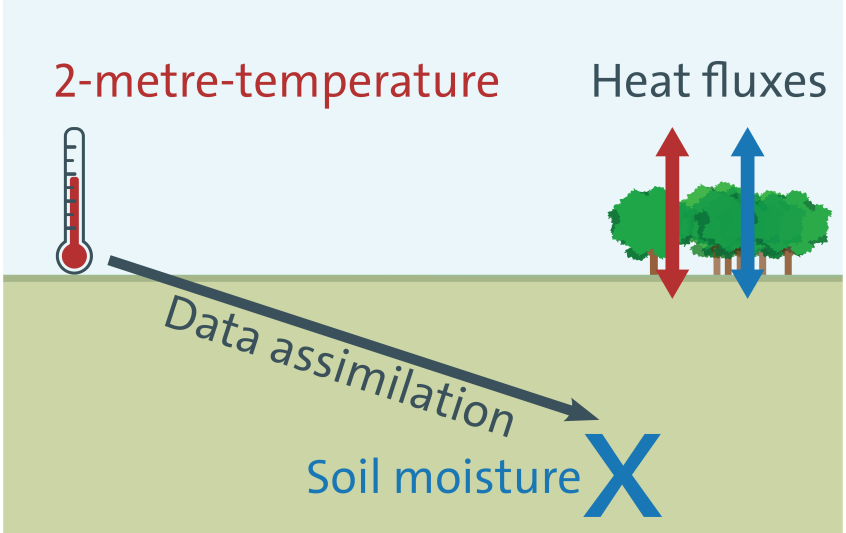
MAX-PLANCK-INSTITUT
FÜR METEOROLOGIE

IMPRS-ESM
International Max Planck Research School on Earth System Modelling

Advancing Non-Linear Methods for Coupled Data Assimilation across the Atmosphere-Land Interface



Tobias Sebastian Finn

Hamburg 2022

Berichte zur Erdsystemforschung
Reports on Earth System Science

251
2022

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.

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

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 PhD theses.

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

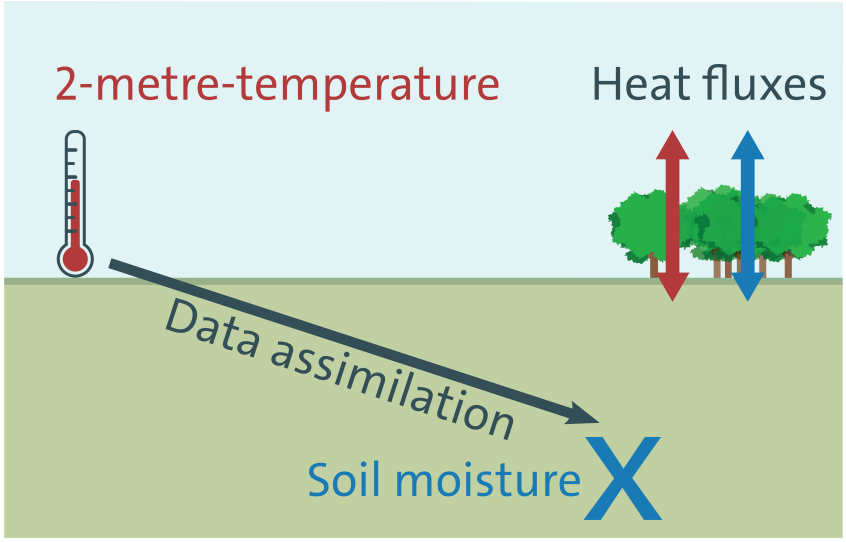
The "Reports on Earth System Science" continue the former "Reports" and "Examensarbeiten" of the Max Planck Institute.

Layout

Bettina Diallo and Norbert P. Noreiks
Communication

Copyright

Advancing Non-Linear Methods for Coupled Data Assimilation across the Atmosphere-Land Interface



Tobias Sebastian Finn

Hamburg 2022

Tobias Sebastian Finn
aus Hannover, Deutschland

Max-Planck-Institut für Meteorologie
The International Max Planck Research School on Earth System Modelling
(IMPRS-ESM)
Bundesstrasse 53
20146 Hamburg

Universität Hamburg
Geowissenschaften
Meteorologisches Institut
Bundesstr. 55
20146 Hamburg

Tag der Disputation: 14. Juni 2021

Folgende Gutachter empfehlen die Annahme der Dissertation:
Prof. Dr. Felix Ament
Dr. Gernot Geppert

Vorsitzender des Promotionsausschusses:
Prof. Dr. Dirk Gajewski

Dekan der MIN-Fakultät:
Prof. Dr. Heinrich Graener

ISSN 1614-1199

Tobias Sebastian Finn

Advancing Non-Linear Methods for Coupled Data Assimilation across the Atmosphere-Land Interface

Abstract

In this thesis, I present two complementary frameworks to improve data assimilation in Earth system models, using the atmosphere-land interface as an exemplary case. As processes and components in the Earth system are coupled via interfaces, we would expect that assimilating observations from one Earth system component into another would improve the initialization of both components. In contrast to this expectation, it is often found that assimilation of atmospheric boundary layer observations into the land surface does not improve the analysis of the latter component. To disentangle the effects on the cross-compartmental assimilation, I take a step back from operational methods and use the coupled atmosphere-land modelling platform TerrSysMP in idealized twin experiments. I synthesize hourly and sparsely-distributed 2-metre-temperature observations from a single "nature" run. I subsequently assimilate these observations into the soil moisture with different types of data assimilation methods. Based on this experimental structure, I test advanced data assimilation methods without model errors or biases.

As my first framework, I propose to use localized ensemble Kalman filters for the unification of coupled data assimilation in Earth system models. To validate this framework, I conduct comparison experiments with a localized ensemble transform Kalman filter and a simplified extended Kalman filter, as similarly used at the ECMWF. Based on my developed environment, I find that we can assimilate 2-metre-temperature observations to improve the soil moisture analysis. In addition, hourly-updating the soil moisture with an ensemble Kalman filter decreases the error within the soil moisture analysis by up to 50 % compared to a daily-smoothing with a simplified extended Kalman filter. As a consequence, observations from the atmospheric boundary layer can be directly assimilated into the land surface model without a need of any intermediate interpolation, as normally used in land surface data assimilation. The improvement suggests that the land surface can be updated based on the same hourly cycle as used for mesoscale data assimilation. My results therefore prove that a unification of methods for data assimilation across the atmosphere-land interface is possible.

As my second framework, I propose to use feature-based data assimilation to stabilize cross-compartmental data assimilation. To validate this framework, I use my implementation of an ensemble Kalman smoother that applies its analysis at the beginning of an assimilation window and resembles 4DEnVar. This smoother takes advantage of temporal dependencies in the atmosphere-land interface and improves the soil moisture analysis compared to the ensemble Kalman filter by 10 %. Subsequently based on this smoother, I introduce fingerprint operators as observational feature extractor into cross-compartmental data assimilation. These fingerprint operators take advantage of characteristic fingerprints in the difference

between observations and model that point towards forecast errors, possibly in another Earth system component. As main finding, this concept can condense the information from the diurnal cycle in 2-metre-temperature observations into two observational features. This condensation makes the soil moisture analysis more robust against a miss-specified localization radius and errors in the observational covariance.

Finally, I provide two new theoretical approaches to automatically learn such observational features with machine learning. In the first approach, I generalize ensemble Kalman filter with observational features to a novel kernelized ensemble transform Kalman filter.automatically This kernelized filter automatically constructs the feature extractor on the basis of the given ensemble data and a chosen kernel function. In the second approach, I show that parameters within the data assimilation can be learned by variational Bayes. In this way, we can find whole distributions for parameters in data assimilation and, thus, determining their uncertainties. Furthermore, I prove the ensemble transform Kalman filter as a special solution of variational Bayes in the linearized-Gaussian case. These results suggest a possibility to specify the feature extractor as neural network and to train it with variational Bayes. These two approaches therefore prove that developments in machine learning can be used to extend data assimilation.

