model parameters that are not observed.

3.7.4 Choice of the anamorphosis function and definition of model space distributions

The transformation or anamorphosis function t can be chosen ad-hoc or constructed numerically from an ensemble of states as well as from an ensemble of observations (Simon and Bertino, 2009; Brankart et al., 2012). Motivated by the application for albedo, we choose the logit function

$$t(x) = \ln(x) - \ln(1 - x) \tag{3.56}$$

that maps (0, 1) to $(-\infty, \infty)$ and whose inverse is the logistic function

$$t(\tilde{x}) = \frac{\exp(\tilde{x})}{\exp(\tilde{x}) + 1}. \tag{3.57}$$

The ad-hoc choice of an anamorphosis function avoids the problem of defining the tails of a numerically constructed function beyond the last data points of the ensemble and simplifies the implementation. The logit transform is also applicable to other double-bounded intervals if the variables are appropriately shifted and scaled. The end points of the interval are excluded because the logarithm is not defined there.

Our choice of the distributions of the state vector and the observation error in model space result from the choice of the transformation and the requirement that the transformed distributions must be Gaussian. This leads to logit-normal distributions (Johnson and Kotz, 1970) in model space . The prior state distribution is chosen such that its mode is at the best available prior estimate and that it has a standard deviation that represents the prior uncertainty. The observation error pdfs are defined such that they have mode zero and a standard deviation that corresponds to the assumed observation error standard deviation.

Numerical calculation of transformed space distributions

Given the mode x_{mode} and the variance σ_x^2 of the logit-normal pdf in model space, we calculate the mean and the variance of the normal pdf in transformed space numerically using (3.55) and

$$\frac{\mathrm{d}p_x(x)}{\mathrm{d}x}\Big|_{x_{\text{mode}}} = 0 \tag{3.58}$$

together with the integral definition of the mean μ_x and the variance σ_x^2,

$$\mu_x = \int_x x p_x(x) \mathrm{d}x, \tag{3.59}$$

$$\sigma_x^2 = \int_x (x - \mu_x)^2 p_x(x) \mathrm{d}x. \tag{3.60}$$

Because of the computational cost, we store the results for a variance of 0.0001 and 0.0016 in model space (corresponding to standard deviations of 0.01 and 0.04, see section 4.2.2) for all modes $0.00001, \ldots, 0.99999$ in a look-up table that enables the efficient conversion of model space distributions to transformed space distributions and the efficient numerical construction of likelihood functions. In order to represent an observation error distribution, the respective distribution in model space whose mode is at the observed value is shifted such that its mode is at zero, equivalent to the assumption that the perfect observation is the most probable one.

3.7.5 Transformation of observations and observation error

The transformation of the observation itself is straightforward, it is an evaluation of the anamorphosis function $t_{\mathbf{y}}$. The observation error covariance of the transformed observation, however, cannot be derived directly from the observation error in model space and no formal derivation has been given so far. In previous applications of Gaussian anamorphosis, the transformed observation error covariance has been derived by ad-hoc assumptions. Doron et al. (2011, 2013) used small observation error covariances that justify similarly small error covariances in the transformed space. Fontana et al. (2013) and Lien et al. (2013) linearised the anamorphosis function locally and scaled the observation error standard deviation with the slope of the anamorphosis function. Schöniger et al. (2012) and Simon and Bertino (2012) suggested to use an ensemble of perturbed observations, transform them and estimate the covariance of the transformed observation error from the transformed ensemble. We formalise this approach, analyse it, and provide an alternative approach that avoids hitherto unnoted shortcomings.

Calculation of transformed observation error covariance from the transformed observation pdf

In section 3.7.3, we have introduced the transformation $t_{\mathbf{y}}$ that yields the transformed observation

$$\tilde{\mathbf{y}} = t_{\mathbf{y}}(\mathbf{y}).$$

The KF requires the covariance of the error of the transformed observation

$$\tilde{\boldsymbol{\varepsilon}} = \tilde{\mathbf{y}} - \tilde{h}(\tilde{\mathbf{x}}),$$

that is, the KF requires the covariance of $\tilde{\boldsymbol{\varepsilon}}$. A direct derivation of the distribution of $\tilde{\boldsymbol{\varepsilon}}$ from the distribution of $\boldsymbol{\varepsilon}$ and from requiring that the distribution of

$$\tilde{h}(\tilde{\mathbf{x}}) + \tilde{\boldsymbol{\varepsilon}} = t_{\mathbf{y}} \circ h \circ t_{\mathbf{x}}^{-1}(\tilde{\mathbf{x}}) + \tilde{\boldsymbol{\varepsilon}}$$

is equal to the distribution of

$$t_{\mathbf{y}}\left(h(\mathbf{x}) + \boldsymbol{\varepsilon}\right)$$

is not possible because $t_{\mathbf{y}}$ is nonlinear (otherwise, the anamorphosis does not improve the Gaussianity nor does it map states and observations to an unbounded domain) and because the distribution of $\tilde{\boldsymbol{\varepsilon}}$ depends on $\mathbf{x}$ or, respectively, $\tilde{\mathbf{x}}$. Therefore, we suggest to derive $\tilde{\boldsymbol{\varepsilon}}$ from the distributions of $\tilde{\mathbf{y}}$ and $\mathbf{y}$ for an assumed true state $\mathbf{x}$. Given $\mathbf{x}$, the distribution of $\mathbf{y}$ is

$$p_{\mathbf{y}}(\mathbf{y}) = p_{\boldsymbol{\varepsilon}}(\mathbf{y} - h(\mathbf{x})),$$

where $p_{\boldsymbol{\varepsilon}}$ is the pdf of $\boldsymbol{\varepsilon}$ (cf. (3.7)). We transform $p_{\mathbf{y}}(\mathbf{y})$ according to the anamorphosis function $t_{\mathbf{y}}$ and get $p_{\tilde{\mathbf{y}}}(\tilde{\mathbf{y}})$. According to

$$p_{\tilde{\mathbf{y}}}(\tilde{\mathbf{y}}) = p_{\tilde{\boldsymbol{\varepsilon}}}(\tilde{\mathbf{y}} - \tilde{h}(\tilde{\mathbf{x}}))$$

and

$$\begin{aligned}\tilde{h}(\tilde{\mathbf{x}}) &= t_{\mathbf{y}} \circ h \circ t_{\mathbf{x}}^{-1}(t_{\mathbf{x}}(\mathbf{x})) \\ &= t_{\mathbf{y}} \circ h(\mathbf{x})\end{aligned}$$

the distribution of the transformed observations $p_{\tilde{\mathbf{y}}}(\tilde{\mathbf{y}})$ is equal to the distribution of the transformed observation error, shifted by $t_{\mathbf{y}} \circ h(\mathbf{x})$, where $p_{\tilde{\boldsymbol{\varepsilon}}}$ is the distribution of $\tilde{\boldsymbol{\varepsilon}}$. Note that because $t_{\mathbf{y}}$ is nonlinear and because of (3.55), the shape of the distribution of $\tilde{\boldsymbol{\varepsilon}}$ depends on $\mathbf{x}$ or, respectively, $\tilde{\mathbf{x}}$. Thus, the transformed observation error covariance is state-dependent, even if the original observation error covariance is constant. As explained in section 3.6.3, this is undesirable because it leads to different approximating joint pdfs for different realisations of the same observation (Figure 3.8).

When assimilating an observation $\mathbf{y}$ with the KF, we do not know its distribution nor do we know the true state that caused $\mathbf{y}$ (which determines the distribution of $\mathbf{y}$). An obvious ad-hoc solution to derive $\tilde{\boldsymbol{\varepsilon}}$ is therefore to use

$$\mathbf{x} = h^{-1}(\mathbf{y})$$

to derive the distribution of $\mathbf{y}$ and to subsequently derive the distribution of $\tilde{\boldsymbol{\varepsilon}}$. This corresponds to the ensemble of perturbed observations used by Simon and Bertino (2012).

We extend their method by using the full observation pdf – given by the observation error pdf shifted such that its mode is at the observation – instead of an ensemble and thus calculate the exact distribution and the exact covariance of the transformed observation error instead of approximating it with an estimate from an ensemble.

We apply this approach to the example from section 3.6.4 and Figure 3.6. The outcome of the transformation and the resulting KF estimation in the transformed space are shown in Figure 3.8. With respect to the desired bi-Gaussianity of the transformed joint pdf, which would improve the results of the linear estimation, the transformation does not yield any improvements compared to Figure 3.6, except that it is now defined on an unbounded domain. The conditional mean is still highly nonlinear and the grey contours in panel a) and b) still indicate a non-Gaussian joint pdf. The true linear regression approximates the conditional mean well over a smaller range considering the model space units. But it approximates the conditional mean much better for small and very small observations close to zero.

Considering the transformation of the observation error covariance, Figure 3.8 shows that the modified method of (Simon and Bertino, 2012) leads to different linear regression approximations for different realisations of the same observation. The different approximations are due to the state-dependent transformed observation error covariance which is statistically inconsistent as explained in section 3.6.3.

Estimation of the transformed observation error covariance by covariance scaling

To avoid the state dependence of the transformed observation error covariance, we suggest a new method for the estimation of the transformed observation error covariance based on a scaling approach.

Consider the sequential assimilation of a scalar observation y and the update step of the EAKF that uses an ensemble of predicted, that is prior, observations to derive observation increments and that maps these increments to state increments (section 3.2.3). The inverse observation error covariance enters the calculation of the observation increments as a weighting factor for the actual observation while the inverse estimated prior observation covariance is used as a weighting factor for the mean of the prior observation ensemble (cf. (3.36)). Thus, the impact of the observation on the observation ensemble, and consequently on the state, is governed by the ratio of the prior observation covariance to the observation error covariance. Therefore, we suggest to use this ratio calculated from the model space values to scale the transformed observation error covariance such that its ratio to the prior observation covariance in the transformed space is equal to the ratio in model space. Let $\tilde{\sigma}_o^2$ be the transformed observation error covariance, $\tilde{\sigma}_p^2$ the covariance of the transformed prior observations and let σ_o^2 and σ_p^2 be the respective covariances in model space. Then, we suggest to estimate $\tilde{\sigma}_o^2$ as

$$\tilde{\sigma}_o^2 = \sigma_o^2 \, \frac{\tilde{\sigma}_p^2}{\sigma_p^2}. \tag{3.61}$$

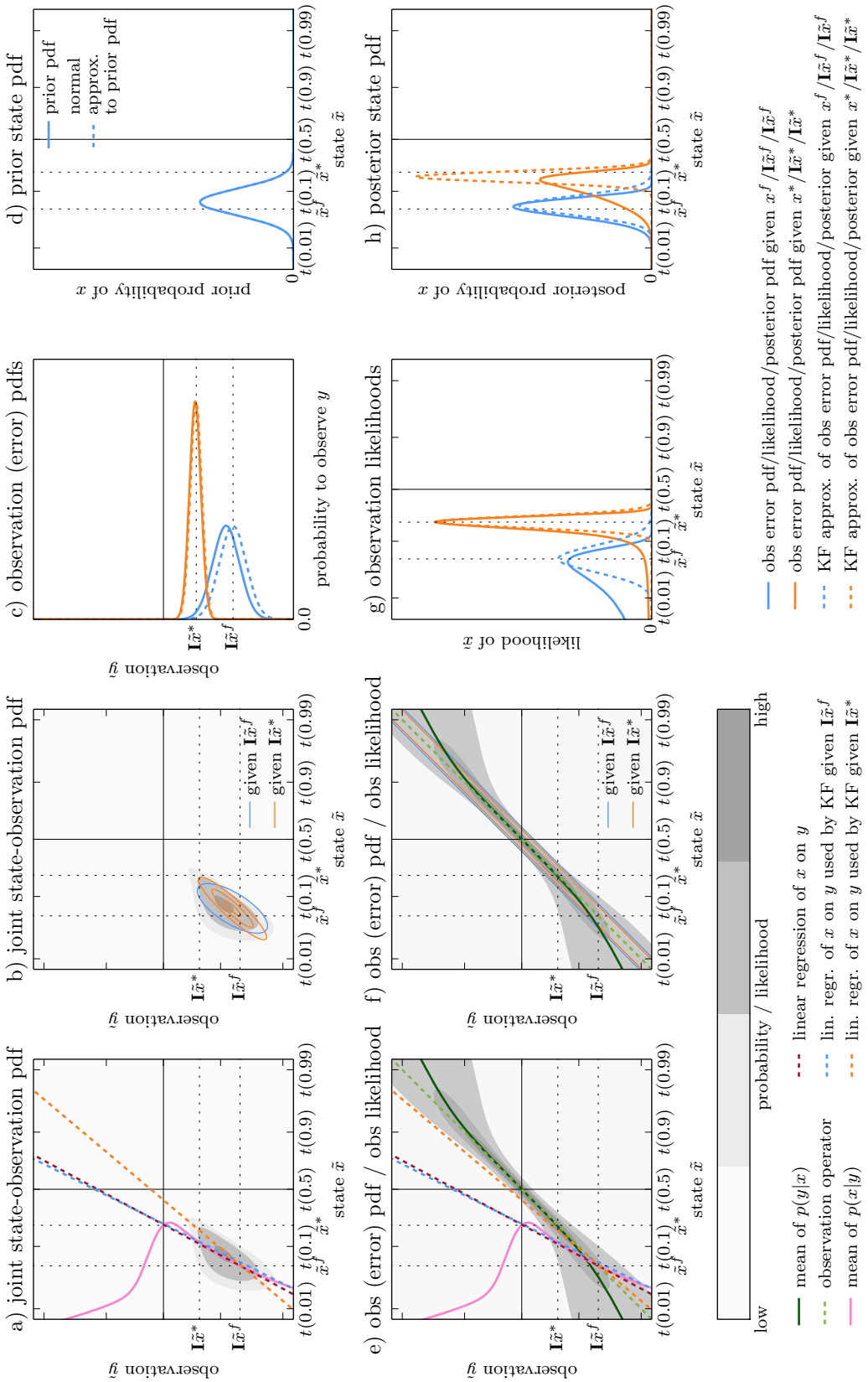


Figure 3.8: KF estimation in transformed space using the logit transform and the modified method of Simon and Bertino (2012).

Figure 3.8: The KF estimate of the linear regression for two different realisations of the same transformed observation (different random observation error) is shown as orange and blue dashed lines in panels a) and e). The true linear regression line in the transformed space is shown as purple dashed line and the curve of the conditional mean of $\tilde{x}$ given $\tilde{y}$ is shown as pink solid line. The transformed true joint pdf is shown as filled grey contours in panel a) and b) and the implicit approximations of the KF are shown as orange and blue contours. The transformed true observation pdf (function of $\tilde{y}$) and likelihood (function of $\tilde{x}$) are shown as filled grey contours in panels e) and f). The transformed observation operator is shown as light green, dashed line and the mean of the transformed observation pdf is shown as dark green, solid line. The implicit approximations of the KF to the transformed true observation pdf/true likelihood are shown as orange and blue contours in panels e) and f). The transformed true observation error pdf and the KF approximation shifted to the transformed perfect observation $\mathbf{I}\tilde{x}$ for two different transformed true states corresponding to the realisations of the transformed observation are shown in panel c). The transformed prior pdf and the KF approximation are shown in panel d). The transformed true likelihood of $\tilde{x}$ given two different realisations of the observation and the KF approximations are shown in panel g). And the transformed true posterior pdf and the KF approximation for two different transformed observations are shown in panel f).

For diagonal observation error covariance matrices, this approach can be extended to the simultaneous assimilation of several observations by

$$\tilde{\mathbf{R}} = \mathbf{R}\tilde{\mathbf{P}}^f_{\text{diag}} \left(\mathbf{P}^f_{\text{diag}}\right)^{-1},$$

where $\tilde{\mathbf{R}}$ is the transformed observation error covariance matrix, $\mathbf{R}$ is its model space equivalent, $\tilde{\mathbf{P}}^f_{\text{diag}}$ is the covariance matrix of the transformed prior observations with all off-diagonal elements set to zero and $\mathbf{P}^f_{\text{diag}}$ is its model space equivalent. Provided that $\mathbf{R}$ is independent of $\mathbf{x}$, this approach ensures that $\tilde{\mathbf{R}}$ is independent of $\tilde{\mathbf{x}}$ and that all observations get the same weight relative to the prior ensemble in the transformed space as they would get in model space.

As a result of the covariance scaling, the estimated covariance matrix $\text{cov}(\tilde{\mathbf{y}}, \tilde{\mathbf{y}})$ is independent of the actual transformed observation $\tilde{\mathbf{y}}$ as is the normal approximation of the transformed joint prior pdf. Figure 3.9 shows that, consequently, the linear regression estimated by the KF is now independent of the realisation of the transformed observation. However, the KF approximation of the linear regression is still different from the true linear regression because the covariance scaling does not yield the expected value $E(\tilde{\mathbf{R}}(\tilde{\mathbf{x}}))$ that defines the true linear regression. Further, the non-zero mean observation errors are not remedied by the variable transformation and, thus, still cause the slope of the estimated linear regression lines to be different from the slope of the true linear regression line (section 3.6.4).

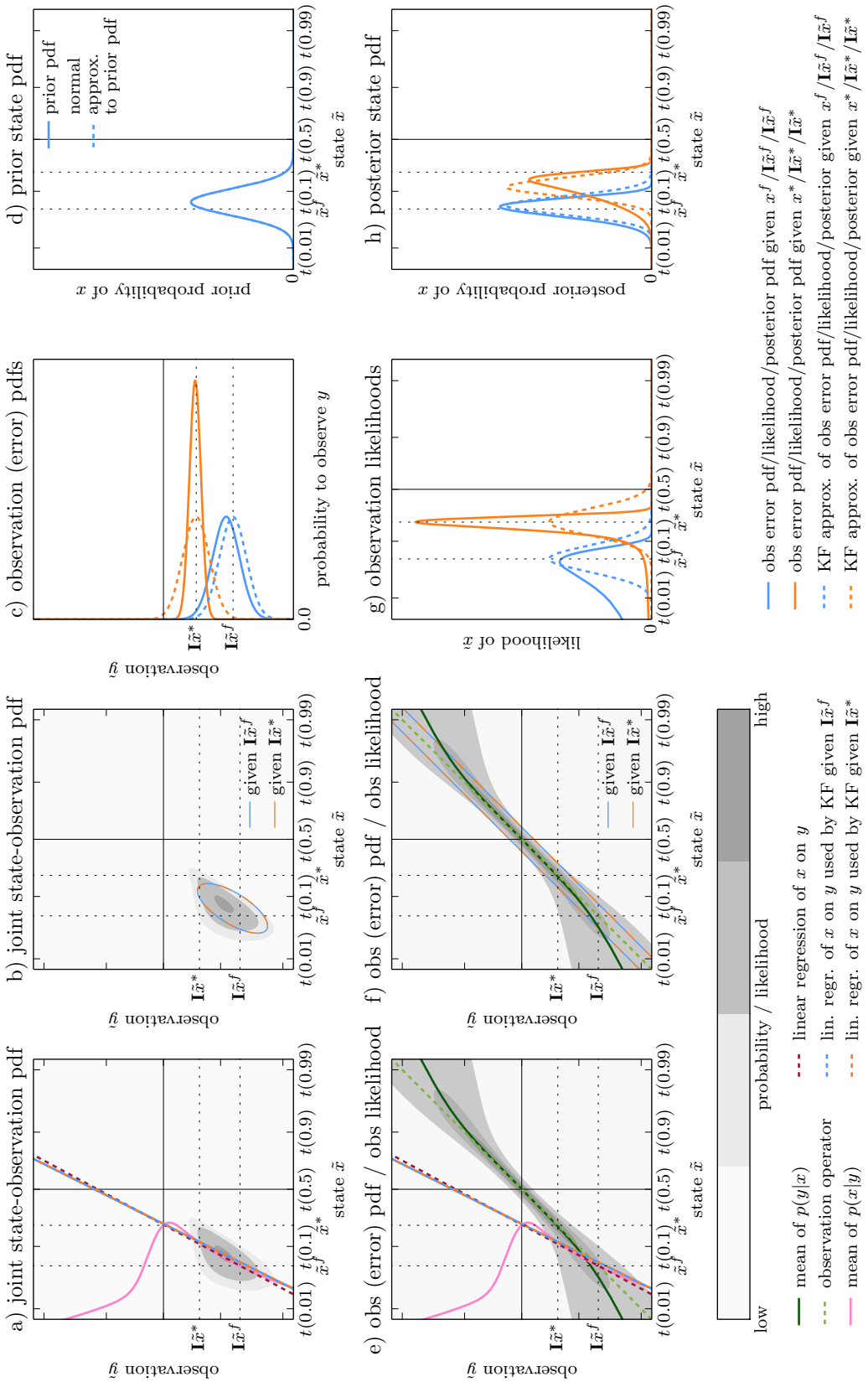


Figure 3.9: KF estimation in transformed space using the logit transform and covariance scaling (for detailed explanation see Figure 3.8).