Zusammenfassung

In dieser Arbeit stelle ich zwei unterschiedliche Frameworks vor, um die Initialisierung in gekoppelten Erdsystemmodellen für die Wettervorhersage zu verbessern. Dabei benutze ich die Schnittstelle zwischen der Atmosphäre und der Landoberfläche als Beispiel. Diese Schnittstelle bietet mir die Möglichkeit zu untersuchen, in wie weit gekoppelte Datenassimilierung möglich ist. Prozesse und Komponenten des Klimasystems sind über verschiedene Schnittstellen miteinander verbunden. Von daher würden wir erwarten, dass Beobachtungen aus der atmosphärischen Grenzschicht, auch die Initialisierung von Bodenmodellen verbessern, allerdings wurde in verschiedenen vorangegangenden Studien gezeigt, dass dies nicht der Fall ist. Um die Einflüsse von unterschiedlichen Fehler-Faktoren auf die Datenassimilierung zu reduzieren, benutze ich Experimente, die im Vergleich zur operationellen Wettervorhersage vereinfacht sind. Hierfür benutze ich das gekoppelte Atmosphären-Land Vorhersagemodel TerrSysMP. All diese Experimente basieren auf einem Lauf ohne Datenassimilierung, den ich als meine "Natur" definiere. Aus diesem Naturlauf extrahiere ich künstliche 2-Meter-Temperatur Beobachtungen, welche dann mit unterschiedlichen Datenassimilierungsverfahren in die Bodenfeuchte assimiliert werden. Mit dieser Art von Experimenten teste ich fortschrittliche und nicht-lineare Datenassimilierungsverfahren für die Atmosphären-Land-Schnittstelle.

Als erstes Framework schlage ich vor, einen lokalisierten Ensemble-Kalman-Filter für eine vereinheitlichte Datenassimilierung in Erdsystemmodellen zu verwenden. Um dieses Framework zu validieren, mache ich Vergleichsexperimente mit dem eben erwähnten lokalisierten Ensemble-Kalman-Filter und einem vereinfachten Extended-Kalman-Filter, der in ähnlicher Form beim Europäischen Zentrum für mittelfristige Wettervorhersage verwendet wird. Basierend auf meiner entwickelten Umgebung zeige ich, dass 2-Meter-Temperatur Beobachtungen dafür verwendet werden können, um die Initialisierung der Bodenfeuchte zu verbessern. Der lokalisierte Ensemble-Kalman Filter reduziert zusätzlich den Fehler in der Initialisierung der Bodenfeuchte um bis zu 50 %, im Vergleich zu dem vereinfachten Extended-Kalman-Filter. Dies zeigt zum ersten Mal, dass Beobachtungen aus der atmosphärischen Grenzschicht, direkt für die Initialisierung der Bodenfeuchte, verwendet werden können, ohne den Umweg einer Interpolierung zu nehmen, wie es bei dem vereinfachten Extended-Kalman-Filter der Fall ist. Darüberhinausgehend legen diese Verbesserungen nahe, dass die Landoberfläche mit der gleichen stündlichen Aktualisierungs-Rate, wie die Atmosphäre, initialisiert werden kann. Deshalb beweisen diese Ergebnisse, dass eine vereinheitlichte Datenassimilierung über die Atmosphären-Land-Schnittstelle hinweg möglich ist.

Als zweites Framework schlage ich vor, anstatt von Beobachtungen, Merkmale

dieser Beobachtung zu assimilieren. Dies kann die Assimilierung, über die Atmosphären-Land Schnittstelle hinweg, verbessern. Um dieses Framework zu validieren, führe ich einen Ensemble-Kalman-Smoother ein. Dieser Ensemble-Kalman-Smoother initialisiert die Bodenfeuchte auf Basis eines Assimilierungsfensters, ähnlich dem variationsgetriebenem vierdimensionellem Verfahren. Mit diesem Ensemble-Kalman-Smoother zeige ich, dass es möglich ist, zeitliche Abhängigkeiten innerhalb der Atmospähren-Land-Schnittstelle in der Datenassimilierung zu verwenden. Die Verwendung dieser Abhängigkeiten verbessert hierbei die Initialisierung der Bodenfeuchte. Auf Basis dieser Methodik, führe ich Operatoren ein, die Fingerabdrücke innerhalb von Beobachtungen ausnutzen. Diese Fingerabdruck-Operatoren nutze ich dafür, um Vorhersage-Fehler in anderen Komponenten des Erdsystems zu finden. Für die 2-Meter-Temperatur zeige ich, dass Informationen aus dem Tagesverlauf der Temperatur in 2 unterschiedliche Merkmale kondensiert werden können. Diese Kondensation macht die Initialisierung der Bodenfeuchte robuster gegen Störungen innerhalb der Lokalisierung und der Beobachtungskovarianzen. Deshalb beweisen diese Ergebnisse, dass die eingeführten Fingerabdruck-Operatoren, die Datenassimilierung über die Atmosphären-Land Schnittstelle hinweg stabilisieren.

Als letzten Punkte führe ich zwei neue, theoretische, Ansätze ein, um solche Beobachtungsmerkmale automatisch mit maschinellem Lernen zu finden. In meinem ersten Ansatz zeige ich, dass der merkmal-basierte Ensemble-Kalman-Filter unter dem Deckmantel des kernbasierten Ensemble-Transform-Kalman-Filter generalisiert werden kann. Hierbei lernt die Datenassimilierung automatisch die wichtigsten Beobachtungsmerkmale auf Basis der Ensemble Daten und einem gewählten Kern. In meinem zweiten Ansatz, zeige ich, dass Parameter des Ensemble-Kalman Filters mit variationsgetriebenen Bayesianischen Methoden erlernt werden können. Mit dieser Bayesianischen Methode kann die gesamte Wahrscheinlichkeitsverteilung der Parameter herausgefunden und so Unsicherheiten, innerhalb dieser, dargestellt werden können. Zusätzlich beweise ich, dass der Ensemble-Kalman-Filters eine spezielle Lösung dieses Ansatze im linear-Gaussischen Fall ist. Als Konsequenz, deute ich an, dass wir die Beobachtungsmerkmale durch neuronale Netzwerke ersetzen können, die mit Hilfe dieses Ansatze erlernt werden. Von daher beweisen diese beiden Ansätze, dass Entwicklungen im maschinellen Lernen dafür genutzt werden können, um Datenassimilierungsmethoden zu erweitern und möglicherweise zu verbessern.

Contents

1
Introduction

On December 26th, 1999, the cyclone Lothar reached France and southern Germany and left a trail of destruction. One day before, the then newly developed weather prediction model of the German Weather Service missed its formation, because observations from a single, restarted, radiosonde at the coast of Newfoundland were used during the initialization of the model at the wrong time (Wergen and Buchhold, 2002). This example highlights the importance of correctly initializing weather prediction models. However, this mistake occurred 20 years ago, and initialization methods for weather prediction models have significantly improved since then.

Now, let's turn to the future of weather prediction in the next ten years. Will weather prediction improve even more in this time frame? The pragmatic answer is yes, the prediction will steadily improve at the same rate as in the past decades, and we will gain roughly one day of accuracy in our weather prediction compared to today (Bauer et al., 2015). The last innovation that led to one day of accuracy was the introduction of ensemble prediction systems over the course of the last decade. What will be then the next innovation that provides us with an additional day of accuracy? Most likely, this next innovation will be the use of fully-coupled Earth system models for weather prediction together with machine learning ("ECMWF Strategy 2021-2030" 2021). However, for this innovation, we do not only have to improve forecast models, but also to take better advantage of available observations when initializing these models. In this thesis, I make my contribution to gain the next day of accuracy. I propose two complementary frameworks to initialize Earth system models for weather prediction by using the interaction at the interface between the land surface and the atmosphere as an exemplary case.

As my first framework, I propose to unify and couple the initialization of Earth system components using data assimilation and localized ensemble Kalman filters. Such a localized ensemble Kalman filter allows me to use observations from a well-observed Earth system component to initialize another, less-observed, component. In my second, subsequent, framework, I build the theoretical foundations for fingerprint operators based on machine learning and introduce this concept into coupled data assimilation for Earth system models. These operators condense the information from observation into observational features and can simplify in such a way the data assimilation problem across Earth system interface.

1.1 Initializing Earth system models

Earth system models contain specialized modules for the representation of each Earth system component, such as the atmosphere, the ocean, the land, and the cryosphere. In the Earth system, processes interact with each other across different interfaces. Hence, in coupled Earth system models, the modules are run interactively during the forecast. In this way, the coupling in these models improves weather prediction up to seasonal scales (Brunet et al., 2015). However, the quality of the weather prediction does not only depend on the quality of the model, but also on the initial conditions. Caused by the chaotic nature of the Earth system, small errors in the initial conditions can have a large impact on later forecasts (Lorenz, 1963). As a consequence, the initial conditions with which the model starts need to be as good as possible too.

Components of the Earth system are differently well observed. The atmosphere, for example, is one of the better observed components, whereas the land is one of the less-well observed ones. As processes interact across interfaces, information is propagated from one component into another component. On this basis, we could expect that our large amount of observations from the atmosphere also contains information from other components. Nevertheless, this cross-compartmental information is difficult to disentangle, and one of the main challenges in initializing Earth system models is the initialization of less-observed components with observations from the atmosphere.

The initialization problem brings us to the central topic of this thesis – data assimilation. On the one hand, we have physical knowledge in form of a numerical model that tells us something about the spatial and temporal connections in the system, but nothing about the current state. On the other hand, we have observations that imperfectly represent the current state, but do not cover each point in space and time. In data assimilation, we want to combine these two sources of information to distill the full current system state trajectory. As starting point for this distillation, we use the numerical model to create a short-term forecast, the so-called *prior forecast*. Afterwards, we correct this prior based on one given set of observations to a so-called *analysis*. The analysis is an average of the prior and the observations, weighted by their uncertainties (Kalnay, 2003; Law et al., 2015; Asch et al., 2016). Initialized with this estimated analysis, we run a new short-term forecast that will again corrected with observations, as schematically shown in Fig. 1.1. By cycling between short-term forecasts and correction, we ideally nudge the state trajectory to the true, but usually unknown, conditions of the system. However, as trivial as it sounds, data assimilation is not straightforward to apply and many challenges related to the combination of observations and Earth system models remain.

One of the main problems for data assimilation are the unresolved processes in Earth system models. This is even the case in recent developments towards a digital twin of Earth (Bauer et al., 2021a,b). This digital twin should replicate the state trajectory of our physical world in a simulated, virtual, world. For this

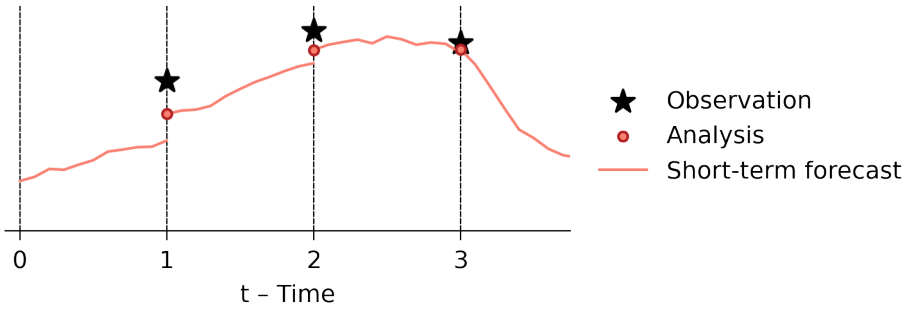


Figure 1.1: A simplified and schematic view on the data assimilation cycle. Data assimilation creates an analysis as a corrected short-term forecast based on a given observation. This analysis is the average of the forecast and the observations, taking the uncertainties of both into account. As a simplification, these uncertainties are not shown in this figure. The analysis is then used to initialize a new short-term forecast, which is then corrected again by an observation one time step later.

replication, the digital twin needs a global horizontal resolution of around 1 km to resolve convection-permitting processes (Palmer and Stevens, 2019; Schär et al., 2020; Wedi et al., 2020). Nevertheless, many other processes act on scales below this threshold of 1 km and, hence, remain unresolved. The effects of these unresolved processes on resolved processes have to be parametrized. These parametrizations result into approximations and simplifications compared to the processes in nature, causing model errors and biases. These model biases are very difficult, if not impossible, to identify, to quantify, and to modify in a data assimilation system.

To minimize the negative impact of these model biases on the analysis, we have to simplify the procedure with its two ingredients. First, we have the numerical model, representing the current physical knowledge of the system. Here, for example, we could limit the number of represented processes within the model, or we could deactivate whole modules that mirror one of the Earth system components. Secondly, we have the data assimilation method, combining observations with model forecast. Here, for example, we could reduce the number of assimilated observations. As weather centers usually want to take advantage of as much physical knowledge as possible for operational forecasting, they use the approach where as many processes as possible are represented within the model. Thus, the data assimilation methods have to be somehow simplified to allow operational forecasting.

As one of these simplifications for data assimilation, interactions between the Earth system components are minimized during the initialization of Earth system models. Every component of the system is initialized by itself, decoupled from the initialization of the other components, as is schematically shown in Fig. 1.2. Nevertheless, observations from one Earth system component are also influenced by other components. This discrepancy between the decoupled initialization and the coupled nature of observations and forecast models prohibits a physically-consistent initialization of Earth system models. This means that the forecast

models are not initialized with the best method available. As my contribution for a physically-consistent initialization of Earth system models, I therefore show that observations from the atmosphere can be assimilated across these interfaces with an ensemble Kalman filter.

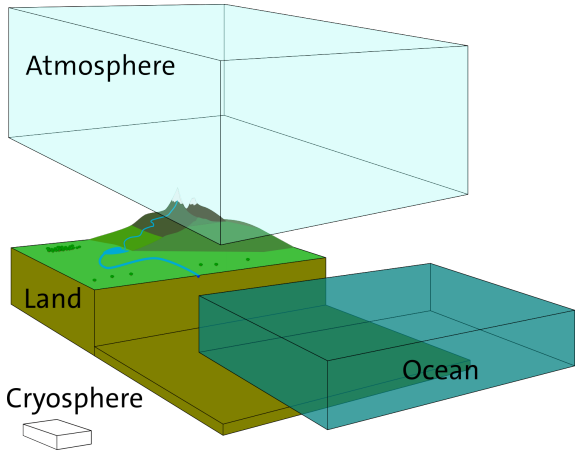


Figure 1.2: A schematic view on the decoupled approach for initializing Earth system models. In operational numerical weather prediction, every here-shown component of the Earth system is initialized independently, whereas the models for the components are interactively run for forecasting.

The interface between atmospheric boundary layer and land surface is one of the few places with operational cross-compartmental data assimilation (Brunet et al., 2010). At this interface, the sensible heat flux and evapotranspiration couple the land surface to the atmospheric boundary layer (Fig. 1.3). These fluxes are driven by the amount of incoming solar radiation at the surface, and hence they are also modulated by clouds. Because their representation in modules for the atmospheric component is inaccurate (Stevens and Bony, 2013; Schär et al., 2020), clouds are one of the main causes for errors in the coupling between the land surface and the atmosphere. In addition, important processes at the land surface act on scales below 1 km and remain unresolved in the current generation of Earth system models (Dirmeyer et al., 2017; Kauffeldt et al., 2015; Orth et al., 2017; Best et al., 2015). These unresolved processes lead to systematic errors, so-called biases, especially in operational land surface models.