3.7.6 Estimation of conditional pdf in model space with Gaussian anamorphosis

After the transformed observation error covariance is determined, we perform the assimilation in the transformed space. The transformed prior ensemble is updated according to the transformed observation and its transformed observation error covariance. The approximate conditional pdf in the transformed space is Gaussian with mean $\tilde{\mathbf{x}}^a$ and covariance matrix $\tilde{\mathbf{P}}^a$ given by the estimates from the updated transformed ensemble. This approximate conditional pdf in transformed space can be mapped back to model space to yield the approximate conditional pdf in model space. The mean and covariance matrix of the approximate conditional pdf in model space agree with the ensemble mean and covariance matrix of the updated model space ensemble as they would for the KF without anamorphosis but the approximate conditional pdf is a logit-normal pdf instead of a normal pdf. This logit-normal pdf is more appropriate to derive estimates of the conditional mode in model space as explained in section 3.8.

The KF with Gaussian anamorphosis adds an additional step to Algorithm 2 that maps the ensemble – as a representation of the prior pdf – to an transformed ensemble for the assimilation. And Gaussian anamorphosis modifies the approximation step of Algorithm 2 such that the approximate conditional pdf in model space results from the inverse transformation of the approximate conditional pdf in the transformed space:

Algorithm 3

1. **initialise** the forecast distribution $p(\mathbf{x}_0)$,
2. **for** i **from** 1 **to** k**:**
   a) **forecast** the prior pdf $p_{\mathbf{x}_i}(\mathbf{x}_i|\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$,
   b) **transform** the transformed prior pdf to $p_{\tilde{\mathbf{x}}_i}(\tilde{\mathbf{x}}_i|\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$,
   c) **estimate** the conditional mean and covariance matrix in transformed space using linear regression in the transformed space,
   d) **approximate** the conditional pdf $p(\mathbf{x}_i|\mathbf{y}_1, \ldots, \mathbf{y}_i)$ in model space with $\hat{p}_{\tilde{\mathbf{x}}_i}(t^{-1}(\tilde{\mathbf{x}}_i)|\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}) \left|\frac{\mathrm{d}t^{-1}(\tilde{\mathbf{x}})}{\mathrm{d}\tilde{\mathbf{x}}}\right|$, where $\hat{p}_{\tilde{\mathbf{x}}_i}$ is a normal approximation of the transformed conditional pdf.

3.7.7 Inflation and Gaussian anamorphosis

As discussed in section 3.3.1, covariance inflation is a necessity in EnKFs to avoid filter divergence. Inflation modifies the ensemble such that its spread, or covariance, increases without changing other characteristics of the ensemble. Gaussian anamorphosis and the use of a transformed ensemble next to the model space ensemble now raise the question which ensemble to inflate and how. Inflating the transformed ensemble, even without changing any other of its characteristics, changes all moments of the model space ensemble

because of the nonlinear inverse transformation, which manifests itself in the derivative term on the right hand side of (3.55). In particular, inflating the transformed ensemble shifts the location of the model space ensemble and changes its estimated conditional mode. Since our goal is to approximate the conditional mode in model space, we need to perform the inflation in a way that does not change the mode after the ensemble has been inflated. For previous applications of Gaussian anamorphosis the use of inflation is not discussed except for Lien et al. (2013) who also use a model space inflation technique. In fact, Simon and Bertino (2012) note that their parameter estimates diverge and the use of inflation still has to be investigated.

We here propose a simple additive inflation scheme that preserves the mode of the approximate conditional pdf in model space. For every ensemble member, we add a random model error term that is sampled from a shifted beta distribution. This beta distribution is chosen such that its mode is zero and its covariance equals a prescribed model error covariance. The beta distribution is shifted such that it is defined on the interval $(-x,\, 1-x)$ where x is the state value that we perturb. This shift of the beta distribution ensures a perturbed state that is physically consistent. We apply the additive inflation only after an update step and not in every model time step of the next forecast cycle (note that the assimilation does not need to take place at every model time step). Otherwise, the ensemble distribution would change towards the distribution of the sum of beta distributions and lose its logit-normal shape. The amount of inflation, that is, the prescribed covariance of the model error, is a tuning parameter of the data assimilation system (section 4.3).

An alternative inflation method that we derived from the relaxation-to-prior-spread method (Whitaker and Hamill, 2012) determines the normal distribution in transformed space that corresponds to a desired distribution in model space and shifts and scales the normally distributed ensemble in transformed space. The desired distribution in model space would be one that has the same mode as the current approximate conditional pdf but a larger covariance. Experiments with this technique led to unsatisfying results with a collapsed ensemble where only one or two members generated the desired ensemble spread.

3.8 Comparison of KF estimates for double-bounded quantities

We evaluate four methods to approximate the conditional mode and the conditional covariance of a scalar quantity x that is restricted to the interval $(0,\, 1)$ from direct observations y of that quantity. The prior pdf and the observation error pdfs are logit-normal distributions with equal covariance, the mode of all observation error pdfs is zero. The true conditional pdf is calculated from the prior pdf of x and the likelihood of x given y using Bayes' Theorem. The conditional mode and the conditional covariance are estimated from the approximate conditional pdf that results from four different applications of the KF or the EnKF assuming an infinite ensemble.

The first method represents the KF applied in model space without Gaussian anamor-

phosis. The KF approximates the prior and the observation error pdfs with normal distributions and the approximate conditional pdf is also a normal pdf. In an EnKF context, this would mean having an ensemble that contains members outside the interval (0, 1) in order to get correct estimates for the mean and covariance of the prior pdf.

The second method uses Gaussian anamorphosis to transform the variables from (0, 1) to (∞, ∞). This makes the transformed prior pdf and the transformed observation error pdfs Gaussian. But because of the state-dependence of the transformed observation error pdf, the transformed observation likelihood is non-Gaussian (cf. Figure 3.1). Applying Bayes' Theorem in the transformed space, the transformed conditional pdf is calculated from the transformed prior pdf and the transformed likelihood. We then approximate the transformed conditional pdf with a normal pdf with equal mean and covariance and this normal pdf is transformed back to model space. This method corresponds to an exact Bayesian update of the mean and the covariance of the prior observation ensemble in the transformed space. Since the prior observation ensemble is a sample of a normal distribution and only the mean and covariance are used to update the ensemble, the updated observation ensemble will also be a sample of a normal distribution. The state increments which are derived from the updated observation ensemble lead to an updated transformed state ensemble that is also sample from a normal distribution. Therefore, the normal approximation of the transformed conditional pdf is transformed back to model space and not the result of Bayes' Theorem.

The third method is the KF with Gaussian anamorphosis where the transformed observation error covariance is estimated with the modified method of Simon and Bertino (2012). And the fourth method is the KF with Gaussian anamorphosis where the transformed observation error covariance is estimated with covariance scaling (for both methods see section 3.7.5).

3.8.1 Comparison of the estimated regression curves and approximate conditional pdfs in model space

The upper two panels of Figure 3.10 show the true conditional mean regression curve and the true linear regression in model space together with the four estimated regression curves for a prior distribution with mode at 0.05 and two realisations of an observations at 0.05 and 0.2. The covariance of the prior pdf and the observation error was 0.0016, which corresponds to an observation error standard deviation of 0.04 (a comparison for prior and observation error covariances of 0.0001 (standard deviation 0.01) is shown in Appendix A).

Gaussian anamorphosis and the assimilation in the transformed space lead to nonlinear regression curves due to the nonlinear inverse transformations $t_{\mathbf{x}}^{-1}$ and $t_{\mathbf{y}}^{-1}$. These curves are also defined by only two parameters as the linear regressions (the slope and the intercept in the transformed space) but allow a much better approximation of the true conditional mean in the vicinity of the prior mode. They quickly diverge from the true

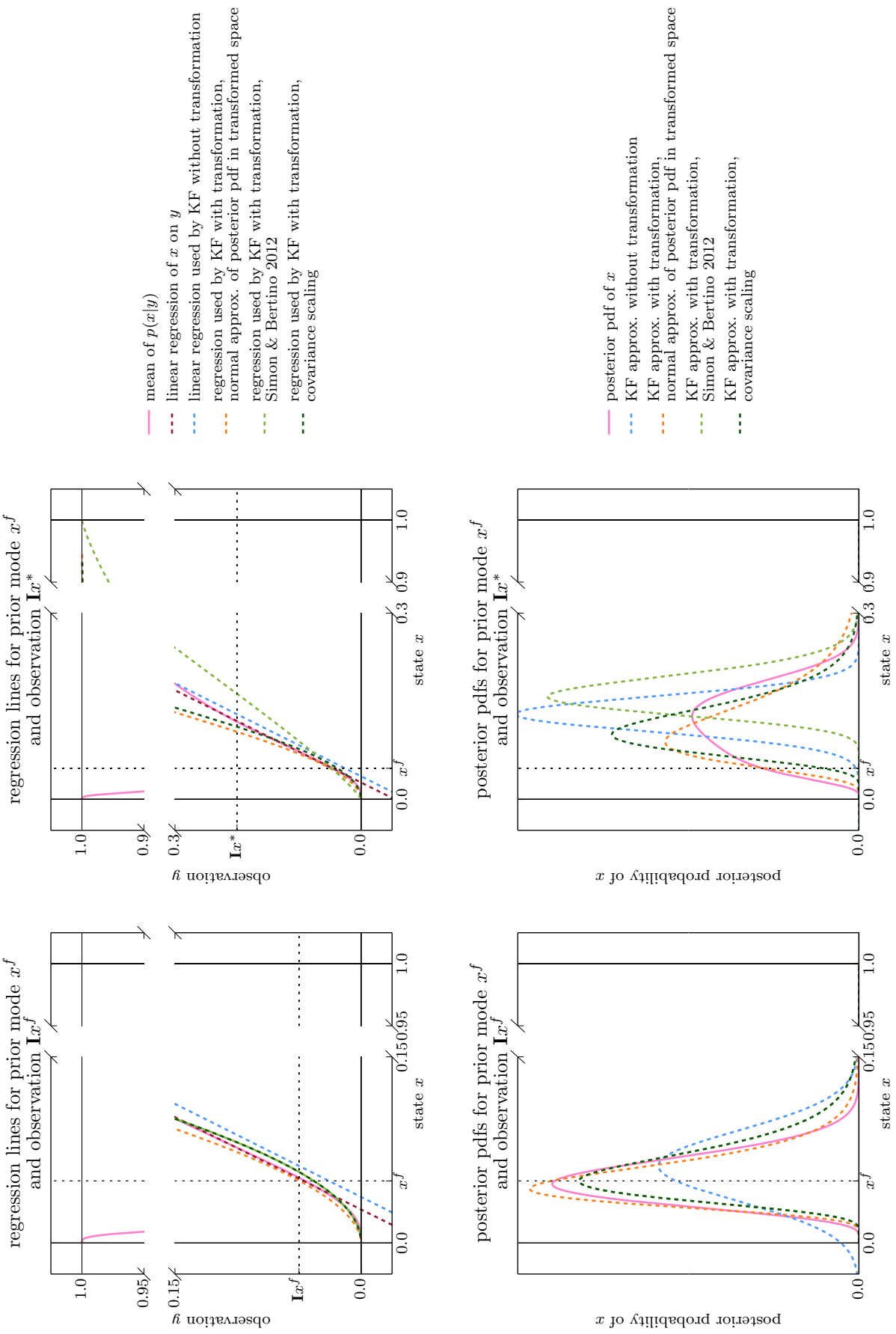


Figure 3.10: Estimated regression curves and approximate conditional pdfs.

conditional mean because the linear regression in transformed space also diverges from the transformed true conditional mean curve. The KF approximation to the linear regression in model space is biased towards the centre of the interval because of the non-zero mean observation errors.

The regression curves estimated from the modified method of Simon and Bertino (2012) and from covariance scaling agree for the first realisation of the observation $\mathbf{I}x^f$ because it is equal to the prior mode and the observation error pdf in this case is identical with the prior pdf. Consequently, the transformed pdfs and their covariances are equal such that the covariance scaling factor is one. For the second realisation, the transformed observation pdf will be different from the transformed prior pdf. Now the covariance scaling enforces the estimated linear regression in the transformed space to be equal to the estimated regression from the first realisation, while the direct calculation from the transformed observation pdf leads to a different transformed observation error covariance and, therefore, to a different approximation of the linear regression in the transformed space. Hence, the two estimated regression curves from the modified method of Simon and Bertino (2012) and covariance scaling differ for the second realisation $\mathbf{I}x^*$.

The different estimated conditional means from the different regression curves lead to different locations of the approximate conditional pdfs in model space shown in the lower two panels of Figure 3.10. The two approximate conditional pdfs from the methods using Gaussian anamorphosis are better approximations to the true conditional for the first realisation of the observation $\mathbf{I}x^f$ because the transformed true conditional pdf in this case is close to a normal pdf (Figure 3.9, panel h)). The normal approximate conditional pdf resulting from the KF without transformation is not a good approximation because it is forced to be symmetric and cannot well approximate the true conditional pdf which is skewed. For the second realisation $\mathbf{I}x^*$, the approximate conditional pdfs from the KFs with transformation are not good approximations because the observation is far from the prior mean and the linear approximation to the transformed conditional mean quickly diverges as the distance to the prior mode increases. The location of the normal approximate conditional pdf from the KF without transformation is better in this case but its covariance does not well approximate the true conditional covariance.

3.8.2 Comparison of estimated conditional modes and covariances

Figure 3.11 shows the error of the estimated conditional mode derived from the approximate conditional pdfs and Figure 3.12 shows the error of the square root of the estimated conditional variance for all possible combinations of prior mode and observation for a prior covariance and observation error covariance of 0.0016, which corresponds to an observation error standard deviation of 0.04 (a comparison for prior and observation error covariances of 0.0001 (standard deviation 0.01) is shown in Appendix A). Figure 3.13 shows the square root of the true conditional covariance, that is, the conditional standard deviation, and serves as a reference for the errors in the estimated conditional mode and

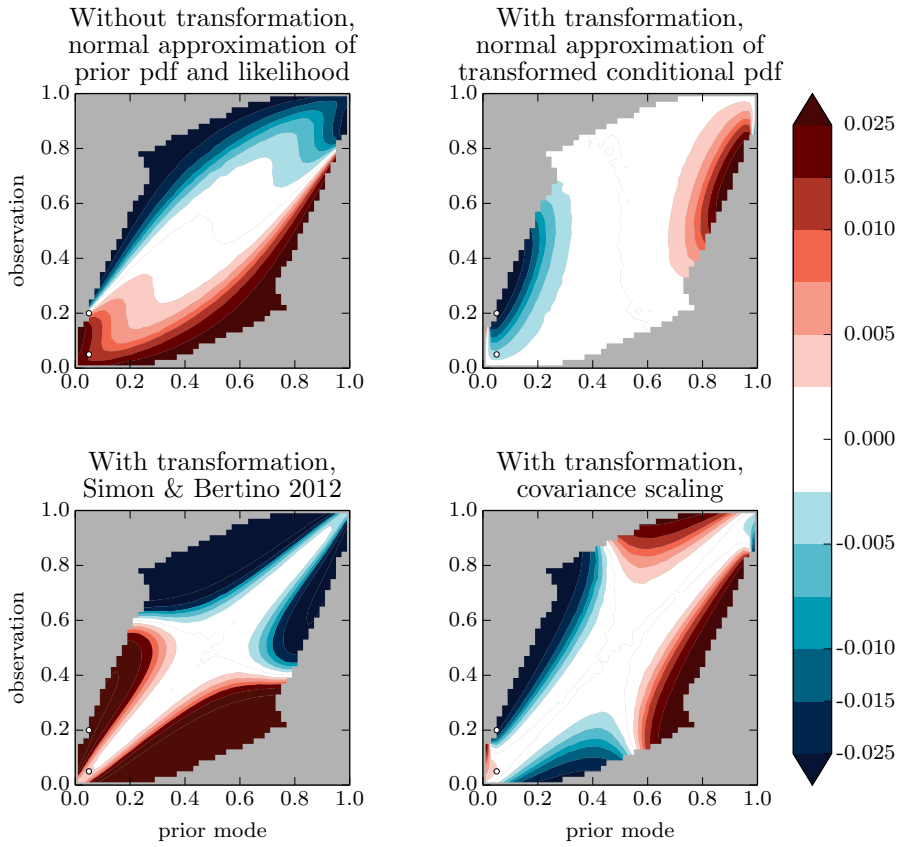


Figure 3.11: Error in the estimated conditional mode as a function of the mode of the prior distribution and the observation. The prior distribution has a covariance that is equal to the covariance of the observation error of 0.0016 (standard deviation 0.04). Grey areas indicate a bimodal conditional pdf. White dots indicate the position of the examples in Figure 3.10.