To resolve these small-scale processes, recent fully-coupled terrestrial system models (Fatichi et al., 2016; Prein et al., 2015; Vereecken et al., 2016) apply a process-based modelling approach for the atmosphere-land interface. Typically, these models represent the water and energy in layers through a layer-specific soil moisture and soil temperature, the main variables of interest in such a model. With their process-based and layered approach, these models can close the water and energy balance at the land surface and within the soil. As the atmosphere and land are coupled in these models, they can simulate the full water and heat transport from the soil and land surface into the atmosphere and vice-versa. Because of their physical consistency, these models seamlessly scale from continental-scales (Kollet et al., 2018) up to metre-scales in soil (Gebler et al., 2017). In this thesis,

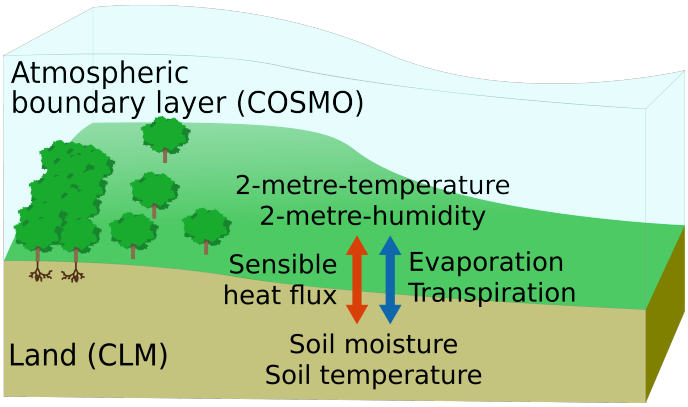


Figure 1.3: A schematic coupling between atmospheric boundary layer and land surface as seen by the Terrestrial System Modelling Platform TerrSysMP with its involved models COSMO and Community Land Model. At this atmosphere-land interface, the sensible heat flux and evapotranspiration couple the soil moisture and soil temperature to the 2-metre-temperature and 2-metre-humidity.

such a model provides me with an useful tool to simulate the interaction between atmosphere and land.

The 2-metre-temperature in the atmospheric boundary layer is locally influenced by the soil moisture (Mahfouf, 1991; Entekhabi et al., 1996; Santanello Jr. et al., 2019; Liu and Pu, 2019), As the soil moisture regulates the sensible heat flux and surface temperature during day-time and influences in this way the 2-metre-temperature. Hence, by observing the 2-metre-temperature, we also indirectly observe the soil moisture. We would therefore expect that by assimilating 2-metre-temperature observations into land surface models, we would also improve the soil moisture analysis.

In reality, the 2-metre-temperature has a negative impact on the soil moisture analysis in operational forecasting (Hess, 2001; Drusch and Viterbo, 2007; Draper et al., 2011; Su et al., 2013; Carrera et al., 2019; Muñoz-Sabater et al., 2019). Unresolved processes and model biases in both, the land surface and the atmospheric boundary layer, heavily impact the model representation of the 2-metre-temperature. Hence, forecast errors in the 2-metre-temperature can be completely unrelated to forecast errors in the soil moisture. Land surface data assimilation nevertheless nudges the predicted 2-metre-temperature to the observed 2-metre-temperature by updating the soil moisture. In the case of unrelated errors, the update of the soil moisture is then simply used as a sink term for errors in the atmospheric boundary layer. Thus, we often only improve the forecast of the atmospheric boundary layer, whereas negatively impacting the soil moisture analysis. In contrast to these previous results, I prove that it is possible to improve the analysis of both components at the same time when the right tools are available.

In operational data assimilation for the atmosphere, ensemble Kalman filters, four-dimensional variational data assimilation methods, or hybrid methods are the state-of-the-art (“IFS Documentation CY47R1 - Part II: Data Assimilation” 2020;

Gustafsson et al., 2018; Schraff et al., 2016; Milan et al., 2020; Lorenc and Jardak, 2018; Wang and Lei, 2014; Kwon et al., 2018). These data assimilation methods take advantage of the full spatio-temporal information during their correction step, and they can take observations at their observational positions into account. Contrary to these advanced methods, unresolved processes and model biases at the land surface force weather centers to still use simplified land surface data assimilation methods (Duan et al., 2019; Fairbairn et al., 2019; Shahabadi et al., 2019).

The simplified land surface data assimilation methods are usually based on the simplified extended Kalman filter (Hess, 2001; Rosnay et al., 2013; Mahfouf et al., 2009; Gómez et al., 2020; Xia et al., 2019; Dharssi et al., 2011; Mucia et al., 2020) or on one-dimensional ensemble Kalman filters (Bonan et al., 2020; Fairbairn et al., 2015; Carrera et al., 2015). Such one-dimensional methods consider only a vertical information flow between the 2-metre-temperature observations and the soil moisture during the correction step. As a consequence, the observations need to be interpolated to the horizontal grid of the land surface model (Shahabadi et al., 2019; "IFS Documentation CY47R1 - Part II: Data Assimilation" 2020). This additional step increases uncertainties in the interpolated observations. The caused noise might overshadow the signal that comes from the soil moisture within the observations. This would reduce the information content for the soil moisture analysis.

To correct the soil moisture with 2-metre-temperature observations, we need to estimate the cross-compartmental sensitivity from the observations to the corrected state variables. In operational land surface data assimilation, these sensitivities are estimated with external model forecasts ("IFS Documentation CY47R1 - Part II: Data Assimilation" 2020; Milbrandt et al., 2016; Gómez et al., 2020) by a finite-differences' or an ensemble approach. Because the external model forecasts are independent from the corrected model forecast, this simplification can lead to physically-implausible or inconsistent sensitivities. These physically-implausible or inconsistent sensitivities can lead to a wrong correction of the soil moisture.

These simplifications for land surface models can prevent in such ways a successful use of 2-metre-temperature observations for the soil moisture analysis. In addition, different methods are used for data assimilation in the atmosphere, leading to a discrepancy in the methods between atmosphere and land surface. To initialize the land surface model together with the atmospheric model, as needed for a physically-consistent initialization of Earth system models, we would need to use the methods for data assimilation in the atmosphere and land surface.

In this thesis, I investigate new possibilities for coupled data assimilation in Earth system models to possibly unify the data assimilation systems. To do so, I use the example of assimilating 2-metre-temperature observations across the atmosphere-land interface. Contrary to the approach in operational forecasting, I simplify my model configuration (see also Chapter 2) to be able to experiment with advanced data assimilation methods. I conduct my experiments by simulating the interac-

tions between atmosphere and land with the limited-area Terrestrial Modelling Platform (TerrSysMP, Shrestha et al. 2014; Gasper et al. 2014). With this model environment, I generate a nature run, a simulation without data assimilation. I define this nature run as my reality and synthesize hourly and sparsely-distributed 2-metre-temperature observations based on its trajectory. I assimilate then these synthetic observations into the soil moisture and atmospheric temperature in different data assimilation experiments, conducted based on the same model configuration as for the nature run. By using this approach, I avoid model errors and uncertainties caused by parametrizations within the model. Hence, this well-controlled environment yields clear indications about the performance of the considered data assimilation methods for coupled data assimilation in a "perfect world" case.

In this context, I compare a simplified extended Kalman filter to a localized ensemble Kalman filter, as similarly used for data assimilation in the atmosphere (for a derivation of the ensemble Kalman filter see also Chapter 3). This advanced data assimilation approach together with the coupled terrestrial system model and idealized experiments provides a unique opportunity for improvements in land surface data assimilation. For the first time, I show that 2-metre-temperature observations can be directly assimilated across the atmosphere-land interface without the need for any intermediate interpolation step (see also Chapter 4). By assimilating these observations at their original positions, I prove that they have a previously unclear potential to improve the analysis in both, the atmosphere and the land surface, at the same time. This allows me to merge the decoupled cycles for the atmosphere and the land surface into one single data assimilation cycle with hourly correction steps. As I merge their data assimilation cycles, I initialize both Earth system components together in a physically-consistent way. Therefore, I propose as my first framework to unify and couple the data assimilation in Earth system models with a localized ensemble Kalman filter.

1.2 Ensemble Kalman filtering

As the Earth is a chaotic system, small initial deviations are exponentially amplified over time. This error amplification cannot be represented by one single deterministic forecast. Thus, a popular approach in weather prediction is the ensemble forecast. Instead of initializing one single forecast, a whole bundle of forecasts is started, all with their own slightly different initial conditions. Because of the exponential amplification, the forecasts deviate more and more from each other over time. In this way, they provide an estimate for the uncertainty in the system. This uncertainty estimate evolves over time and goes literally with the flow in the system. It therefore ideally represents the true uncertainties in the system that dynamically depends on the current state of the system.