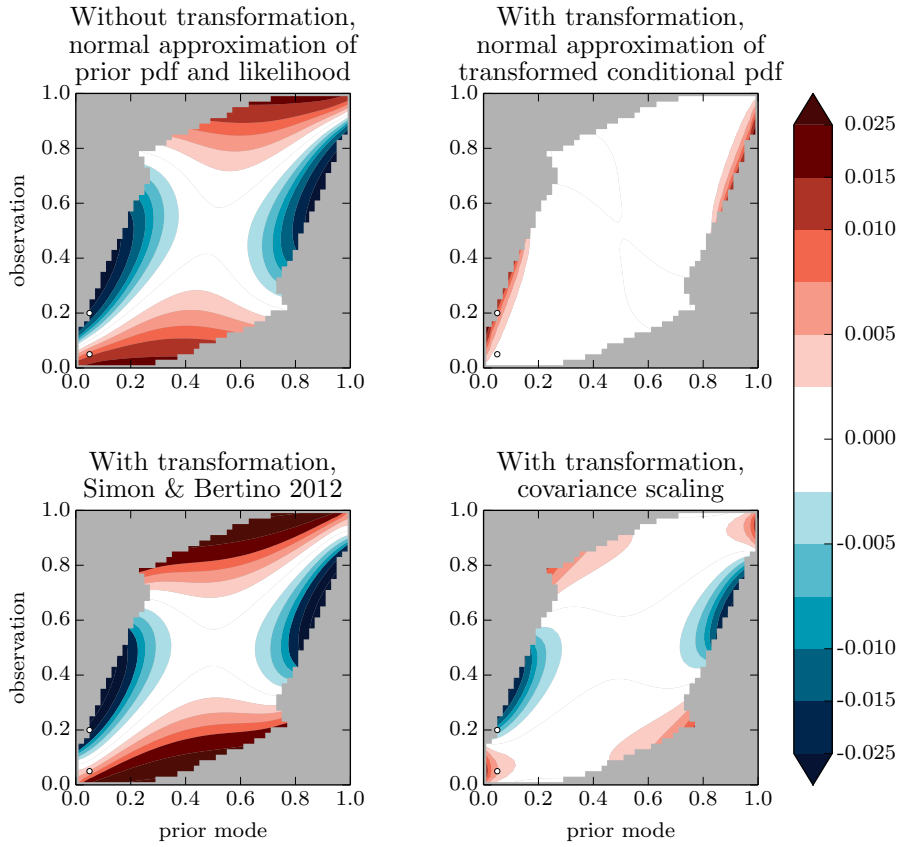


Figure 3.12: Error in the square root of the estimated conditional variance as a function of the mode of the prior distribution and the observation. The prior distribution has a covariance that is equal to the covariance of the observation error of 0.0016 (standard deviation 0.04). Grey areas indicate a bimodal conditional pdf. White dots indicate the position of the examples in Figure 3.10.

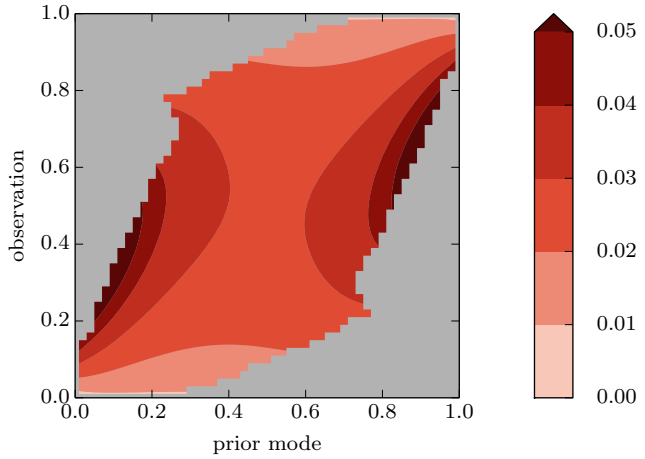


Figure 3.13: Square root of the true conditional covariance.

covariance. The conditional standard deviation approximately represents the width of the conditional pdf and errors in the conditional mode have to be judged in relation to this width. Consequently even the small magnitude of the errors in the conditional mode is a sign of substantial errors. The grey areas indicate bimodal conditional pdfs which are not discussed.

The quality of the estimates of all four methods deteriorates as the prior mode approaches the bounds of the interval. This due the normal approximation of the conditional pdf for the KF without transformation. For the two methods that estimate the transformed observation error covariance, the errors originate in the normal approximation of the transformed conditional pdf and the nonlinear inverse transformation. The normal approximation of the true transformed conditional pdf has the correct mean and covariance and the errors in the estimates from this method are only due to the nonlinear inverse transformation. Both sources of error exacerbate as the prior mode approaches the bound of the interval because the normal approximation becomes less valid in model space as well as in transformed space and the nonlinearity of the transformation increases.

The KF without transformation causes biases in the estimated conditional mode for all combinations of prior mode and observation that are not at the centre of the joint domain. The KF with transformation where we approximate the true transformed conditional pdf with a normal pdf shows a large tolerance to differences in the prior mode and the observation if the prior mode is between 0.25 and 0.75. But this method causes biases even when prior mode and observation agree when the bound is approached. The modified method of Simon and Bertino (2012) and covariance scaling show a generally similar behaviour with opposite sign for the errors. Both yield good estimates if the observation is close to the prior mode but the covariance scaling is more tolerant to differences between the prior mode and the observation.

The comparison yields qualitatively similar results for a prior and observation error covariance of 0.0001 (standard deviation 0.01) in model space, except for the KF without transformation. This approach performs better for the small covariances because the pdfs are extremely narrow and the non-Gaussianity only matters in a very small region close to the interval bound.

3.8.3 Comparison of the approximating joint pdfs and observation error pdfs in model space

For the four compared methods, Figure 3.14 shows the approximating joint pdfs in model space (upper row) and the approximating likelihoods and observation error pdfs in model space, respectively as a function of x or as a function of y (lower row). The approximating joint pdf of the KF without transformation does not approximate the true joint pdf well. Most notably, it extends beyond the bounds of the domain $(0,\ 1)\times(0,\ 1)$ and the contours of the approximate pdf narrow where the contours of the true pdf broaden. For the modified method of Simon and Bertino (2012), the quality of the approximating joint pdf depends on

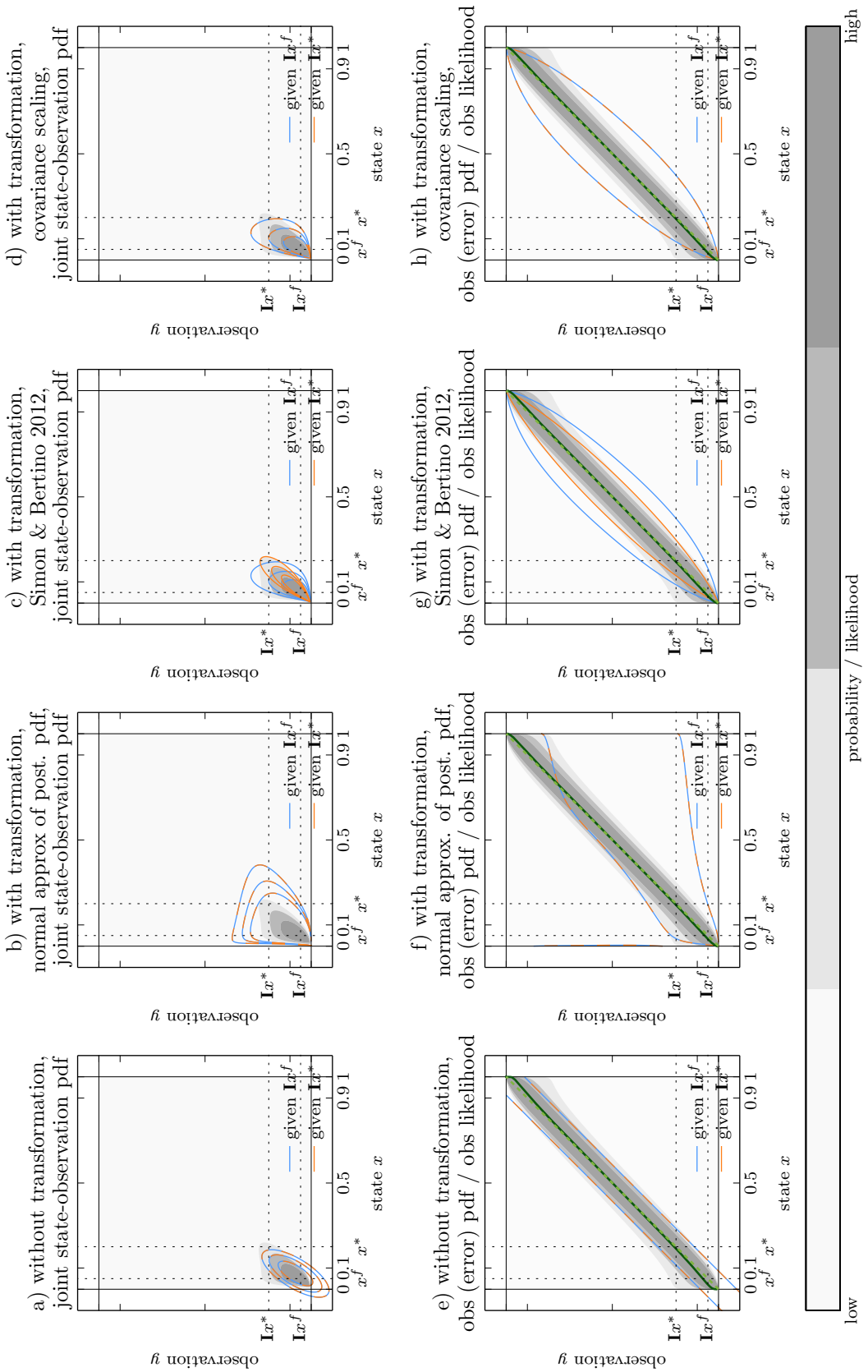


Figure 3.14: Joint pdf of x and y and approximations used by the KF in model space (upper row) and observation pdfs (function of y for fixed x) and observations likelihood (function of x for fixed y) and approximations used by the KF (lower row) for two realisations $\mathbf{I}x^f$ and $\mathbf{I}x^*$ of an observation.

the realisation of the observation. The approximating joint pdfs of the covariance scaling approach and the approach with a normal approximation of the transformed conditional pdf capture the broadening shape of the true joint pdf well. This is in agreement with Figures 3.11 and 3.12 where these two approaches showed the best performance.

Regarding the observation error pdfs, the modified method of Simon and Bertino (2012) and covariance scaling lead to an approximation with symmetrical observation error distributions whose covariance decreases as the observation approaches the bounds of the interval. The most important difference, again, is that the implicit approximations of the observation pdf and the observation likelihood, respectively, are independent of the realisation of the observation for the covariance scaling approach, as opposed to the modified method of Simon and Bertino (2012).

We note here, that the assumption of normal observation error distributions with constant covariance in transformed space as it is made in the covariance scaling approach corresponds to an approximation of the true observation errors in model space with relative observation errors, where the error magnitude scales with increasing distance from the bounds of the interval. This explains the bent shape and the congruence of the blue and orange contours in panel h) of Figure 3.14.

3.8.4 Summary and discussion of KF estimates for double-bounded quantities

We interpret the KF as an approximate version of the sequential filtering algorithm (Algorithm 1 in section 3.1) that yields an approximate conditional pdf of the state vector $\mathbf{x}$ given the observations $\mathbf{y}$. The estimated conditional pdf is a good approximation of the true conditional pdf as long as the prior and observation error distributions are approximately Gaussian. Double-bounded quantities imply non-Gaussian distributions for the prior state and, in case of direct observations, for the observation error. These non-Gaussian distributions cause the joint pdf of $\mathbf{x}$ and $\mathbf{y}$ to be non-Gaussian and make the conditional mean of the state $\mathbf{x}$ given the observations $\mathbf{y}$ a nonlinear function of $\mathbf{y}$.

The KF uses a linear regression approach to estimate the conditional mean and the conditional covariance and, subsequently, approximates the conditional pdf with a normal pdf. The estimation of the linear regression, however, is susceptible to non-zero mean and state-dependent observation errors. The quality of the approximation of the conditional mean can be improved with a nonlinear regression that results from the KF in conjunction with Gaussian anamorphosis. This, however, requires the estimation of the observation error covariance in transformed space which is, in general, state-dependent.

We compare four methods, one without and three with Gaussian anamorphosis, to estimate the mode of the non-Gaussian conditional pdf of a state x given an observation y, where both x and y are restricted to the bounded interval (0, 1). The KF without Gaussian anamorphosis is sub-optimal because the normal approximation of the conditional pdf in model space is inadequate for bounded quantities, in particular close to the bounds of the interval. The estimates of the conditional mode from this approach are biased towards the

centre of the interval for observations that approximately agree with the prior mode. The KF without Gaussian anamorphosis only yields acceptable estimates if the prior mode and the observations are close to the centre of the interval because, in this case, the prior distribution and the observation error distributions are close to Gaussian distributions. Moreover, without Gaussian anamorphosis, physically inconsistent estimates that have to be manually corrected may occur. The transformation of the state by an anamorphosis function avoids this problem (section 3.7.1).

The second method uses Gaussian anamorphosis, exactly calculates the transformed conditional pdf using Bayes' Theorem from the Gaussian transformed prior pdf and the non-Gaussian transformed observation likelihood, and approximates the transformed conditional pdf with a normal pdf with equal mean and covariance. This approach is computationally expensive because the application of Bayes' Theorem requires the construction of the transformed likelihood, the point-wise multiplication of the transformed prior pdf and the transformed likelihood, and numerical integrations to calculate the mean and the covariance of the transformed conditional pdf. This approach performs well if the prior mode is close to the centre of the interval, independent of the observation. But, as the prior mode approaches the bounds, the estimates of the conditional mode are biased towards the bounds of the interval even for observations that are consistent with the prior mode.

The third method calculates the transformed observation error covariance from an approximation of the transformed observation error pdf (modified method of Simon and Bertino, 2012). Finding the transformed pdf, however, is numerically expensive (section 3.7.4). This approach performs well in terms of the estimated conditional mode and the estimated conditional covariance in model space over the whole interval, given that the observation is close to the prior mode. When the prior mode or the observation approaches the bounds, the estimates of the conditional mode are biased towards the centre of the interval and the quality of the estimates of the conditional covariance also deteriorates.

The fourth method uses the new covariance scaling technique proposed in this thesis to approximate the transformed observation error covariance. This method performs similar to the modified method of Simon and Bertino (2012). But the range in which the prior mode and the observation may differ to still yield acceptable estimates is much larger than for the modified method of Simon and Bertino (2012). When the prior mode or the observation approaches the bounds, the estimates of the conditional mode are biased towards the bounds of the interval while the estimates of the conditional covariance remain acceptable.

Assuming unbiased observations, the most frequent combinations of prior mode and observations are these around the one-to-one line in Figures 3.11 and 3.12. Thus, small estimation errors in this area are particularly important. In this respect, the method of approximating the true transformed conditional pdf with a normal pdf, the modified method of Simon and Bertino (2012), and covariance scaling perform similar. The approximation of the true transformed conditional pdf with a normal pdf, however, is prohibitively expen-

sive. When comparing the modified method of Simon and Bertino (2012) and covariance scaling, covariance scaling is easier to implement and less costly. More important, however, is that covariance scaling uses the same statistical approximations independent of the observed value while the estimation process itself of the modified method of Simon and Bertino (2012) is sensitive to the observation.

Chapter 4

Data assimilation experiments with synthetic observations

4.1 A sequential data assimilation framework for JSBACH

Motivated by the need for a climatology of JSBACH canopy albedo parameters (section 2.3), we set up a flexible sequential data assimilation framework for JSBACH based on the Data Assimilation Research Testbed (DART; Anderson et al., 2009). DART uses a parallel implementation of the EAKF to update the state vector from a sequence of scalar observations (Anderson and Collins, 2007). For the assimilation experiments with DART, we map the visible and near-infrared grid box albedos and the visible and near-infrared canopy albedo parameters to the DART state vector and assimilate scalar observations of visible and near-infrared grid box albedo. We integrated JSBACH into DART with full restart capabilities to be able to perform longer assimilation experiments with computationally expensive update algorithms as described in section 3.8.

4.1.1 Model setup and forcing

We use the offline version of JSBACH with a time step of 30 minutes on a T63 Gaussian grid, corresponding to a resolution of $1.875^\circ \times 1.875^\circ$ at the equator (Dalmonech and Zaehle, 2013). The model forcing consists of 6 years of daily data for surface wind speed, shortwave and longwave incoming radiation, precipitation, and minimum and maximum air temperature. To generate the forcing, we conservatively remapped ERA-Interim reanalysis data for the years 2005 to 2010 (Dee et al., 2011). To correct for errors in ERA-Interim precipitation values, we rescaled these values such that their monthly means match the monthly values from the Global Precipitation Climatology Project (GPCP; Huffman et al., 2009) according to Balsamo et al. (2010). Further, CO_2 forcing is taken from the RCP 4.5 scenario and rises from 379 ppm in 2005 to 388 ppm in 2010 (Moss et al., 2010).

We run JSBACH with one tile per grid box and with a constant spatial distribution of cover types as shown in Figure 4.1. The vegetated fraction for each grid box is also constant and is shown in Figure 4.2. We use different PFTs for the northern hemisphere (NH) and the southern hemisphere (SH) because of the assumed opposing seasonal cycles of the

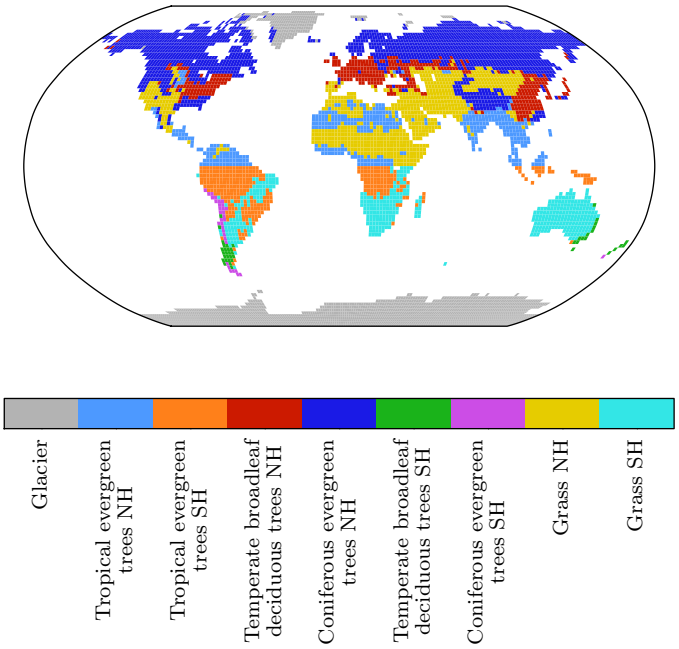


Figure 4.1: Spatial distribution of cover types (PFTs).

canopy albedo parameters (see section 2.3). Otherwise, NH and SH PFTs are identical. The visible and near-infrared canopy albedo parameters are prescribed as constant in time for one set of experiments. In this case, the parameters are equal for NH and SH PFTs. In a second set of experiments, we prescribe a seasonal cycle for the parameters. The values for the constant parameters in the first set of experiments and the prescribed seasonal cycles for the second set of experiments are shown in Figure 4.3. We selected the PFTs such that they exhibit different values for LAI and thus for canopy cover fraction. And we included evergreen and deciduous vegetation types to simulate different seasonal cycles of LAI and canopy cover fractions (Figures 4.4 and 4.5).

4.1.2 Extensions of the Data Assimilation Research Testbed

We integrated JSBACH into DART such that they run as single executable. DART uses a predefined model interface to which we coupled JSBACH such that DART controls the advancement of the model on a model time step basis. Between observations, DART repeatedly advances an ensemble of model states until the next observation time is reached. At this point, the model state is mapped to the DART state vector and DART performs the assimilation using the EAKF. Subsequently, the updated DART state vector is mapped back to a model state and the next forecast cycle starts.