This ensemble approach makes ensemble Kalman filters easier to implement than other popular methods in atmospheric data assimilation (Kalnay et al., 2007a; Bannister, 2017) such as four-dimensional variational data assimilation. The flow-dependent uncertainty estimate from the ensemble can be directly used for a

time-dependent weighting in the data assimilation cycle (Lorenc, 2003; Hamill and Snyder, 2000). As the coupling strength of cross-compartmental interfaces in the Earth system depends on the current state of the system, time-dependent weighting is especially important for coupled data assimilation. Data assimilation systems with a climatological uncertainty estimate have difficulties to represent this time-dependency (Lin et al., 2017; Lin and Pu, 2018; Frolov et al., 2016; Smith et al., 2017). In contrast, the estimate evolves on the basis of the coupled forecasting models with an ensemble approach. Consequently, the ensemble estimate includes all important cross-compartmental process that are also represented within the models.

The ensemble Kalman filter additionally solves the problem of estimating the sensitivities from the assimilated observations to the state variables (Evensen, 1994; Evensen and Leeuwen, 1996; Burgers et al., 1998; Houtekamer and Mitchell, 1998; Hunt et al., 2007). Normally, the modelled state variables are discretized on a grid and averaged over a whole grid box. On the opposite, observations are represented by positions of their observational sites. The assimilated observations might thus have no direct equivalent among the state variables. For the correction step, we have nevertheless to estimate the sensitivity of the state variables to the observations. To estimate the sensitivity, non-linear observation operators are used to translate the state variables into their observational equivalent. The sensitivity is then given as way back from the observations to the state variables. A non-linear sensitivity is difficult to represent in data assimilation and makes an analytical solution for the analysis nearly impossible.

As one solution, the sensitivity is represented by an additional adjoint model in variational data assimilation methods. On this basis, the analysis is iteratively searched as a deterministic and constrained problem. Because the adjoint model has to be developed and maintained (Bannister, 2017), variational data assimilation methods are difficult to implement for Earth system models. As another solution, ensemble Kalman filters represent the current state of the system as the mean of the ensemble forecast. By linearizing the observation operator around this ensemble mean, ensemble Kalman filters circumvent the problem of non-linear sensitivities and analytically solve the analysis based on the prior ensemble. In this way, we do not have to develop and maintain an adjoint model. Hence, ensemble Kalman filters are a linearized and approximative solution to the data assimilation problem with non-linear observation operators.

A challenge for ensemble Kalman filters are temporal-dependencies of observations to state variables of interest. These temporal dependencies emerge because Earth system components are temporally dependent on each other. As an example, information from the land surface into the atmospheric boundary layer is propagated by the sensible heat flux and evapotranspiration with a time lag. In an ensemble Kalman filter, only observations from the same time as the forecast are assimilated during the correction step, as schematically shown in Fig. 1.4. Thus, time-dependencies between components are only simulated by the propagation in the short-term forecast and not considered during the correction step. This

limitation inhibits observations from unfolding their full assimilation impact for the cross-compartmental initialization of Earth system models.

To take temporal dependencies into account, it is theoretically advantageous to assimilate observations not at a single time but in a given assimilation window, as schematically shown in Fig. 1.4. By assimilating observations within a window, a smoothing method targets trajectories spanning the whole time period, instead of filtering a forecast at a single time step. Hence, the ensemble Kalman smoother can take future observations into account and make use of temporal dependencies between observations and trajectory that should be corrected. By utilizing temporal dependencies between components in the Earth system, these smoothing methods are a promising way to extract more information out of existing observations for coupled data assimilation. I show with an ensemble Kalman smoother that this advantage is not only theoretically the case, but also leads to improvements in the soil moisture analysis.

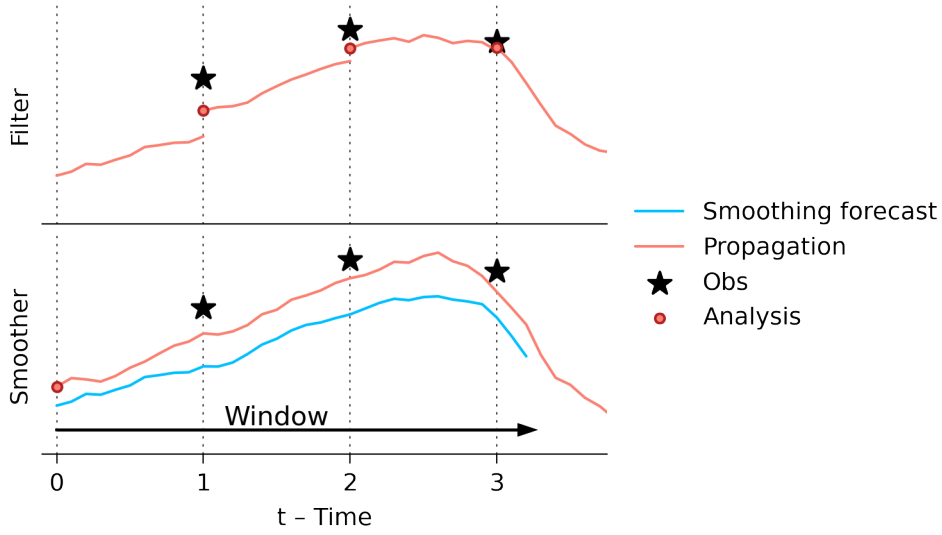


Figure 1.4: A simplified and schematic view on the difference between ensemble Kalman filters and ensemble Kalman smoothers. In filtering, a short-term forecast is corrected to an analysis at each observational time. The filtered analysis is propagated to the next time step. In smoothing, one longer forecast trajectory is propagated up to the end of the assimilation window. Based on this forecast, all observations within the window are used to correct the conditions at the beginning of the window. These conditions are propagated a second time through the assimilation window to get the corrected and smoothed trajectory.

Another challenge for ensemble Kalman methods is the limited number of ensemble members. Only a limited ensemble size is affordable, because the ensemble-based short-term forecast is the most expensive step in these methods. The number of ensemble members upper-bounds the represented degrees of freedom in an ensemble. As a consequence, the limited number of members causes multiple problems in ensemble Kalman methods 1) the ensemble approximation induces spurious correlations between state variables and observations, 2) the sampling of ensemble members causes sampling errors in the uncertainty estimate. Because of these sampling errors, the estimated uncertainty is likely too small to represent the

true uncertainty of the system, leading to an underestimation of the observational impact on the data assimilation. To tackle these problems, we need techniques like localizing the influence of observations on the state variables (Houtekamer and Mitchell, 1998, 2001; Ott et al., 2003; Hunt et al., 2007) or an artificial inflation for the ensemble uncertainties (Anderson and Anderson, 1999; Mitchell and Houtekamer, 2000). These tricks make ensemble Kalman methods feasible for operational forecasting.

The limited number of ensemble members also constrains the usable information from observations (Tsyrulnikov, 2013; Hotta and Ota, 2020). This problem can be logically understood from a weighting point of view in data assimilation. For each grid point and observation, we estimate one weighting factor. If more observations are assimilated, then also more weighting factors are needed. Because the degrees of freedom are upper-bounded, the ensemble does not provide enough information content to estimate the weighting factors. Because of their regulative character, ensemble Kalman methods reduces their weights for single observations. As a consequence, single observations have a lowered observation impact. To make more out of the existing observations, we have thus to reduce the assimilated quantity of information.

To reduce the assimilated quantity of information in operational forecasting, the assimilated observations can be thinned out (Hamrud et al., 2015). Another approach is to constrain the number of observations that are considered for the analysis at a single grid point (Schraff et al., 2016). However, these are only heuristic tricks to reduce the assimilated quantity of information; they do not solve the problem in a principle way.