We extended DART with the option to transform the elements of the model state vector with the logit function when mapping them to the DART state vector (section 3.7). And we extended the implementation of the update step of the EAKF to use any of the four methods explained in section 3.8, those are the KF without transformation, the

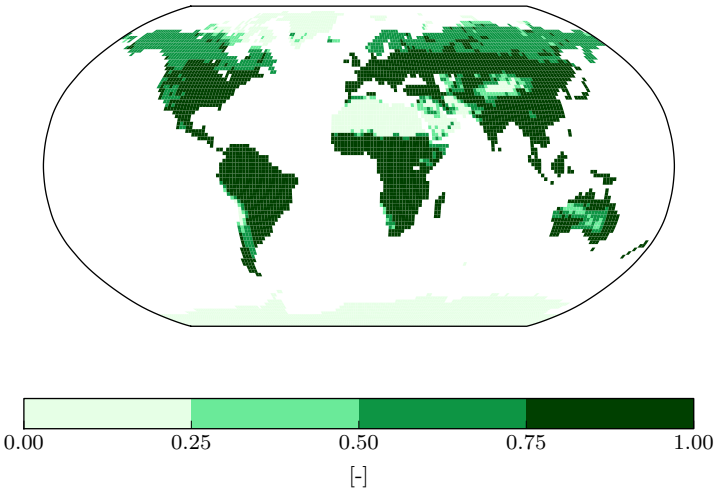


Figure 4.2: Vegetated fraction V_{max} of model grid boxes.

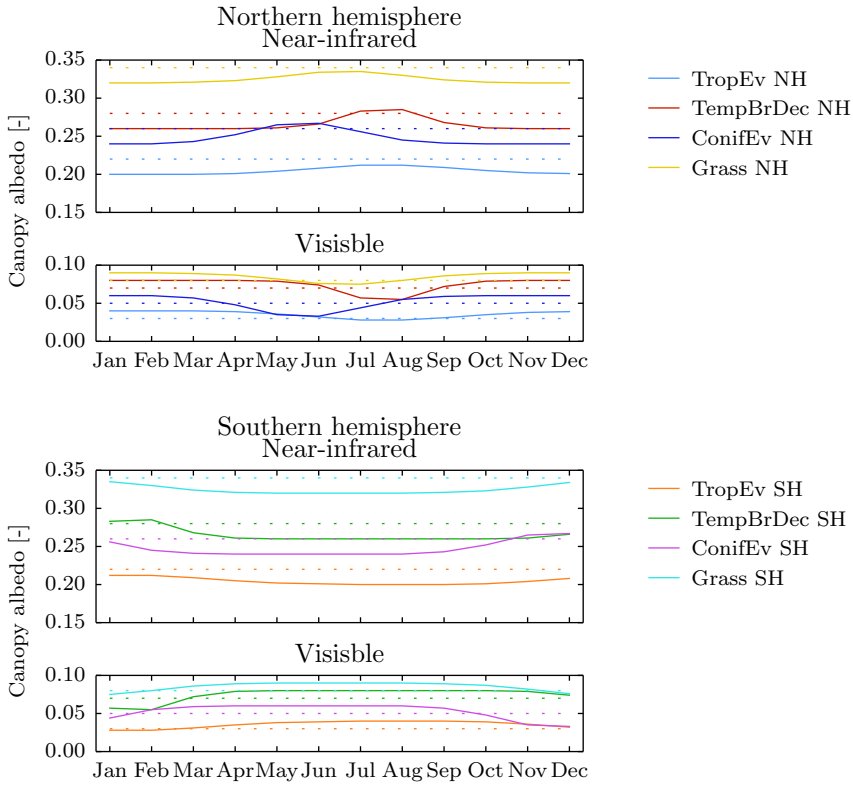


Figure 4.3: Prescribed constant (dashed) and seasonal (solid) canopy albedo parameters.

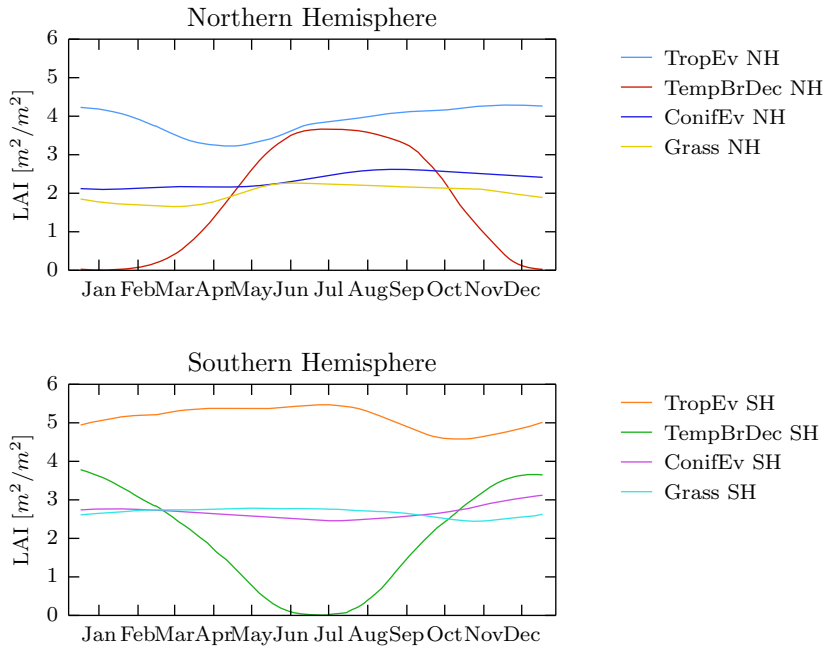


Figure 4.4: Mean seasonal cycle of LAI, averaged of grid boxes that are covered by the same PFT.

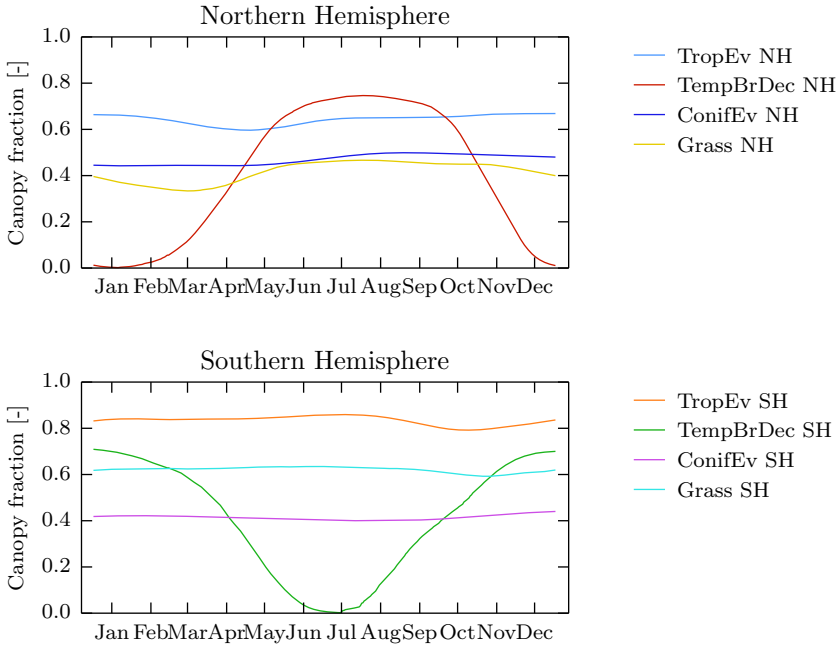


Figure 4.5: Mean seasonal cycle of canopy fraction f_c, averaged of grid boxes that are covered by the same PFT.

approximation of the transformed conditional pdf with a normal pdf, the modified method of Simon and Bertino, 2012, and covariance scaling. For the KF without transformation, we ensure physically consistent values by setting updated states and parameters to zero or one if they are less than zero or greater than one, respectively.

For the approximation of the transformed conditional pdf with a normal pdf, we calculate the transformed conditional pdf by pointwise multiplication of the transformed prior pdf and the transformed observation likelihood according to Bayes' Theorem. To calculate the mean and covariance of this pdf, we added a numerical integration scheme to the update step. We construct the required observation likelihood from the observation pdfs as described in section 3.1 (cf. Figure 3.1). All these calculations are done with univariate pdfs because DART assimilates observations sequentially.

The modified method of Simon and Bertino (2012) requires the covariance of the transformed observation pdf. We assume logit-normal observation error distributions in model space and use pre-calculated look-up tables as described in section 3.7.4 to retrieve the mean and covariance of this pdf from the observed value and its prescribed observation error covariance.

For the covariance scaling, we calculate the ensemble covariance in the transformed space and in model space. Together with the prescribed observation error covariance in model space, we use these two covariance estimates to calculate the transformed observation error covariance according to (3.61).

Further we added an option for additive inflation in model space to DART. After the update step, the ensemble is transformed back to model space. Before we start the next forecast cycle, we add a random error term to each canopy albedo parameter for each ensemble member. The random error term is drawn from a beta distribution as described in section 3.7.7. The parameters of the beta distribution are found from the equations for the mode and the variance of the beta distribution (Johnson and Kotz, 1970).

4.2 Setup of assimilation experiments

We performed assimilation experiments with synthetic observations generated from a control run of JSBACH that represents a virtual truth. The use of synthetic observations generated from a virtual truth allows us to analyse the errors of the estimated parameters and to draw conclusions for the assimilation of real observations.

4.2.1 Generation of initial ensembles

We generate initial ensembles with 64 members by perturbing the canopy albedo parameters of the 8 PFTs shown in Figure 4.1. For the initial uncertainty, we assume a variance of 0.0025 (standard deviation of 0.05) for the parameter distribution in model space. We also shift the mode of the initial parameter distribution in model space randomly from the true parameter value. We calculate the transformed, normal distribution that corresponds

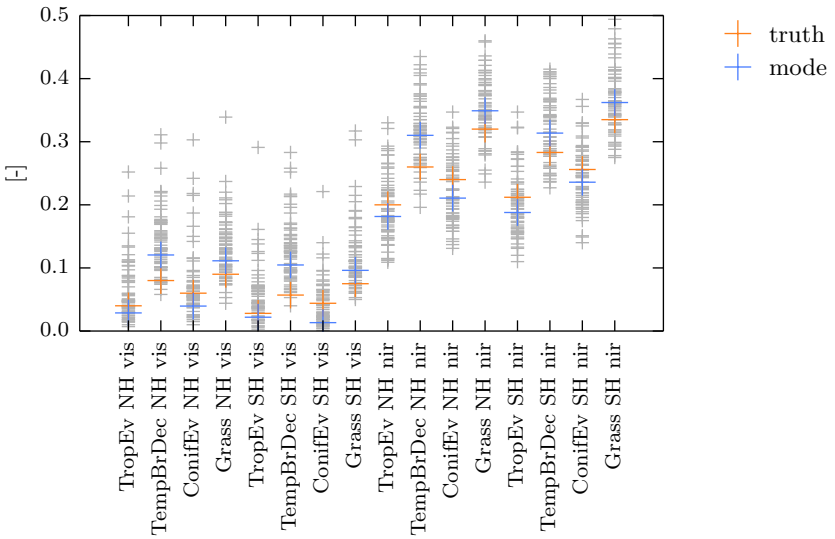


Figure 4.6: Initial ensembles of visible (left) and near-infrared (right) canopy albedo parameters. Ensemble members are shown in grey, the virtual true value in orange and the mode of the initial ensemble in blue.

to the shifted mode and the assumed variance, sample the 64 parameter values from that normal distribution, and transform the ensemble back to model space. Figure 4.6 shows the initial ensembles for the canopy albedo parameters generated in this fashion. The initial distributions for the visible parameters, which are very close to zero, are highly skewed with a sharp peak and a long tail. This shape of the pdf causes the outliers and the clustering on the lower end of the visible parameter ensembles. The distributions for the near-infrared parameters, on the other hand, are more symmetrical. This is reflected by the more symmetrical distribution of the near-infrared ensemble members around the mode.

Finally, we use the generated parameter ensembles to simulate an ensemble of initial model states in a one-year model spin-up for each ensemble member. This one-year simulation uses the given parameters as constant parameters without a seasonal cycle. The assimilation then starts from these 64 initial model states.

4.2.2 Generation of synthetic observations

In the assimilation experiments we use synthetic observations which we generate from a virtual truth. This is particularly important for the estimation of effective model parameters because the virtual truth serves as a direct reference for the evaluation of the estimated parameters. Using real observations, the estimated parameters would have to be used in an additional simulation step to predict observations which could then be compared to independent observations for the validation of the estimated parameters. But this introduces additional sources of error. The assimilation of synthetic observations in a twin experiment is therefore an essential step for the evaluation of an assimilation framework.

We generate the synthetic truth with a five-year control run of JSBACH that starts after a one-year spin up. We prescribe constant canopy albedo parameters for one set of experiments and seasonal canopy albedo parameters for another set as shown in Figure 4.3. From the model state of the control run, we extract visible and near-infrared grid-box albedo values every eight days and add a random error to generate the synthetic observations. The added observation error is sampled from a shifted beta distribution with mode zero and a prescribed variance. The distribution is shifted such that the perturbed observation lies in the interval (0, 1), similar to the model error distribution described in section 3.7.7. For the assimilation experiments, we generated observations with an observation error covariance of 0.0016 (standard deviation 0.04) and 0.0001 (standard deviation 0.01).

The chosen observation error covariances and observation frequency correspond to the order of magnitude of the errors and the observation frequency of land surface albedo observations from MODIS (Liu et al., 2009). We note, however, that the scale of the observations differs significantly (500 m for MODIS, approximately 200 km at the equator for the T63 grid).

4.3 Data assimilation experiments

We compare the results of data assimilation experiments for prescribed constant and seasonal canopy albedo parameters. The model forcing in all assimilation experiments is the same forcing that we used for the generation of the synthetic observations. In all experiments we return only the updated parameters to the model for the next forecast cycle. To update the parameters, we use the four methods described in section 3.8 (KF without transformation, approximation of the transformed conditional pdf with a normal pdf, the modified method of Simon and Bertino, 2012, and covariance scaling).

We compare the assimilation results for the visible and near-infrared canopy albedo parameters of JSBACH. The estimates of the conditional mode and the conditional covariance in model space are calculated directly from the univariate approximate conditional pdf in model space. The mode is given by the location of the maximum of that pdf and the covariance is found from the integral definition of the covariance in (3.60). We derive the approximate conditional pdf in model space from the normal pdf in the transformed space that is given by the mean and the covariance of the transformed ensemble of parameter values (see section 3.7.6).

For both, constant and seasonal canopy albedo parameters, we first compare the effects of different magnitudes of inflation followed by the comparison of the four update methods.

4.3.1 Experiments with constant canopy albedo parameters

Figure 4.7 shows the conditional mode and the ensemble spread (given as the square root of the conditional covariance) in model space for experiments without inflation, with an added

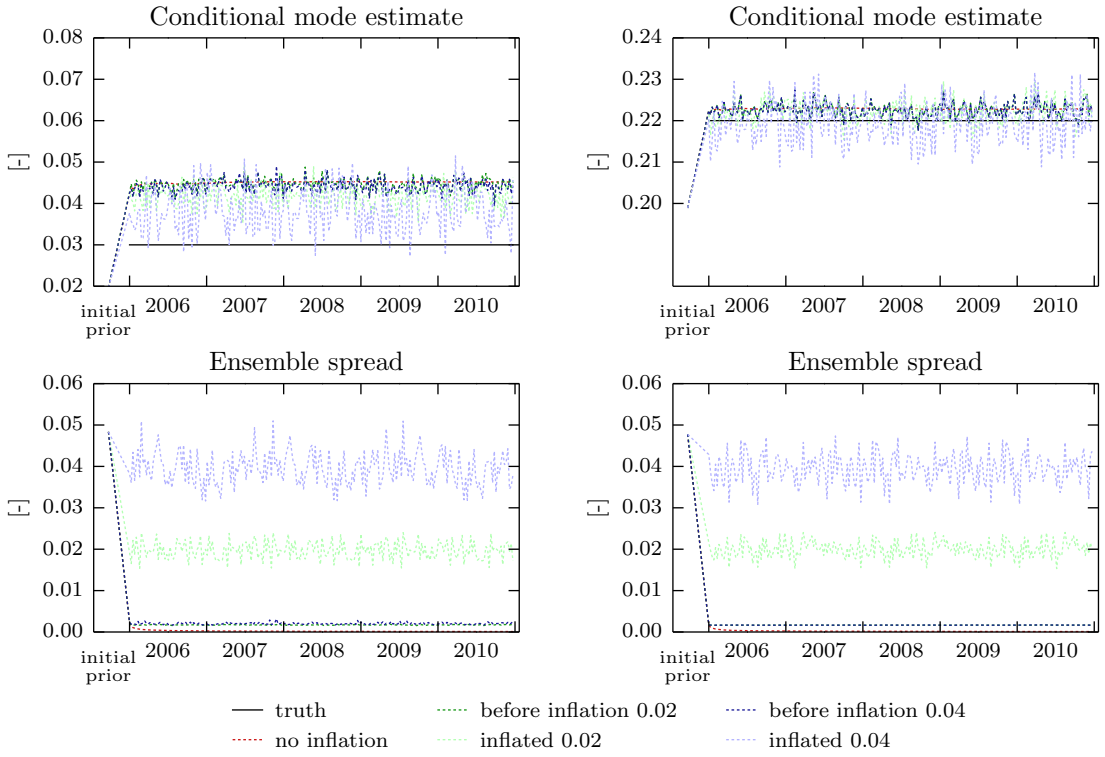


Figure 4.7: Assimilation results for one fixed canopy albedo parameter in the visible (left) and the near-infrared (right) domain. Dark-coloured lines show the results before inflation and the light-coloured lines show the results after inflation.

random model error that has covariance 0.0004 (standard deviation 0.02), and with an added random model error with covariance 0.0016 (standard deviation 0.04). We refer to the standard deviation of the added random model error as inflation magnitude. Without inflation, the ensemble collapses to nearly zero spread during the first few assimilation steps. The estimate of the conditional mode without inflation, however, is equal to the estimates with inflation (within their random variations).

Using inflation, the ensemble spread is increased to the inflation magnitude after every update step (difference between dark- and light-coloured lines in Figure 4.7). The estimated conditional mode varies randomly around a constant value. The variations are small and the estimates before inflation approximately agree with the estimate without inflation. After the inflation, the variability of the estimated mode of the inflated ensembles increases and the estimated mode is on average smaller than before the inflation. This applies in particular for the parameter in the visible domain that is much closer to zero. The magnitude of the variations in the estimated mode and of the difference between the estimates before and after inflation depends on the inflation magnitude. A larger inflation magnitude leads to larger variations and a larger difference. The results for other vegetation types than the one shown in Figure 4.7 are qualitatively similar and are summarised in Figure 4.8. The experiments in the subsequent comparison of the four methods used

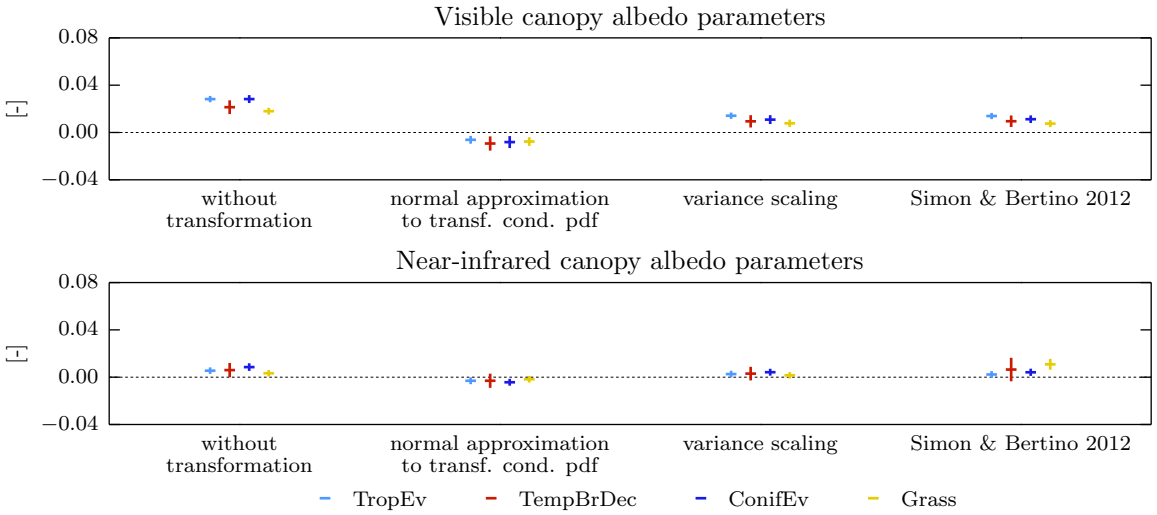


Figure 4.8: Bias and error variability of the conditional mode for the estimation of fixed canopy albedo parameters. Horizontal lines indicate the mean error, vertical bars above and below indicate one standard deviation of the errors.

	RMSE	bias	std of errors
without transformation	0.025	0.015	0.010
normal approx. to cond. pdf	0.010	-0.005	0.004
covariance scaling	0.012	0.007	0.005
Simon and Bertino (2012)	0.014	0.008	0.005

Table 4.1: Overall root mean square error (RMSE), bias and standard deviation of errors for the estimation of fixed canopy albedo parameters.

an inflation magnitude of 0.04. Figure 4.8 shows the mean error (bias) and the standard deviations of the errors for the estimated visible and near-infrared canopy albedo parameters for the four update methods. Table 4.1 gives the combined values over all parameters for root mean square error, bias, and standard deviation of the errors. The errors have a consistent pattern in the visible and the near-infrared domain. For both domains, the normal approximation of the transformed conditional pdf leads to an underestimation of the conditional mode while the other three methods overestimate the parameter.

In the visible domain, the KF without transformation causes the largest absolute errors and has the largest variability in the errors. The other three methods yield almost equal results, with the normal approximation of the transformed conditional pdf having marginally smaller absolute errors. In the near-infrared domain the results are similar although with an overall smaller magnitude of the errors.

4.3.2 Experiments with seasonal canopy albedo parameters

Figure 4.9 shows the conditional mode and the ensemble spread in model space for experiments without inflation and for inflation with magnitudes 0.02 and 0.04 for the estimation

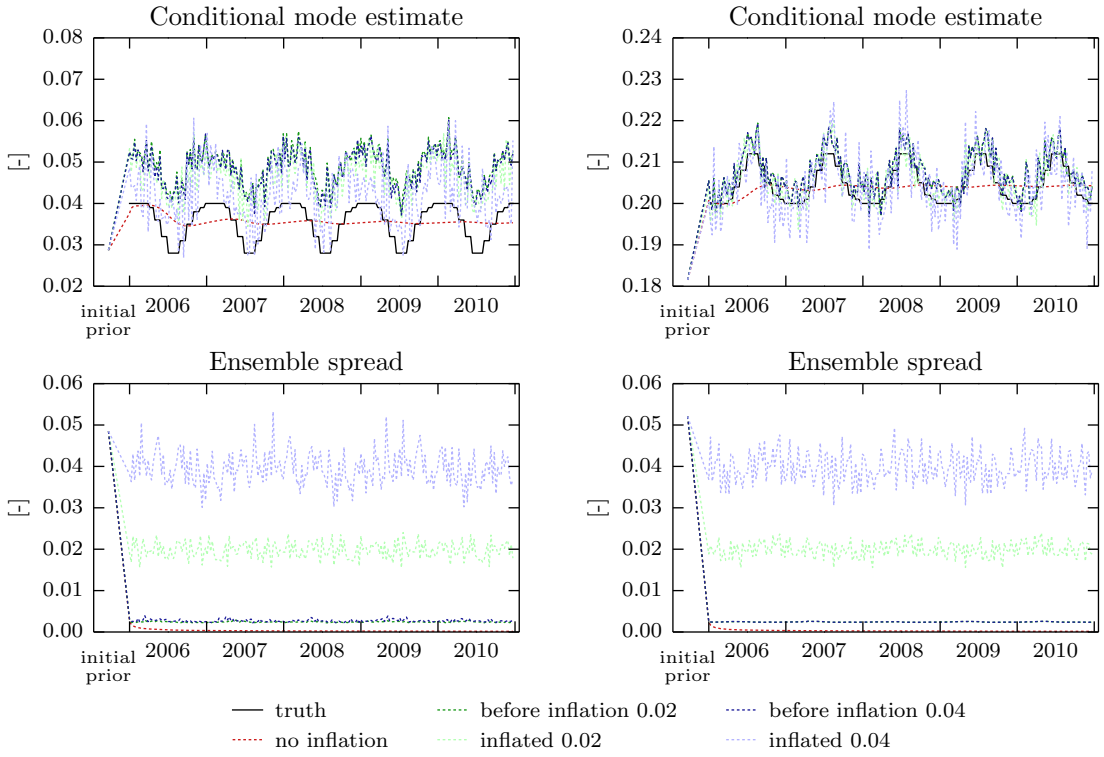


Figure 4.9: Assimilation results for one seasonal canopy albedo parameter in the visible (left) and the near-infrared (right) domain. Dark-coloured lines show the results before inflation and the light-coloured lines show the results after inflation.

of a seasonal canopy albedo parameter. Without inflation, the ensemble collapses during the first few assimilation steps. As a result, the conditional mode estimate diverges and cannot follow the seasonality of the parameter. The estimates from experiments with inflation follow the seasonal cycle of the parameter. The effects of the inflation are the same as described for the estimation of a fixed parameter in the previous section.

As before, the experiments in the subsequent comparison of the four methods used an inflation magnitude of 0.04. Figure 4.10 illustrates the effects of the different update methods on the estimates of the conditional mean and the conditional covariance in model space. First, all four methods yield estimates that follow the prescribed seasonal cycle up to random variations. But all four methods are shifted by a constant value from the true parameter value. The results are qualitatively the same in the visible and near-infrared domain, although the estimates for the near-infrared parameter are much closer to each other. In both domains, the KF without transformation yields the largest estimates of the conditional mode and the normal approximation to the transformed conditional pdf yields the smallest estimates. The estimates of the covariance scaling method and the modified method of Simon and Bertino (2012) are nearly identical and lie in between the other two methods. The estimated conditional covariance, shown by the ensemble spread, of the method based on Simon and Bertino (2012) is larger than for the other three methods.

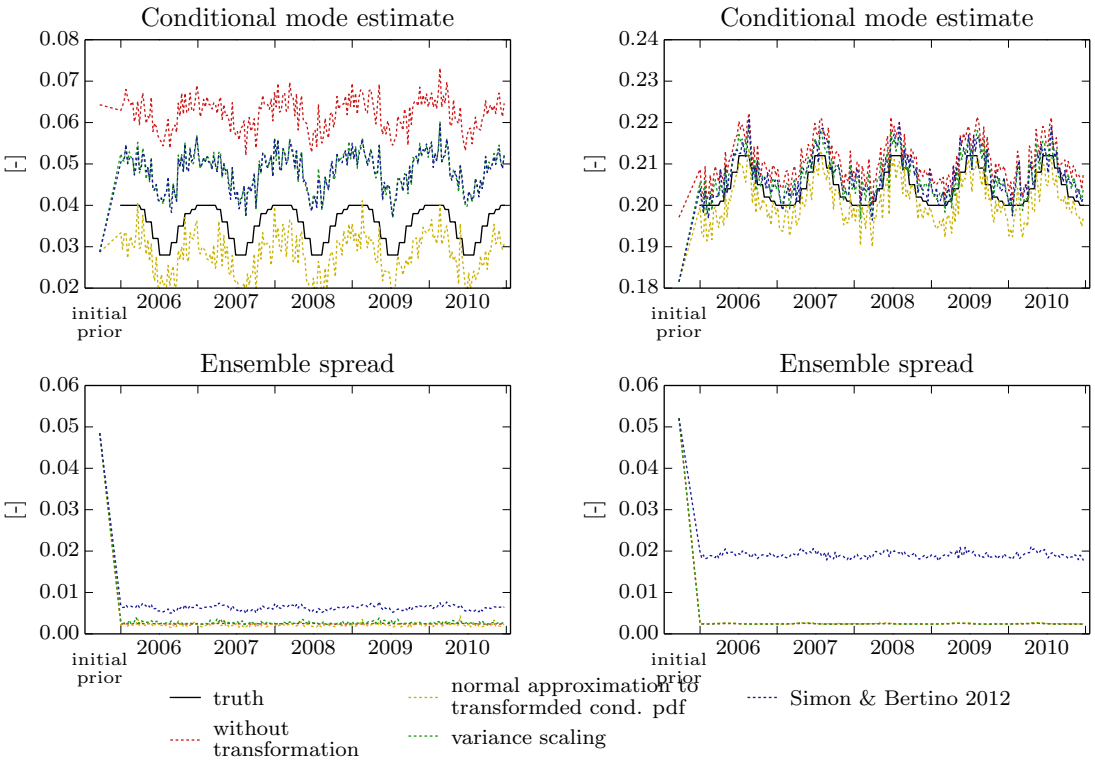


Figure 4.10: Assimilation results for one seasonal canopy albedo parameter of an evergreen PFT in the visible (left) and the near-infrared (right) domain. The ensemble spread is the spread after the update before the ensemble is inflated.

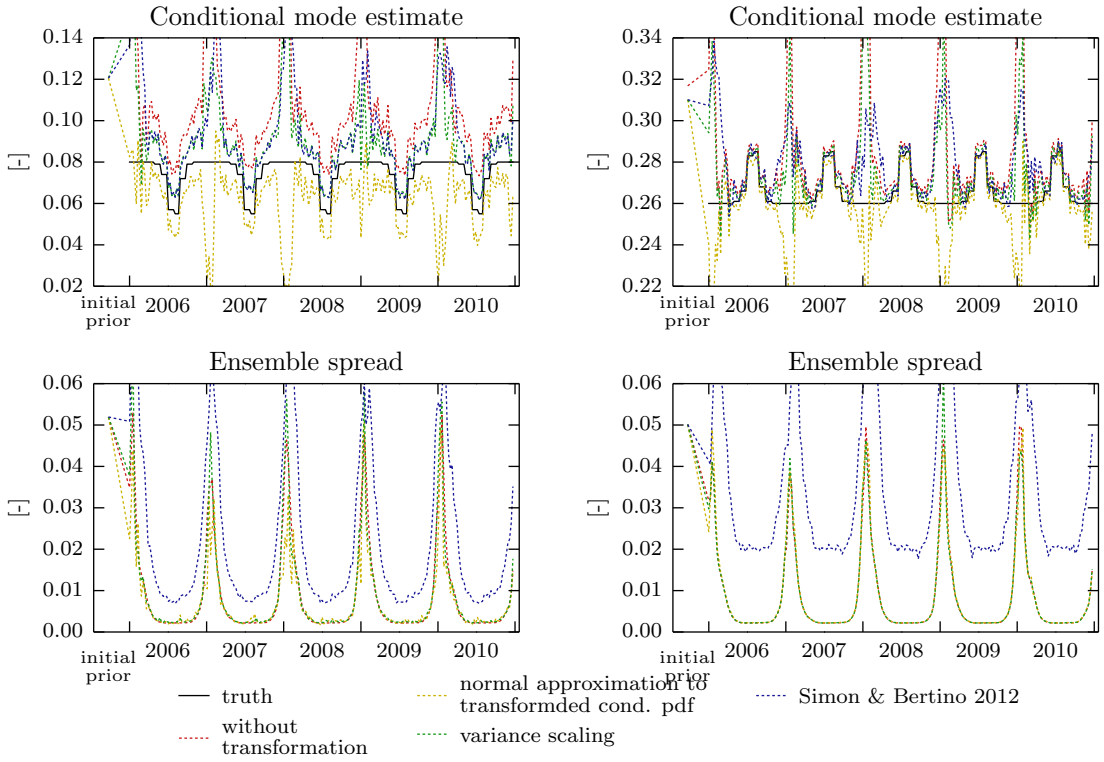


Figure 4.11: Assimilation results for one seasonal canopy albedo parameter of a deciduous PFT in the visible (left) and the near-infrared (right) domain. The ensemble spread is the spread after the update before the ensemble is inflated.

Figure 4.11 is similar to Figure 4.10 and shows the same quantities but the estimated parameter belongs to a deciduous vegetation type (Temperate Broadleaf Deciduous) whereas the parameter in Figure 4.10 was that of an evergreen vegetation type (Tropical Evergreen). In conjunction with Figures 4.4 and 4.5, we see that all four methods are able to constrain the parameter only when observations of the canopy are possible. When the canopy fraction decreases due to a decrease in LAI, the conditional mode diverges from the truth. During phases with a small or no observable canopy fraction, the ensemble spread also grows continuously. When the canopy fraction starts to increase again, the ensemble spread decreases and the conditional mode approaches the true value again. Further, for the deciduous PFT, the differences in the estimated conditional covariance between the modified method of Simon and Bertino (2012) and the other three update methods are larger than for the parameter of the evergreen PFT.

The results for other vegetation types are qualitatively similar to either Figure 4.10 or Figure 4.11 and are summarised in Figure 4.12 and discussed in section 4.4. The results for the visible and the near-infrared canopy albedo parameters are qualitatively the same, with the normal approximation to the transformed conditional pdf underestimating the parameters while the other three methods overestimate them.

As for the estimation of the constant parameters, the KF without transformation shows

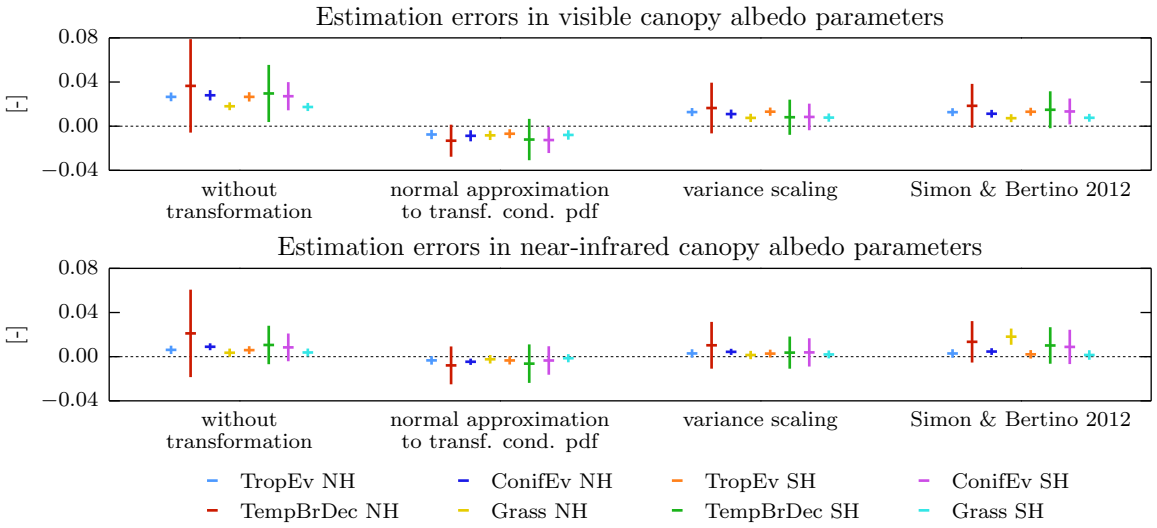


Figure 4.12: Bias and error variability of the conditional mode for the estimation of seasonal canopy albedo parameters. Horizontal lines indicate the mean error, vertical bars above and below indicate one standard deviation of the errors.

	RMSE	bias	std of errors
without transformation	0.052	0.017	0.020
normal approx. to cond. pdf	0.024	-0.007	0.010
covariance scaling	0.026	0.007	0.011
Simon and Bertino (2012)	0.030	0.010	0.011

Table 4.2: Overall root mean square error (RMSE), bias and standard deviation of errors for the estimation of seasonal canopy albedo parameters.

the largest absolute errors while the other three methods perform similarly. But contrary to the example with constant parameters, we see different error variations for the different vegetation types. The estimates for the deciduous vegetation type have a higher error variability than the for the evergreen types. And the estimates for the SH coniferous type also have a higher error variability than for the NH coniferous type.

Figure 4.13 shows the time-series correlations of the estimated parameters with the true values and the ratios of their standard deviations (for an explanation of the diagram see Taylor, 2001). The PFTs can be divided into two groups, with the deciduous types and the coniferous type on the southern hemisphere in one group and the other types in the second group. The first group exhibits low correlation values ($\sim$ 0.6 and below) and a higher variability in the estimated time series than in the true time series. The second group has high correlation values ($\sim$ 0.8 and above) and approximately the same temporal variability as the true time series.

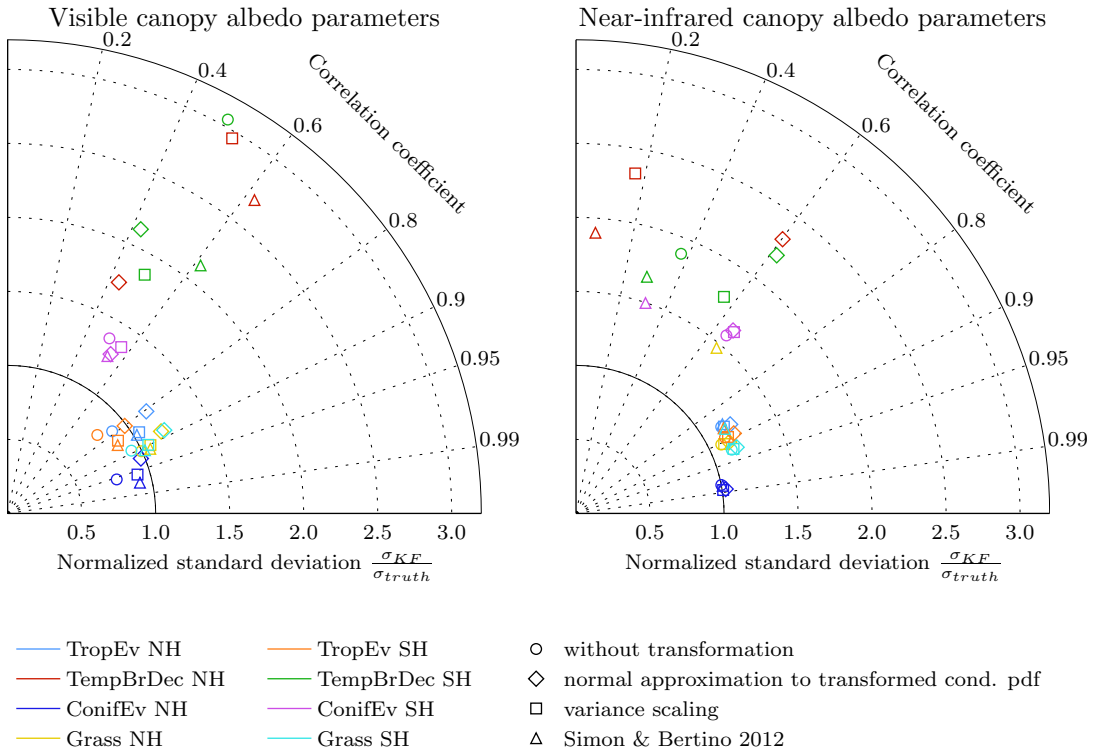


Figure 4.13: Taylor diagram of the estimated time series of seasonal canopy albedo parameters.