Another concept to reduce the quantity of information is to use features of observations (Morzfeld et al., 2018). The idea is that observations have less degrees of freedom than there are observational points. Let's take for example a satellite image: the image itself has $\mathcal{O}(10^6)$ pixels, but only 100 different objects are visible within the image. If we build an object detection algorithm, we can condense the information from $\mathcal{O}(10^6)$ pixels into $\mathcal{O}(100)$ features, representing the objects and their position. Instead of assimilating the raw pixels, we would then assimilate these features. Thus, feature-based data assimilation allows us to reduce the assimilated information content.

Feature-based data assimilation can also help for data assimilation across interfaces. Observations from one Earth system component could have characteristic fingerprints that point toward errors in another component. To understand this fingerprint concept, let's stick again to an exemplary case: during daytime, the soil moisture modulates the diurnal cycle of the 2-metre-temperature via the sensible heat flux. However, the 2-metre-temperature is typically measured on an hourly basis, where it exhibits fluctuations that might be not related to the soil moisture. If we assimilate the hourly measured 2-metre-temperature into a coupled forecast model, we would also assimilate noisy fluctuations, which could overshadow the signal coming from the soil moisture. Instead, we can design features that are

representative for the diurnal cycle of the 2-metre-temperature. These features can then filter out unrelated fluctuations such that only the signal related to the soil moisture remains. In this way, we can make data assimilation more robust against noise in the ensemble and observations. Feature-based data assimilation can be therefore a general procedure for coupled data assimilation in Earth system models.

Building upon my first proposed framework of coupled data assimilation, I assimilate 2-metre-temperature observations with an ensemble Kalman smoother and a 24 hour window (see also Chapter 5). In my idealized experiments, I compare the performance of such an ensemble Kalman smoother to two different ensemble Kalman filters, where the analysis is conditioned to current and past observations. My results prove that ensemble Kalman smoothers can take temporal dependencies of the atmospheric boundary layer on the land surface into account. As a result, they further improve the soil moisture analysis compared to ensemble Kalman filters.

I introduce the novel concept of feature-based data assimilation into data assimilation across interfaces by using fingerprint operators (see also Chapter 5). I design two fingerprint operators for 2-metre-temperature observations in a 24 hour window to correct errors in the soil moisture forecast. In these fingerprint operators, I take advantage that the soil moisture modulates the diurnal cycle of 2-metre-temperature observations. These introduced fingerprint operators condense the information from 24 2-metre-temperature into two features. By assimilating these features, I obtain similar results for the soil moisture analysis as using the raw observations in an ensemble Kalman smoother. Therefore, I propose as my second framework to use such a feature-based ensemble data assimilation for the initialization of Earth system models to make more out of the existing observations across cross-compartmental interfaces.

As it can be time consuming to define such features for every single problem in Earth system models, I introduce two novel approaches to define these features based on data-driven methods (see also Chapter 6). In machine learning, linear regression or linear classification methods are applied on explicitly extracted features from raw data (Hastie et al., 2009; Murphy, 2012) to improve the linear methods. Feature-based data assimilation with fingerprint operators can be seen in the same way. I extract possibly non-linear features with explicitly defined fingerprint operators from observations in the atmosphere. Afterwards, I linearly assimilate these features with ensemble Kalman methods into the land surface. In my two additional approaches, I generalize this idea of feature extraction as first step and applying a linear data assimilation in a second step.

As first approach, I derive a novel generalization of fingerprint operators in ensemble Kalman methods by kernel methods. In these kernel methods, the linear core mechanic of creating similarities between vectors is replaced by a kernel (Schölkopf and Smola, 2002; Rasmussen and Williams, 2006; Murphy, 2012), which implicitly represents the assimilated feature space in a reproducing

kernel Hilbert space. In data assimilation terms, I automatically construct the feature extractor based on the ensemble data by choosing a specific kernel. As second approach, I bring in deep-learning-based methods (LeCun et al., 2015; Goodfellow et al., 2016) on the basis of variational Bayes (Jordan et al., 1999; Beal, 2003; Hinton and van Camp, 1993) as pre-processing step for ensemble Kalman methods. I prove the ensemble transform Kalman filter (Bishop et al., 2001; Hunt et al., 2007), used throughout this thesis, as a special solution of variational Bayes. This allows me to use variational Bayes as general method to tune parameters in the ensemble Kalman filter in a consistent Bayesian way. As this general method can fit parameters with stochastic gradient descent and differentiate through the data assimilation procedure, it can be additionally used to replace hand-crafted observational features by a multi-layered neural network and learn the network on the basis of training data. In this case, the neural network could learn observational features from past data and we could use the network for non-linear data assimilation with ensemble Kalman methods. Based on these two approaches, I show in this thesis a way to integrate data-driven learning into data assimilation methods and especially ensemble Kalman methods.

2

Idealized twin experiments for the atmosphere-land interface

In this Chapter, I present my idealized setup that allows me to study the interface between the atmospheric boundary layer and the land surface. This idealized setup is based on so-called twin experiments, on which I elaborate more in detail in Section 2.1. Throughout this thesis, I use the Community Land Model as land surface model and COSMO as atmospheric model in a Platform called TerrSysMP; their components, coupling, and configuration are shown in Section 2.2. Additionally to the explained model configuration, I construct an ensemble of initial states that are used as initial conditions for my runs. I show how I construct this ensemble based on a spun-up run in Section 2.3. I define a single model run as my reality, which is subsequently called nature run. This run and its weather conditions are presented in Section 2.4. Based on this run, I generate synthetic 2-metre-temperature observations at 99 observational sites, which are unfolded in Section 2.4.2.

2.1 What are my idealized twin experiments?

In real-world data assimilation, we do not know the true evolution of states in the system. Our only source of information are disturbed representations of these unknown true states, in form of observations with an observational error and model simulations with a model error. Because of these errors, the performance of data assimilation has to be evaluated against proxy data. This proxy data is often a non-assimilated set of observations or a different data assimilation product from another weather center. Because the performance of different data assimilation methods heavily depends on this proxy data, it is difficult to compare different data assimilation methods.

Intractable factors and errors influence observations and model. This problem especially prevails in the atmospheric boundary layer and at the land surface. Here, processes acting at small-scales have a large impact on the temporal development of the cross-compartmental interface. Nevertheless, the resolution of our model is too coarse to resolve these processes. As a consequence, these unresolved processes cause model errors that lead to biases in the representation of the atmospheric boundary layer and the land surface. These small-scale processes additionally influence observations from the atmospheric boundary layer, like 2-metre-temperature observations. Because these processes are unresolved, it is difficult to construct a model equivalent of atmospheric boundary layer observa-

tions. If a model equivalent is found, then its observational errors are often large so that single boundary layer observations have only a low impact on the data assimilation.

To circumvent the problems of an unknown true state and intractable errors, I employ idealized twin experiments. In these twin experiments, I conduct a single, deterministic model run that I define as my reality, the so-called nature run. Based on this known reality, I extract synthetic observations with a perfectly-known observation operator and a pre-defined observational error. Hence, I exactly know the observational error distribution, which would not be the case for real-world observations. I additionally use the same model configuration for my experiments as for my nature run. In this way, I deactivate any kind of model error that might lower the performance of the data assimilation. As additional simplification in my experiments, I only perturb the initial conditions for the land surface model, whereas I use the same initial and lateral boundary conditions for the atmospheric model. All differences in the experiments are therefore only a result of the perturbed initial conditions in the land surface. With this strategy, I can control and backtrack the uncertainties and errors in the experiments.

I can simply compare my experiments against my nature run without the need of any interpolation or proxy data, because I know the reality on the same grid as for my experiments. This simple verification, together with a control on the uncertainties, allows me to get robust results about the performance of my data assimilation methods on a simulated time period of seven days. On the opposing side, I simplify my experiments compared to real-world data assimilation experiments. I use a perfect model without any model error in my experiments. As a consequence, it is difficult to draw conclusions about the performance of my data assimilation methods in the case of imperfect models. In addition, I perfectly know the true state and hence also the synthetic observations in these experiments. Therefore, it is difficult to make general statements about the robustness of my data assimilation methods in the case of only partial knowledge. All in all, my idealized twin experiments are a convenient way to compare data assimilation methods without any external source of noise, but the results with these experiments are in some sense only proof of concepts. They are not yet validate in real-world data assimilation.