4.4 Summary and discussion of assimilation experiments

The comparison of experiments with and without inflation shows, that inflation is necessary to maintain a sufficient ensemble spread that allows the observations to have an impact on the estimate. We therefore conclude that the estimation of seasonal parameters requires the inflation of the updated ensemble. In our experiments, we inflate the ensemble after all observations at one observation time have been assimilated. This causes a risk of loosing ensemble spread already early during the sequential assimilation of the single, scalar observations and leaves room for improvement.

The low correlation and high error variability values of the deciduous types are due to the fact that the canopy albedo parameters are unconstrained when there is no canopy to observe. During these times, model error builds up in the parameter estimates and they diverge from the true parameter values. This effect could be damped by using adaptive inflation methods (Anderson, 2009b). The results for the coniferous type on the southern hemisphere are also rooted in the lack of observations of the canopy, but in this case due to the small global fraction of this type (see Figure 4.1).

The comparison of the four update methods yields similar results for the estimation of fixed and seasonal parameters. The covariance scaling method performs marginally better than the modified method of Simon and Bertino (2012) and the normal approximation of the transformed conditional pdf is marginally better than these two. Lastly, the KF without transformation leads to distinctly larger estimation errors than the other three methods.

The different signs of the mean errors and also the different magnitudes agree with the results from the comparison of the the linear and nonlinear regression curves in section 3.8.1. Figure 4.14 shows the estimation errors of the conditional mode (see section 3.8.2), overlaid with contours showing a two-dimensional histogram of truth-observation pairs from the assimilation experiments with seasonal canopy albedo parameters. We use the true values as approximations of the prior modes occurring in the four experiments. This approach is justified by the small estimation errors compared to the size of the interval (0, 1). The contours then approximately show the conditions occurring during the assimilation experiments. The regions overlaid by the bulk of the truth-observation pairs indicate different expected errors for the conditional mode in model space from each of the four update methods. Our experiments confirm these expectations. In agreement with Figure 4.14, the KF without transformation caused the largest errors. The normal approximation of the transformed conditional pdf caused too small estimates. The covariance scaling and the modified method of Simon and Bertino (2012) perform similarly while covariance scaling is slightly better.

The concentration of the truth-observations pairs at the lower edge of the plot, that is in the area of small observation values, in conjunction with the errors in the estimated conditional covariance (see Figure 3.12) also explains the larger ensemble spread for the modified method of Simon and Bertino (2012) in Figures 4.10 and 4.11.

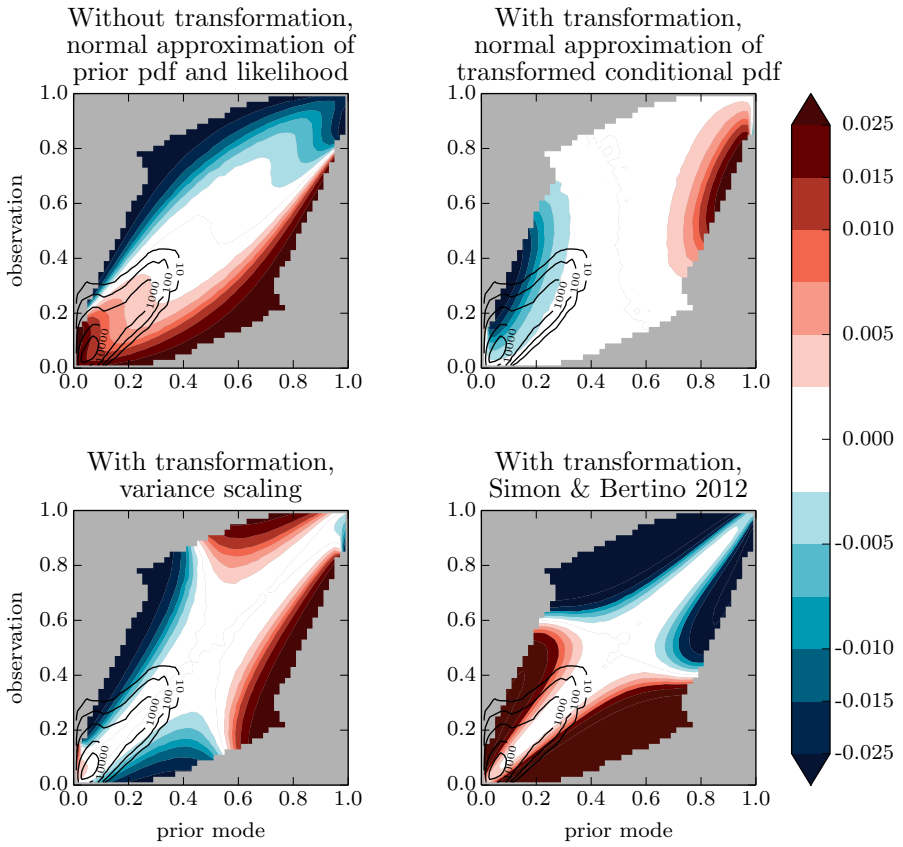


Figure 4.14: Error in the estimated conditional mode as a function of the mode of the prior distribution and the observation. Overlaid are contours showing a two-dimensional histogram of truth-observation pairs from the assimilation experiments (observation of glacier grid boxes are excluded). The histogram shows counts per 0.02×0.02 box and the innermost contour indicates a count value of 30000.

The results of our experiments for an observation error covariance of 0.0001 (standard deviation 0.01) are shown in Appendix B. They are qualitatively similar to the results presented in this section. But the results for the smaller observation error covariance indicate that the choice of the update method is less critical for smaller observation errors. This is due to the fact that all involved distributions are very narrow or become very narrow after the first assimilation. Thus, the distributions are also more symmetrical and closer to Gaussian distributions. And the bounds of the interval are of less importance because the involved pdfs drop off sharply well before they reach the bounds.

4.5 The step to real observations

The next step would be to assimilate real land surface albedo observations. First, this would require a careful adjustment of the observation error model to the characteristics of the observations (for example, Liu et al. (2009) state a small negative bias for MODIS observations).

The larger challenges, however, lie on the side of the model. The assimilation of albedo observations as described in this section will adjust the parameters such that they compensate for all sources of error in the predicted observations. These sources include an incorrect background albedo map, a mismatch in the phenological cycles of the model and the observations, and a mismatch in the assumed and the true vegetation distribution. Dalmonech and Zaehle (2013), for example, compared the JSBACH phenology to satellite-derived proxies for vegetation activity and found shifts in the phenological cycles. And Brovkin et al. (2013) identified problems in the distribution of bare ground and different vegetation types.

A technical point to consider is the extension to several tiles within a model grid box of JSBACH. This extension complicates the estimation of canopy albedo parameters from grid box observations because the joint distribution of parameters and observations will become less Gaussian. Consequently, the linear regression approximation of the relation between observed states and unobserved parameters will be less valid and the quality of the estimates will deteriorate. A possible solution to the problem of exacerbated non-Gaussianity could be the use of multivariate Gaussian anamorphosis. Multivariate Gaussian anamorphosis aims to transform the state vector such that the transformed joint pdf is truly multivariate Gaussian and linear estimation techniques are close to optimal.

Chapter 5

Summary, conclusions, and outlook

5.1 Summary

The motivation for this thesis emerged from the analysis of the new JRC-TIP data set of radiative transfer parameters for vegetation canopies. The analysis of this data set shows that the effective canopy single scattering albedo in the visible and in the near-infrared domain follows a seasonal cycle. We therefore speculate that the canopy albedo parameters in JSBACH should also follow a seasonal climatology. The derivation of such a climatology requires a time series of parameter values which we suggest to derive with the EnKF.

The application of the EnKF for bounded quantities like albedo causes physically inconsistent estimates, on the one hand. The reasons for these estimates are the purely statistical nature of the EnKF's update step, the approximations of the nonlinear state-observation relationship with a linear relationship, and sampling errors due to the finite ensemble size. On the other hand, the application of the EnKF for bounded quantities, like albedo, causes biased estimates. The reasons for these errors are the inherent non-Gaussian properties of the bounded distributions of the state variables and the observation errors as well as the – assumed – state-dependent and non-zero mean observation errors.

For the first time, we analyse the influence non-Gaussian state and observation error distributions, nonlinear observation operators, and state-dependent, non-zero mean observation errors on the EnKF using a linear regression framework. Linear regression has been previously related to the KF and the EnKF (Duncan and Horn, 1972; Lei and Bickel, 2011) but has so far not been used to understand the estimation errors. We find that the total error arises from errors in approximating a nonlinear regression function with a linear regression function, from errors in the estimation of this linear regression function, and from errors due to the approximation of a non-Gaussian conditional pdf with a Gaussian pdf.

We extend the analysis of estimation errors and the linear regression framework to the EnKF with Gaussian anamorphosis. Gaussian anamorphosis transforms the state variables and the observations from the model space to a transformed space. The transformation function, or anamorphosis function, is chosen such that the transformed state and the transformed observation error follow Gaussian distributions. Further, the state variables

are transformed from a bounded to an unbounded domain. This transformation to Gaussian distributions improves the quality of the EnKF estimates because the linear regression approximation used by the EnKF is a better approximation to the state-observation relationship in the transformed space than it is in model space. The transformation to an unbounded domain additionally ensures physically consistent values for the inversely transformed estimates in model space.

The estimation of the transformed conditional mean and the transformed conditional covariance requires the transformed observation error covariance. For the first time, we derive approximations of the transformed observation error covariance based on the transformation of the observation error pdf, explicitly stating the assumptions used in the approximation. Our derivation is an extension of an ensemble-based method to estimate the transformed observation error covariance (Simon and Bertino, 2012). Using the linear regression framework, we find that the estimate of the linear regression function and, consequently, the estimates of the conditional mean and covariance from the method of Simon and Bertino (2012) depend sensitively on the actual realisation of an observation, rather than on the statistical properties of the observation. We then suggest a new approximation of the transformed observation error covariance based on a scaling approach that relates the transformed observation error covariance to the sample covariance of the transformed ensemble.

We compare the method of Simon and Bertino (2012), our newly suggested covariance scaling, a direct approach that approximates the true transformed conditional pdf with a normal pdf, and the KF without transformation with respect to the estimated conditional mode and the estimated conditional covariance in model space. For this comparison we introduce the approximate conditional pdf. This pdf is defined by the normal distribution with mean and covariance given by the ensemble mean and covariance in model space for the KF without transformation. For the other three methods, the approximate conditional pdf results from the inverse transformation of the approximate transformed conditional pdf. The approximate transformed conditional pdf is given by a normal distribution with mean and covariance equal to the ensemble mean and covariance of the transformed ensemble. The comparison of the estimated conditional modes and covariances shows that the covariance scaling method and the method of Simon and Bertino (2012) perform best for typical assimilation conditions. Numerically and statistically, however, covariance scaling is more favourable.

We confirm this finding experimentally by setting up a sequential data assimilation framework based on the ensemble adjustment Kalman filter and the dynamic global vegetation model JSBACH. We generate synthetic observations from a virtual truth and assimilate these observations to retrieve constant and seasonal canopy albedo parameters, respectively. In our experiments, we find that the canopy albedo parameters can be retrieved from the synthetic observations, given that there is a sufficiently large fraction of canopy that contributes to the observations.

Regarding the four compared methods, all retrieve the seasonal cycles of the parameters

equally well. But they differ in the absolute magnitude of the estimated parameters for both the constant and the seasonal parameters. The ranking of the magnitudes of the errors for the four methods in our experiments agrees with the expected errors from our theoretical considerations in the linear regression framework. The numerically expensive method that approximates the true transformed conditional pdf is marginally better than the method of Simon and Bertino (2012) and covariance scaling, which perform similarly and the KF without variable transformation falls behind. Our results are qualitatively similar for the visible and the near-infrared canopy albedo parameters. But the differences in the absolute values between the four update methods are greatly reduced for the near-infrared parameters.

5.2 Conclusions

Our motivating research question was

- Can we retrieve a climatology of canopy albedo parameters from observations of land surface albedo with the ensemble Kalman filter and Gaussian anamorphosis?

In a twin experiment where only the perturbed canopy albedo parameters and random observation errors are the source of the deviations of the assimilated observations from the model state the answer is yes. We can retrieve such a climatology from land surface albedo observations. We show in section 4.3 that the EnKF with Gaussian anamorphosis can retrieve the seasonality in the parameters, independent of the chosen method for the observational update.

Our second research questions was

- What is the best method (out of these four) to assimilate albedo observations with the ensemble Kalman filter from a theoretical point of view?

The answer to this question is rooted in the theoretical considerations for the estimation of bounded quantities in chapter 3. We confirm theoretically – and later experimentally – that using the EnKF with Gaussian anamorphosis to transform the state and the observations yields better estimates than the EnKF without transformation. The EnKF with Gaussian anamorphosis, however, requires an estimate of the transformed observation error covariance which leaves the question which of the remaining three methods performs best. From our theoretical examination and from the data assimilation experiments, we find that our new covariance scaling method is the best choice. It performs only marginally better than our modification of the method of Simon and Bertino (2012). But the covariance scaling technique is easier to implement because it does not require the transformed observation pdf which either causes high computational cost during the assimilation or requires pre-calculated look-up tables. Further, covariance scaling is statistically more consistent because the estimated linear regression function does not depend on the realisation of the assimilated observation.

Our findings are relevant for quantities whose numerical value is close to the bounds of their physical domain. The comparison of the experimental results for visible and near-infrared canopy albedo parameters and the theoretical results confirm the intuition that the estimation will be the more difficult the closer the values are to the bounds of the domain. From the quantitatively different results for the visible and the near-infrared canopy albedo parameters as well as from the results for the experiments with a smaller observation error covariance in Appendix B, we conclude that the importance of treating the bounds depends on the relation of the widths of the involved pdfs, characterised by their covariances or standard deviations, to the distance of the peak of the pdf from the bounds. As a vague generalisation we state that, if the numerical values of the quantities of interest are apart from the bounds of the domain by about four to five times their standard deviation, the effects of the boundedness and non-Gaussianity become nearly negligible compared to other error sources. This holds at least for unimodal logit-normal pdfs considered in this thesis.

5.3 Outlook

The logical next step regarding the estimation of parameters is the assimilation of real observations. But leaving the idealised world of twin experiments with their isolated error sources makes this step somewhat adventurous. In principle we see two approaches to cope with the multitude of errors in assimilation experiments with real data:

1. assimilate one type of observations after another to estimate different types of variables one by one or

2. assimilate many types of observations to estimate many different types of variables simultaneously.

The first approach offers a great deal of control on the assimilation results and is similar to our approach in this thesis. But this approach attributes most of the errors to the first few types of variables that are estimated. This is because the deviations of the predicted observations from the actual observations will be caused by deficiencies in several types of variables while the assimilation will correct only one of them. This approach requires a careful selection of the order in which the observations are assimilated and of appropriate localisation methods, which are not discussed in this thesis.

The second approach poses large challenges on the generation of initial ensembles that exhibit desirable covariance structures which minimise spurious correlations. For mildly nonlinear models and nearly multivariate Gaussian distributions, this approach appears feasible and as the more promising one. But if Gaussian anamorphosis is required for several types of variables and observations, ensuring multivariate Gaussian distributions in the transformed space will be a major challenge.

From our point of view of the EnKF as a linear regression method, the combination of Gaussian anamorphosis with the EnKF essentially turns the linear regression into a

nonlinear regression approach. Nott et al. (2011) have already noted the link between a non-Gaussian extension of the EnKF by Lei and Bickel (2011) and nonlinear regression methods. Lei and Bickel (2011) extend the EnKF to higher moments. This can also be seen as extending the linear regression in the EnKF to regression methods that use more than two parameters (slope and intercept) to approximate the conditional mean. With respect to this thesis, the covariance scaling can possibly be transferred to higher moments or other characteristics of the ensembles to derive regressions with more than just two parameters.

Appendix A

Comparison of KF estimates for model space prior and observation error covariance 0.0001

The true conditional pdf in model space is bimodal in most cases due to the narrow prior and observation error distributions.

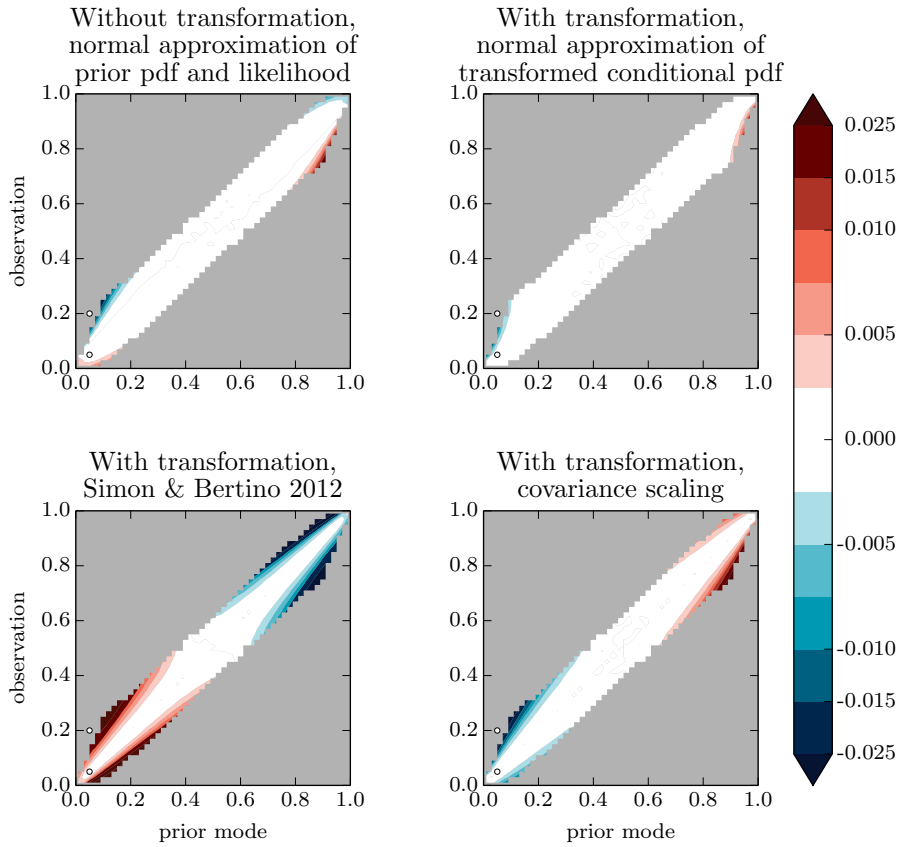


Figure A.1: Error in the estimated conditional mode as a function of the mode of the prior distribution and the observation. The prior distribution has a covariance that is equal to the covariance of the observation error of 0.0001 (standard deviation 0.01). Grey areas indicate a bimodal conditional pdf.

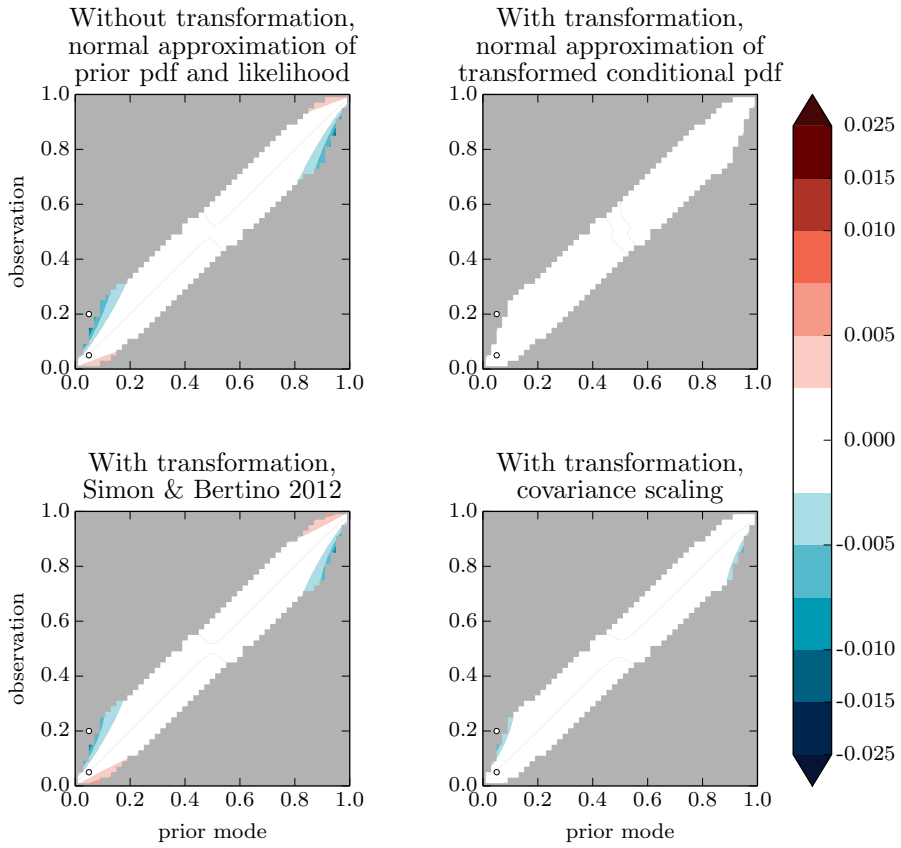


Figure A.2: Error in the square root of the estimated conditional variance as a function of the mode of the prior distribution and the observation. The prior distribution has a covariance that is equal to the covariance of the observation error of 0.0001 (standard deviation 0.01). Grey areas indicate a bimodal conditional pdf.

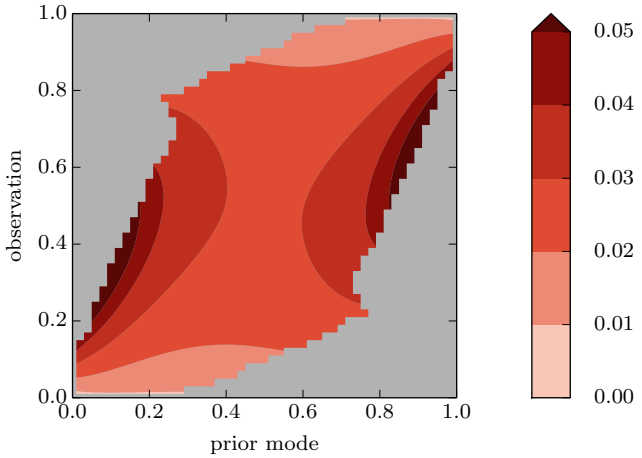


Figure A.3: Square root of the true conditional covariance.

Appendix B

Data assimilation experiments for observation error covariance 0.0001

B.1 Experiments with fixed canopy albedo parameters

The covariance of the initial prior ensemble of canopy albedo parameters in model space was 0.0025 (standard deviation 0.05) as for the experiments in chapter 4. The experiments used an inflation magnitude of 0.04.

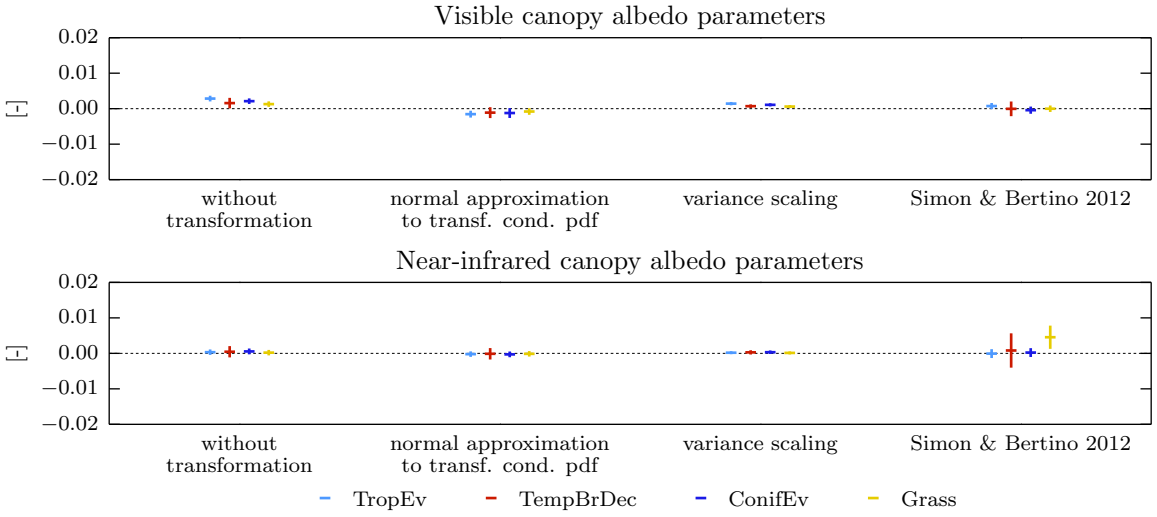


Figure B.1: Assimilation results for one fixed canopy albedo parameter in the visible (left) and the near-infrared (right) domain. Dark-coloured lines show the results before inflation and the light-coloured lines show the results after inflation.

	RMSE	bias	std of errors
without transformation	0.002	0.001	0.001
normal approx. to cond. pdf	0.002	-0.001	0.001
variance scaling	0.001	0.001	0.001
Simon and Bertino (2012)	0.004	0.001	0.003

Table B.1: Overall root mean square error (RMSE), bias and standard deviation of errors for the estimation of fixed canopy albedo parameters.

B.2 Experiments with fixed seasonal albedo parameters

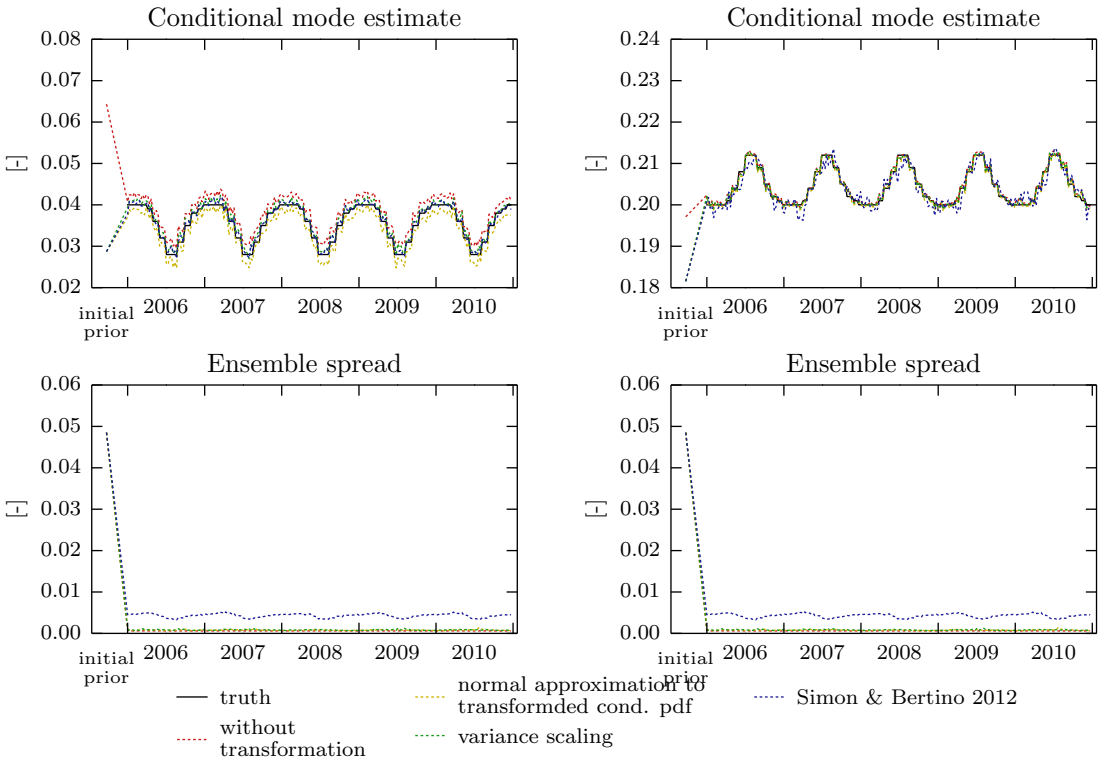


Figure B.2: Assimilation results for one seasonal canopy albedo parameter in the visible (left) and the near-infrared (right) domain. The ensemble spread is the spread after the update before the ensemble is inflated.

	RMSE	bias	std of errors
without transformation	0.008	0.002	0.004
normal approx. to cond. pdf	0.006	-0.001	0.003
variance scaling	0.006	0.000	0.003
Simon and Bertino (2012)	0.013	0.001	0.006

Table B.2: Overall root mean square error (RMSE), bias and standard deviation of errors for the estimation of seasonal canopy albedo parameters.

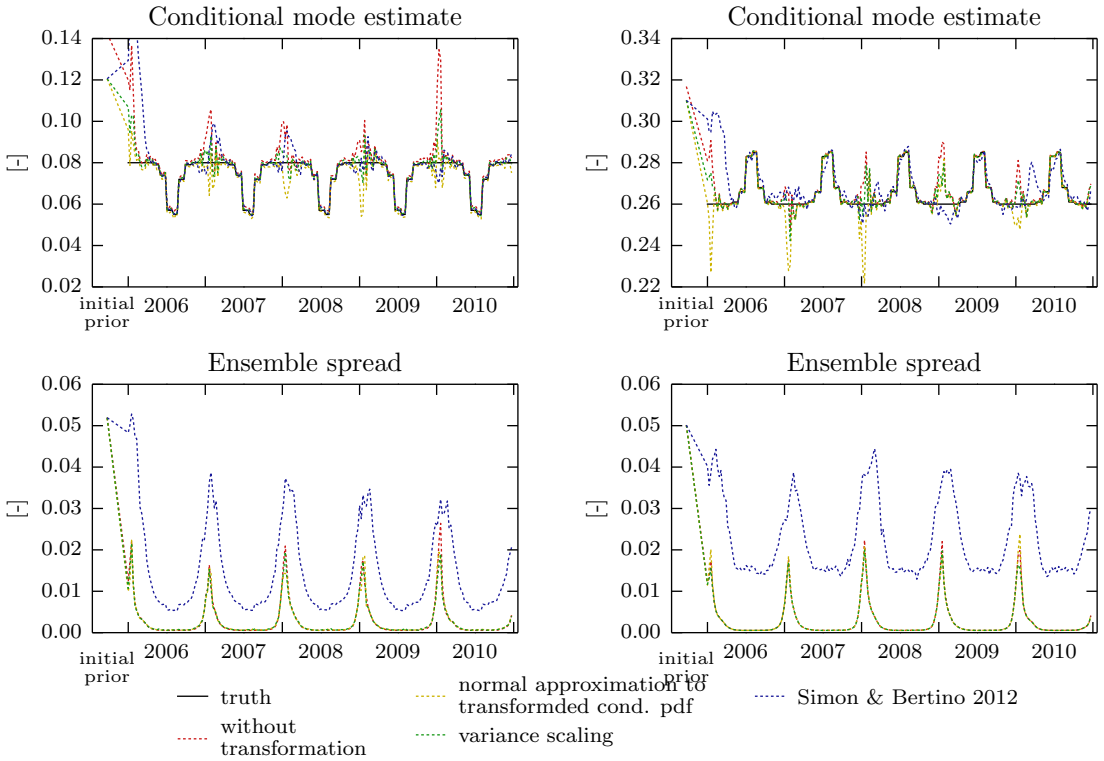


Figure B.3: Assimilation results for one seasonal canopy albedo parameter of a deciduous PFT in the visible (left) and the near-infrared (right) domain. The ensemble spread is the spread after the update before the ensemble is inflated.

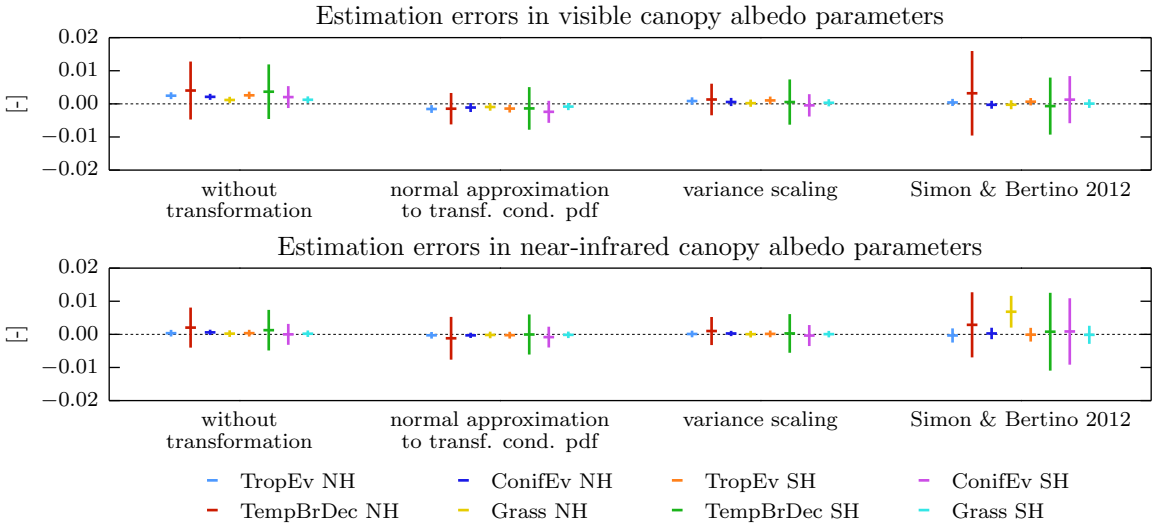


Figure B.4: Bias and error variability of the conditional mode for the estimation of seasonal canopy albedo parameters. Horizontal lines indicate the mean error, vertical bars above and below indicate one standard deviation of the errors.

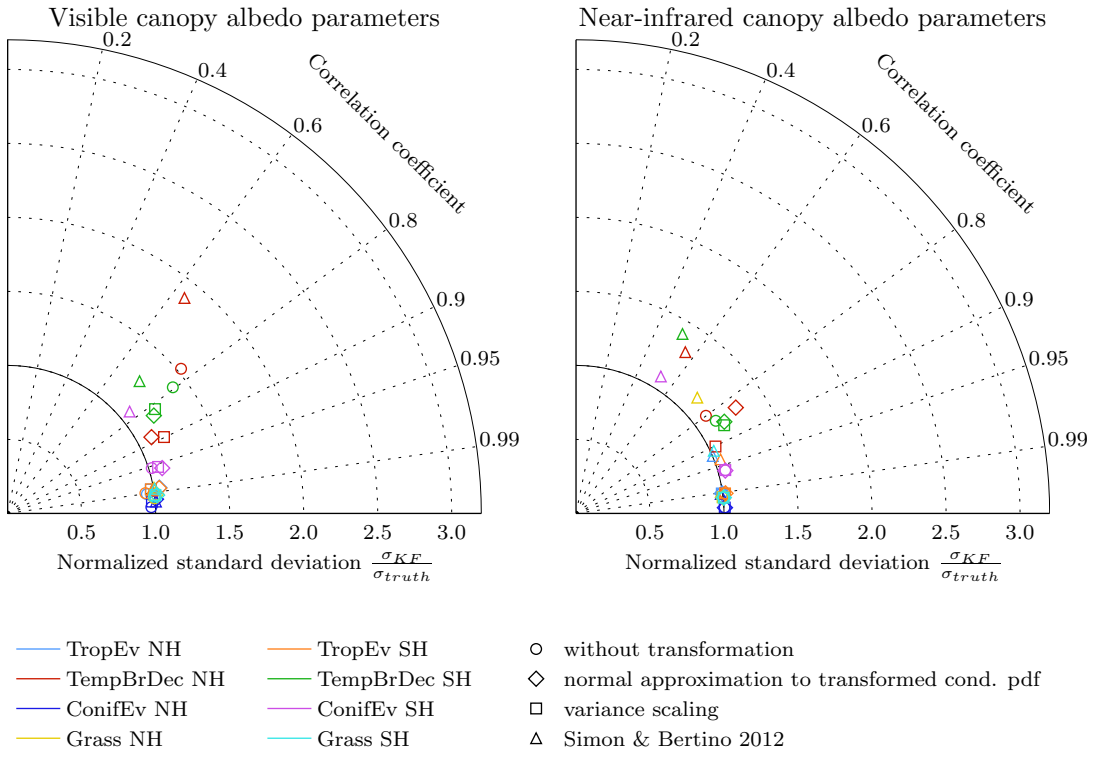


Figure B.5: Taylor diagram of the estimated time series of seasonal canopy albedo parameters.

Bibliography

Anderson, J. L., 2001: An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, **129 (12)**, 2884–2903.

Anderson, J. L., 2003: A Local Least Squares Framework for Ensemble Filtering. *Monthly Weather Review*, **131 (4)**, 634–642.

Anderson, J. L., 2009a: Ensemble Kalman filters for large geophysical applications. *IEEE Control Systems Magazine*, **29 (3)**, 66–82.

Anderson, J. L., 2009b: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A*, **61 (1)**, 72–83.

Anderson, J. L., 2010: A Non-Gaussian Ensemble Filter Update for Data Assimilation. *Monthly Weather Review*, **138 (11)**, 4186–4198.

Anderson, J. L., 2012: Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation. *Monthly Weather Review*, **140 (7)**, 2359–2371.

Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Monthly Weather Review*, **127 (12)**, 2741–2758.

Anderson, J. L. and N. Collins, 2007: Scalable Implementations of Ensemble Filter Algorithms for Data Assimilation. *Journal of Atmospheric and Oceanic Technology*, **24 (8)**, 1452–1463.

Anderson, J. L., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A Community Facility. *Bulletin of the American Meteorological Society*, **90 (9)**, 1283–1296.