2.2 TerrSysMP

The Terrestrial System Modelling Platform (TerrSysMP, Gasper et al. 2014; Shrestha et al. 2014) is a coupled atmosphere-land modelling system. This coupled modelling system simulates in a physically consistent way the transport of water and energy from the land surface into the atmosphere and vice versa.

TerrSysMP includes three different models: 1) the COnsortium for Small-scale MOdelling model (COSMO) for the atmosphere, 2) the Community Land Model (CLM, Oleson et al. 2004; Oleson K. W. et al. 2008) for the land surface, and 3) ParFlow for the sub-surface flow within the soil. In this thesis, I concentrate

only on the relationship between atmospheric boundary layer and land surface. I therefore omit ParFlow for my experiments and solely use CLM, Version 3.5, to simulate processes in the land surface and soil. In this column-based land model, soil and vegetation processes are represented by vertical fluxes of moisture and heat in form of the soil moisture and soil temperature.

COSMO (here-used as Version 4.21) as non-hydrostatic mesoscale weather model evolved out of the non-hydrostatic limited-area model (Steppeler et al., 2003). This convection-permitting model was operationally used at the Deutscher Wetterdients (DWD, Germany Meteorological Service, Baldauf et al. (2011)) for many years. It is continuously used at other weather services (e.g. MeteoSwiss) with horizontal resolutions of up to 1.1 km. Furthermore, COSMO can model processes in the atmospheric boundary layer for up to sub-kilometre-scales (Finn et al., 2020).

In my setup, CLM is coupled to COSMO via the OASIS3 coupler (Valcke, 2013) and acts as lower boundary condition for COSMO. This lower boundary condition is represented as heat and momentum flux from the land surface into the atmospheric boundary layer. In addition, the ground temperature and surface specific humidity in COSMO is replaced by their counterparts from CLM. For the calculation of the radiative fluxes in COSMO, the direct and diffuse albedo and the outgoing longwave radiation is transmitted from CLM to COSMO. As CLM needs atmospheric forcing data, temperature, wind speed, specific humidity, and pressure from the lowest model level is send from COSMO to CLM. To close the water balance in CLM, the convective and grid-scale precipitation fields are gathered from COSMO, whereas the incoming short- and long-wave radiation is used to close the surface energy balance in CLM.

The horizontal grid in CLM is based on a Cartesian coordinate system, whereas COSMO uses a rotated pole system. Furthermore, processes at the land surface act on smaller scales than in the atmosphere. Hence, I use CLM with a horizontal resolution of 1 km, whereas COSMO has a coarser resolution of 2.8 km. As a consequence, the fields have to be interpolated before they can be exchanged between CLM and COSMO. The information coming from CLM and going to COSMO is interpolated by a distance weighted nearest neighbor interpolation, whereas a bilinear interpolation method is used for the information flow from COSMO to CLM, as described by Shrestha et al. (2014).

I use COSMO with 50 full vertical levels. Hence, my configuration of COSMO is similar to the configuration that was used for the operational COSMO-DE runs at the DWD until 2018. CLM has a fixed number of 10 vertical soil levels. On every level, the tendency for water and heat are estimated based on the state in the level above and below the current level. An additional unconfined aquifer stores the ground water in CLM and is the lower boundary condition for CLM. In this thesis, I do not use the 5 additional snow levels in CLM, because my simulations are for a mid-latitude summer period without any snow and ice. The model orographies for COSMO and CLM are based on the European digital elevation model (Figure 2.1, European Environment Agency 2013). The model area has its origin on the Neckar

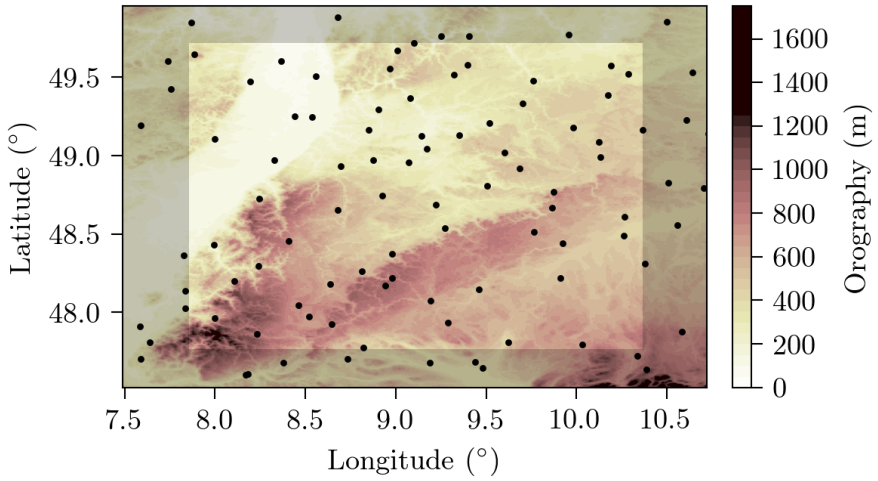


Figure 2.1: The model orography, extracted from CLM, is based on the European digital elevation model (European Environment Agency, 2013). The black points are representing the 99 measurement sites, used for synthetic observations of the 2-metre-temperature.

catchment in Baden-Württemberg and spans a region of ~ 300 km in latitudinal direction and ~ 280 km in longitudinal direction.

The main surface type is defined as single surface type per grid point in CLM. To further simplify the setup, the plant functional types (PFTs) are restricted to a single PFT per grid point. Over the whole area seen, the land use is restricted to five different plant types: 1) broad-leaf forests, 2) needle-leaf forests, 3) grassland, 4) cropland, and 5) bare soil. The land use is classified with the CORINE Land Cover inventory (Keil et al., 2011). Water surfaces, like lakes or rivers, and urban areas are represented as bare soil. For the four plant types, without bare soil, the plant leaf area index (LAI) is calculated based on MODIS (Myneni et al., 2002) and post-processed for bias correction. Based on the LAI, the steam area index is estimated following Lawrence and Chase (2007) and Zeng et al. (2002). For CLM, the additionally needed soil types are extracted and interpolated from the soil map of Germany ('Bodenübersichtskarte', Bundesanstalt fuer Geowissenschaften und Rohstoffe 2016).

As time step in COSMO, I use a time step of 10 s. To satisfy the Courant-Friedrich-Levy criterion (CFL), the wind speed should be as a consequence not faster than 280 m s^{-1}. Since these wind speed values are hardly reached in the atmosphere, even in the polar jet stream, this time step satisfies the CFL criterion. For CLM and for the coupling rate in OASIS3, I apply a longer time step of 90 s, representing that processes within the land surface act on longer time-scales than in the atmosphere.

2.3 An initial ensemble of states

In my experiments, I specifically investigate the sensitivity of the 2-metre-temperature on perturbations in the soil moisture. For this investigation, I reduce the noise coming from other possible sources and perturb only the initial soil conditions

compared to my nature run. All other parameters, like the lateral boundary conditions and the model configuration, are for my ensemble experiments the same as for my nature run. Differences in 2-metre-temperature between different runs are therefore only a consequence of these generated perturbations in the initial soil conditions.

The conditions in the land surface and soil should be synchronized to the climatic conditions in the atmosphere. In addition, processes in soil act on much longer time-scales than processes in the atmosphere. I build my ensemble on the basis of a single, spun-up, run with a similar model configuration and a spin-up of six years. This spun-up run provides the initial conditions for the atmosphere at 2015-07-30 00:00 UTC. The needed lateral boundary conditions for the atmosphere are generated based on one member of the COSMO-DE EPS ensemble from the DWD with the same horizontal and vertical resolution as my COSMO configuration.

Compared to the initial conditions provided by the spun-up run, I disturb the initial soil conditions for 2015-07-30 00:00 UTC. Although I am mainly interested in the relationship between the 2-metre-temperature and the soil moisture, I perturb the soil moisture together with the soil temperature, which induces some additional noise into the atmospheric boundary layer. In initial experiments, I found that only using soil moisture perturbations can lead to very high and physically-implausible sensitivities of perturbations in the soil moisture to perturbations in the 2-metre-temperature. With these implausible sensitivities, data assimilation would make too big update steps, which would have a negative assimilation impact. The additional perturbation of the soil temperature stabilizes the sensitivities without reducing the signal coming from the soil moisture on the 2-metre-temperature.