Balsamo, G., S. Boussetta, P. Lopez, and L. Ferranti, 2010: Evaluation of ERA-Interim and ERA-Interim-GPCP-rescaled precipitation over the U.S.A. ECMWF, ERA Report Series.

Bertino, L., G. Evensen, and H. Wackernagel, 2002: Combining geostatistics and Kalman filtering for data assimilation in an estuarine system. *Inverse Problems*, **18 (1)**, 1–23.

Bertino, L., G. Evensen, and H. Wackernagel, 2003: Sequential Data Assimilation Techniques in Oceanography. *International Statistical Review*, **71 (2)**, 223–241.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter . Part I : Theoretical Aspects. *Monthly Weather Review*, **129 (3)**, 420–436.

Bocquet, M., C. A. Pires, and L. Wu, 2010: Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation. *Monthly Weather Review*, **138 (8)**, 2997–3023.

Brankart, J.-M., C.-E. Testut, D. Béal, M. Doron, C. Fontana, M. Meinvielle, P. Brasseur, and J. Verron, 2012: Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations. *Ocean Science*, **8 (2)**, 121–142.

Brovkin, V., L. Boysen, T. Raddatz, V. Gayler, a. Loew, and M. Claussen, 2013: Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations. *Journal of Advances in Modeling Earth Systems*, **5 (1)**, 48–57.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, **126 (6)**, 1719–1724.

Charney, J., W. J. Quirk, S. Chow, and J. Kornfield, 1977: A Comparative Study of the Effects of Albedo Change on Drought in Semi–Arid Regions. *Journal of the Atmospheric Sciences*, **34 (9)**, 1366–1385.

Chatterjee, a. and a. M. Michalak, 2013: Technical Note: Comparison of ensemble Kalman filter and variational approaches for CO¡sub¿2¡/sub¿ data assimilation. *Atmospheric Chemistry and Physics*, **13 (23)**, 11 643–11 660.

Chilès, J.-P. and P. Delfiner, 1999: *Geostatistics. Modeling Spatial Uncertainty*. Wiley, New York.

Cohn, S. E., 1997: An introduction to estimation theory. *Journal of the Meteorological Society of Japan*, **75 (1B)**, 257–288.

Dalmonech, D. and S. Zaehle, 2013: Towards a more objective evaluation of modelled land-carbon trends using atmospheric CO2 and satellite-based vegetation activity observations. *Biogeosciences*, **10 (6)**, 4189–4210.

Dee, D. P. and A. M. Da Silva, 1998: Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, **124 (545)**, 269–295.

Dee, D. P., et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597.

Doron, M., P. Brasseur, and J.-M. Brankart, 2011: Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical–biogeochemical model: Twin experiments. *Journal of Marine Systems*, **87 (3-4)**, 194–207.

Doron, M., P. Brasseur, J.-M. Brankart, S. N. Losa, and A. Melet, 2013: Stochastic estimation of biogeochemical parameters from Globcolour ocean colour satellite data in a North Atlantic 3D ocean coupled physical–biogeochemical model. *Journal of Marine Systems*, **117-118**, 81–95.

Doucet, A., N. de Freitas, and N. Gordon, 2001: An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo methods in practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds., Springer, New York.

Dumont, M., Y. Durand, Y. Arnaud, and D. Six, 2012: Variational assimilation of albedo in a snowpack model and reconstruction of the spatial mass-balance distribution of an alpine glacier. *Journal of Glaciology*, **58 (207)**, 151–164.

Duncan, D. B. and S. D. Horn, 1972: Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis. *Journal of the American Statistical Association*, **67 (340)**, 815–821.

Durand, M. and S. a. Margulis, 2007: Correcting first-order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme. *Journal of Geophysical Research*, **112 (D13)**, 1–15.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99 (C5)**, 10 143.

Evensen, G., 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53 (4)**, 343–367.

Evensen, G., 2009a: *Data Assimilation. The Ensemble Kalman Filter*. 2nd ed., Springer, Berlin.

Evensen, G., 2009b: The Ensemble Kalman Filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, **29 (3)**, 83–104.

Fontana, C., P. Brasseur, and J.-M. Brankart, 2013: Toward a multivariate reanalysis of the North Atlantic Ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data. *Ocean Science*, **9 (1)**, 37–56.

Freitag, M. A. and R. W. E. Potthast, 2013: Synergy of inverse problems and data assimilation techniques. *Large Scale Inverse Problems*, M. Cullen, M. A. Freitag, S. Kindermann, and R. Scheichl, Eds., De Gruyter, Berlin, Boston, 1–54.

Gardiner, C. W., 2004: *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*. 3rd ed., Springer, Berlin.

Gauthier, P. and J.-N. Thépaut, 2001: Impact of the Digital Filter as a Weak Constraint in the Preoperational 4DVAR Assimilation System of Météo-France. *Monthly Weather Review*, **129 (8)**, 2089–2102.

Gelb, A., (Ed.) , 1974: *Applied optimal estimation*. The MIT Press, Cambridge, MA.

Giorgetta, M. a., et al., 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, **5 (3)**, 572–597.

Gitelson, A. a. and M. N. Merzlyak, 1996: Signature Analysis of Leaf Reflectance Spectra: Algorithm Development for Remote Sensing of Chlorophyll. *Journal of Plant Physiology*, **148 (3-4)**, 494–500.

Gordon, N., D. Salmond, and A. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, **140 (2)**, 107–113.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, **129 (11)**, 2776–2790.

Hendricks Franssen, H. J. and W. Kinzelbach, 2008: Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, **44 (9)**, n/a–n/a.

Ho, Y. and R. Lee, 1964: A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, **9 (4)**, 333–339.

Hollinger, D. Y., et al., 2010: Albedo estimates for land surface models and support for a new paradigm based on foliage nitrogen concentration. *Global Change Biology*, **16 (2)**, 696–710.

Houtekamer, P. L., B. He, and H. L. Mitchell, 2013: Parallel Implementation of an Ensemble Kalman Filter. *Monthly Weather Review*, 130923140955002.

Houtekamer, P. L. and H. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, **129 (1)**, 123–137.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, **126 (3)**, 796–811.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP Version 2.1. *Geophysical Research Letters*, **36 (17)**, L17 808.

Janjić, T., D. McLaughlin, S. E. Cohn, and M. Verlaan, 2014: Conservation of Mass and Preservation of Positivity with Ensemble-Type Kalman Filter Algorithms. *Monthly Weather Review*, **142 (2)**, 755–773.

Jaynes, E. T., 2007: *Probability Theory: The Logic of Science*. Cambebridge University Press, Cambridge.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Johnson, N. and S. Kotz, 1970: *Distributions in statistics: Continuous univariate distributions*. Wiley, New York.

Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, **82 (1)**, 35.

Kalman, R. E. and R. S. Bucy, 1961: New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, **83 (1)**, 95.

Kaminski, P., A. Bryson, and S. Schmidt, 1971: Discrete square root filtering: A survey of current techniques. *IEEE Transactions on Automatic Control*, **16 (6)**, 727–736.

Kaminski, T., et al., 2013: The BETHY/JSBACH Carbon Cycle Data Assimilation System: experiences and challenges. *Journal of Geophysical Research: Biogeosciences*, **118 (Clm)**, n/a–n/a.

Kemp, S., M. Scholze, T. Ziehn, and T. Kaminski, 2014: Limiting the parameter space in the Carbon Cycle Data Assimilation System (CCDAS). *Geoscientific Model Development Discussions*, **7 (1)**, 659–689.

Kitagawa, G., 1996: Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, **5 (1)**.

Knyazikhin, Y., et al., 2013: Hyperspectral remote sensing of foliar nitrogen content. *Proceedings of the National Academy of Sciences of the United States of America*, **110 (3)**, E185–92.

Krzysztofowicz, R., 1997: Transformation and normalization of variates with specified distributions. *Journal of Hydrology*, **197 (1-4)**, 286–292.

Lauvernet, C., J.-M. Brankart, F. Castruccio, G. Broquet, P. Brasseur, and J. Verron, 2009: A truncated Gaussian filter for data assimilation with inequality constraints: Application to the hydrostatic stability condition in ocean models. *Ocean Modelling*, **27 (1-2)**, 1–17.

Lawson, W. G. and J. A. Hansen, 2004: Implications of Stochastic and Deterministic Filters as Ensemble-Based Data Assimilation Methods in Varying Regimes of Error Growth. *Monthly Weather Review*, **132 (8)**, 1966–1981.

Lei, J. and P. Bickel, 2009: Ensemble Filtering for High Dimensional Non-linear State Space Models. University of California Berkeley, Statistics Technical Report 779.

Lei, J. and P. Bickel, 2011: A Moment Matching Ensemble Filter for Nonlinear Non-Gaussian Data Assimilation. *Monthly Weather Review*, **139 (12)**, 3964–3973.

Lei, J., P. Bickel, and C. Snyder, 2010: Comparison of Ensemble Kalman Filters under Non-Gaussianity. *Monthly Weather Review*, **138 (4)**, 1293–1306.

Lewis, P., J. Gómez-Dans, T. Kaminski, J. Settle, T. Quaife, N. Gobron, J. Styles, and M. Berger, 2012: An Earth Observation Land Data Assimilation System (EO-LDAS). *Remote Sensing of Environment*, **120**, 219–235.

Lien, G.-Y., E. Kalnay, and T. Miyoshi, 2013: Effective assimilation of global precipitation: simulation experiments. *Tellus A*, **65**, 1–16.

Liu, J., et al., 2009: Validation of Moderate Resolution Imaging Spectroradiometer (MODIS) albedo retrieval algorithm: Dependence of albedo on solar zenith angle. *Journal of Geophysical Research*, **114 (D1)**, D01 106.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **112 (474)**, 1177–1194.

Malik, M. J., R. van der Velde, Z. Vekerdy, and Z. Su, 2012: Assimilation of Satellite-Observed Snow Albedo in a Land Surface Model. *Journal of Hydrometeorology*, **13 (3)**, 1119–1130.

Mandel, J., L. Cobb, and J. D. Beezley, 2011: On the convergence of the ensemble Kalman filter. *Applications of Mathematics*, **56 (6)**, 533–541, `0901.2951`.

Marin, J.-M. and C. P. Robert, 2007: *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer, New York.

Meinhold, R. J. and N. D. Singpurwalla, 1983: Understanding the Kalman Filter. *The American Statistician*, **37 (2)**, 123–127.

Mitchell, H. L. and P. L. Houtekamer, 2000: An Adaptive Ensemble Kalman Filter. *Monthly Weather Review*, **128 (2)**, 416.

Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser, 2005: Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, **28 (2)**, 135–147.

Moss, R. H., et al., 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463 (7282)**, 747–56.

Nerger, L. and W. Hiller, 2013: Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Computers & Geosciences*, **55**, 110–118.

Nielsen-Gammon, J. W., X.-M. Hu, F. Zhang, and J. E. Pleim, 2010: Evaluation of Planetary Boundary Layer Scheme Sensitivities for the Purpose of Parameter Estimation. *Monthly Weather Review*, **138 (9)**, 3400–3417.

Nott, D. J., L. Marshall, and T. M. Ngoc, 2011: The ensemble Kalman filter is an ABC algorithm. *Statistics and Computing*, **22 (6)**, 1273–1276.

Nowak, W., 2009: Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. *Water Resources Research*, **45 (4)**, n/a–n/a.

Ollinger, S. V., et al., 2008: Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks. *Proceedings of the National Academy of Sciences of the United States of America*, **105 (49)**, 19 336–41.

Paige, C. C. and M. A. Saunders, 1977: Least squares estimation of discrete linear dynamic systems using orthogonal transformations. *SIAM Journal on Numerical Analysis*, **14 (2)**.

Pan, M. and E. F. Wood, 2006: Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter. *Journal of Hydrometeorology*, **7 (3)**, 534–547.

Papoulis, A., 1991: *Probability, random variables, and stochastic processes*. 3rd ed., McGraw-Hill, New York.

Pfeiffer, P. E., 1990: *Probability for applications*. Springer, New York.

Pinty, B., T. Lavergne, R. E. Dickinson, J.-L. Widlowski, N. Gobron, and M. M. Verstraete, 2006: Simplifying the interaction of land surfaces with radiation for relating remote sensing products to climate models. *Journal of Geophysical Research*, **111 (D2)**, D02 116.

Pinty, B., et al., 2011a: Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 1. Effective leaf area index, vegetation, and soil properties. *Journal of Geophysical Research*, **116 (D9)**, D09 105.

Pinty, B., et al., 2011b: Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 2. Fractions of transmitted and absorbed fluxes in the vegetation and soil layers. *Journal of Geophysical Research*, **116 (D9)**, D09 106.

Pires, C. a., O. Talagrand, and M. Bocquet, 2010: Diagnosis and impacts of non-Gaussianity of innovations in data assimilation. *Physica D: Nonlinear Phenomena*, **239 (17)**, 1701–1717.

Raddatz, T. J., et al., 2007: Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? *Climate Dynamics*, **29 (6)**, 565–574.

Rayner, P. J., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann, 2005: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Global Biogeochemical Cycles*, **19 (2)**, n/a–n/a.

Rechid, D., T. J. Raddatz, and D. Jacob, 2009: Parameterization of snow-free land surface albedo as a function of vegetation phenology based on MODIS data and applied in climate modelling. *Theoretical and Applied Climatology*, **95 (3-4)**, 245–255.

Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe, 2007: Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research*, **112 (D9)**, D09 108.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002: Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Monthly Weather Review*, **130 (1)**, 103–114.

Reick, C. H., T. Raddatz, V. Brovkin, and V. Gayler, 2013: Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, **5 (3)**, 459–482.

Sacher, W. and P. Bartello, 2008: Sampling Errors in Ensemble Kalman Filtering. Part I: Theory. *Monthly Weather Review*, **136 (8)**, 3035–3049.

Sakov, P., D. S. Oliver, and L. Bertino, 2012: An Iterative EnKF for Strongly Nonlinear Systems. *Monthly Weather Review*, **140 (6)**, 1988–2004.

Schirber, S., D. Klocke, R. Pincus, J. Quaas, and J. L. Anderson, 2013: Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world. *Journal of Advances in Modeling Earth Systems*, **5 (1)**, 58–70.

Schöniger, A., W. Nowak, and H.-J. Hendricks Franssen, 2012: Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography. *Water Resources Research*, **48 (4)**, 1–18.

Sellers, P., et al., 1995: Remote sensing of the land surface for studies of global change: Models — algorithms — experiments. *Remote Sensing of Environment*, **51 (1)**, 3–26.

Sellers, P. J., 1985: Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, **6 (8)**, 1335–1372.

Sellers, P. J., 1997: Modeling the Exchanges of Energy, Water, and Carbon Between Continents and the Atmosphere. *Science*, **275 (5299)**, 502–509.

Simon, E. and L. Bertino, 2009: Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Science*, **5 (4)**, 495–510.

Simon, E. and L. Bertino, 2012: Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model. *Journal of Marine Systems*, **89 (1)**, 1–18.

Stöckli, R., T. Rutishauser, I. Baker, M. A. Liniger, and A. S. Denning, 2011: A global reanalysis of vegetation phenology. *Journal of Geophysical Research*, **116 (G3)**, G03 020.

Stöckli, R., T. Rutishauser, D. Dragoni, J. O'Keefe, P. E. Thornton, M. Jolly, L. Lu, and a. S. Denning, 2008: Remote sensing data assimilation for a prognostic phenology model. *Journal of Geophysical Research*, **113 (G4)**, G04 021.

Sud, Y. C. and M. Fennessy, 1982: A study of the influence of surface albedo on July circulation in semi-arid regions using the glas GCM. *Journal of Climatology*, **2 (2)**, 105–125.

Talagrand, O., 1997: Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, **75 (1B)**, 191–209.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, **106 (D7)**, 7183.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble Square Root Filters*. *Monthly Weather Review*, **131 (7)**, 1485–1490.

Vamborg, F. S. E., V. Brovkin, and M. Claussen, 2011: The effect of a dynamic background albedo scheme on Sahel/Sahara precipitation during the mid-Holocene. *Climate of the Past*, **7 (1)**, 117–131.

van Leeuwen, P. J., 1999: Comment on "Data Assimilation Using an Ensemble Kalman Filter Technique". *Monthly Weather Review*, **127 (6)**, 1374–1377.

van Leeuwen, P. J. and G. Evensen, 1996: Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. *Monthly Weather Review*, **124 (12)**, 2898–2913.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, **130 (7)**, 1913–1924.

Whitaker, J. S. and T. M. Hamill, 2012: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Monthly Weather Review*, **140 (9)**, 3078–3089.

Williams, M., P. a. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius, 2005: An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, **11 (1)**, 89–105.

Yilmaz, M. T., T. DelSole, and P. R. Houser, 2011: Improving Land Data Assimilation Performance with a Water Budget Constraint. *Journal of Hydrometeorology*, **12 (5)**, 1040–1055.

Yuan, H., R. E. Dickinson, Y. Dai, M. J. Shaikh, L. Zhou, W. Shangguan, and D. Ji, 2014: A 3D Canopy Radiative Transfer Model for Global Climate Modeling: Description, Validation, and Application. *Journal of Climate*, **27 (3)**, 1168–1192.

Zehnwirth, B., 1988: A Generalization of the Kalman Filter for Models With State-Dependent Observation Variance. *Journal of the American Statistical Association*, **83 (401)**, 164.

Zhou, H., J. J. Gómez-Hernández, H.-J. Hendricks Franssen, and L. Li, 2011: An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources*, **34 (7)**, 844–864.

Acknowledgements

First of all, I sincerely thank Alex and Felix for their support and advice during the last years. I enjoyed the honest discussions and the critical questioning of my ideas. Thank you for letting me develop and defend my own ideas while shaping them into a thesis. Right next, I thank Martin for chairing the advisory panel. Your questioning of the goals of my thesis helped me focus my work into one direction. I have learned a lot from all three of you.

I would like to thank the IMPRS-ESM for the opportunity to pursue my PhD research at the Max Planck Institute in Hamburg. This thanks needs also to be addressed personally: to Antje, to Connie and to Wiebke. Your organisation and administration of the school is excellent.

Next, I thank Sönke and Gregor from Jena for their efforts concerning the offline version of JSBACH. Without the offline version, my thesis would have gone a different, probably less favourable, way.

Regarding technical support, I thank Stiig and Reiner for answering all kinds of FORTRAN-related questions.

The next group of people to thank are the organisers of the Les Houches summer school on Advanced Data Assimilation for Geosciences as well as the participants. You boosted my understanding of data assimilation and I very much enjoyed the school in 2012 as well as seeing you again at conferences.

A great thank you for proof reading (as far as time allowed) goes to Rika, Laura and Sebastian. You improved the readability of this thesis significantly.

Personal thanks goes to Michael and all other members of the TRS group, to the students of the IMPRS-ESM and to the people on the 17th floor. To name but a few that made my time here most enjoyable, I thank Philipp, Rika, Jessica, Laura, Sebastian, Werner, Florian, Vera, and Katrin.

The last words of thanks go to my parents and Doris. For their patience and support all along.

Eidesstattliche Versicherung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich versichere, dass die wörtlich oder inhaltich den benutzten Werken entnommenen Stellen von mir kenntlich gemacht wurden.

Gernot Geppert