For the soil moisture perturbations, I perturb the soil moisture saturation, which is the volumetric soil moisture scaled by the saturation point that depends among others on the soil type, whereas I directly perturb the soil temperature. Additionally, I bound the resulting soil moisture saturation to the physically-plausible area between 0 and 1. As perturbation strengths, I use Gaussian noises with a mean of zero and standard deviations of 0.06 for the soil moisture saturation and a standard deviation of 1 K for the soil temperature, across all layers. For the soil moisture saturation similar values were applied in (Schraff et al., 2016), and in an initial experiment, I found that these values generate a reasonable ensemble spread for the 2-metre-temperature, as shown in Fig. 2.2.

I independently correlate the perturbations for the soil moisture and soil temperature in horizontal and vertical dimensions on the basis of two different Gaussian covariance functions. As horizontal covariance function, I use a truncated Gaussian kernel with a scale of 14 grid points, which equals roughly 14 km, whereas the kernel is truncated after 42 grid points. A similarly truncated covariance function is used in vertical dimensions with a scale of 0.5 m and a truncation after 1 m. These correlation scales are again similar to those used in Schraff et al., 2016, and they generate realistic looking soil moisture patterns.

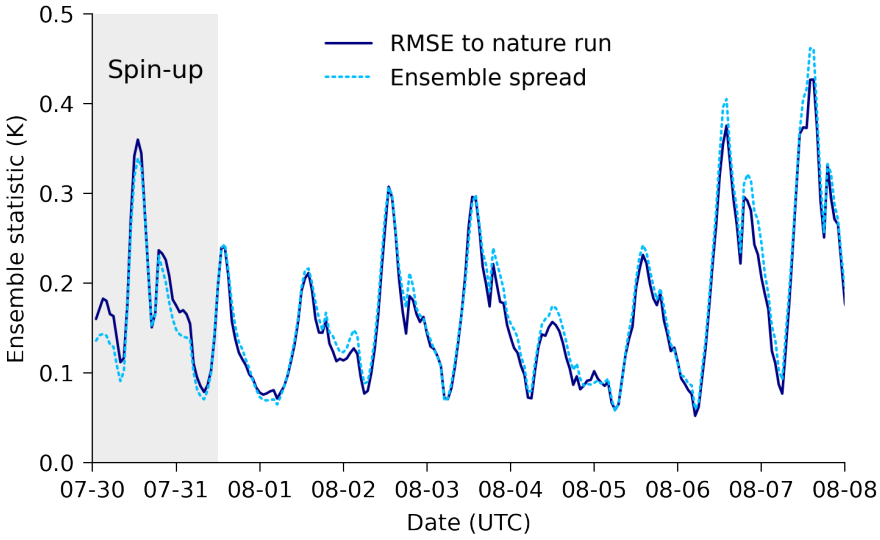


Figure 2.2: The root-mean-squared-error of the ensemble mean compared to the nature run and the ensemble spread, both as the square-root of the area-averaged quadratic statistics, for the 2-metre-temperature in the ensemble without data assimilation. The grey-shaded area indicates the spin-up phase, which is not used for the experiments. The soil conditions as solely driving factor for perturbations in the 2-metre-temperature cause the diurnal cycle in the differences. The shown 2-metre-temperature is a diagnostics quantity in COSMO, interpolated between lowest surface temperature from CLM and lowest mode layer temperature (10 meters height) from COSMO. As a consequence, the perturbed surface temperature drives the initial ensemble spread and forecast error in the 2-metre-temperature.

Based on the Gaussian noise and the Gaussian covariance functions, I generate an ensemble of 40 different initial soil conditions for my data assimilation experiments with 40 ensemble members; the same number of members is operationally used at the DWD for their ensemble data assimilation system, which provides the initial conditions in their ICON-D2-EPS (Reinert et al., 2021). I use the initial soil conditions of a hypothetical 41-th ensemble member as the initial conditions for my nature run. By initializing the nature run in this way, I make the ensemble spread representative for the forecast error of the ensemble mean compared to the nature run, as shown in Fig. 2.2. In theory, there should be no difference if I perturb my nature run or if I center my ensemble around a perturbed ensemble mean. Nevertheless, the soil moisture is bounded between $0\,\mathrm{m}^3\,\mathrm{m}^{-3}$ and the saturation point. As a consequence, there would be small differences between perturbing the nature run or perturbing the ensemble mean. If I recenter my ensemble around a perturbed ensemble mean, I influence the fields of all 40 ensemble members such that there might be edge cases where the soil moisture in all ensemble members collapses to $0\,\mathrm{m}^3\,\mathrm{m}^{-3}$ or to the saturation point. To avoid these edge cases, it is easier for me to perturb the initial conditions of the nature run instead of perturbing the initial conditions of the ensemble mean.

For all runs, including my nature run, I run the first 36 hours of simulation as spin-up phase, until 2015-07-31 12:00 UTC. Within this spin-up phase, I expect that perturbations in the soil conditions near the land surface are propagated into

the atmospheric boundary layer. This propagation should result in considerable differences in the 2-metre-temperature. If the ensemble with its 40 ensemble members is representative for the forecast errors of the ensemble mean, then the averaged ensemble spread should be similar to the root-mean-squared-error of the ensemble mean to the nature run. The comparison in Fig. 2.2 shows that the ensemble is indeed representative for the forecast error. Even more, the perturbations are already propagated to the 2-metre-temperature within the spin-up phase. The 2-metre-temperature is estimated as average between the surface temperature and lowest model layer (in my case 10 metres height) in COSMO, as I will describe in Section 3.5.1. Caused by the initial surface temperature perturbations, also the diagnostic 2-metre-temperature has already perturbations within the initial time step.

After this spin-up phase, starting at 2015-07-31 12:00 UTC, I start with my experiments. Because my data assimilation environment (for more information see also Section 3.5.3) communicates on a file-basis with TerrSysMP, the model is in some experiments restarted hourly. This resets processes in the turbulent kinetic energy schema in COSMO, which can be seen as some kind of model error. To mitigate this possible model error source in all of my experiments, I restart after 2015-07-31 12:00 UTC all runs hourly, including my nature run.

My nature run is together with the other 40 ensemble members the basis for my experiments in Chapter 4 and Chapter 5. Because of its importance, I explain more details about the nature run in the following.

2.4 The nature run

With the previously described configuration, I run my nature run for more than seven days from 2015-07-31 12:00 UTC to 2015-08-08 00:00 UTC. I define the state trajectory of this nature run as my reality. In addition, I synthesize out of this trajectory 2-metre-temperature observations. These synthetic 2-metre-temperature observations are assimilated into my data assimilation experiments. Furthermore, I compare my data assimilation experiments with this state trajectory. In the following, I describe the weather and soil conditions that allow me to do data assimilation across the atmosphere-land interface. Subsequently, I explain how I generate my synthetic 2-metre-temperature observations.

2.4.1 Weather and soil conditions

One of the first requirements to do data assimilation across the atmosphere-land interface is that these two Earth system components are coupled. They are coupled through the sensible heat flux and evapotranspiration. Since these fluxes are mainly driven by the incoming solar radiation at the surface, also the coupling between atmosphere and land is mainly driven by incoming solar radiation.

In addition, perturbations from the soil moisture have their largest impact on the sensible heat flux for soil conditions that are neither too dry nor too moist. If

the soil moisture is too dry and in the region of the welting point, then plants cannot transpire. If, on the other side, the soil moisture is too moist and in the region of the saturation point, then plants have their maximum transpiration. In both cases, differences in the soil moisture have only a small impact on the strength of the transpiration. As a consequence, the sensible heat flux is not influenced by differences in the soil moisture, reducing the sensitivity of the 2-metre-temperature to soil moisture perturbations. My simulated time period with its weather conditions satisfies both conditions as I show in the following (Figure 2.3).

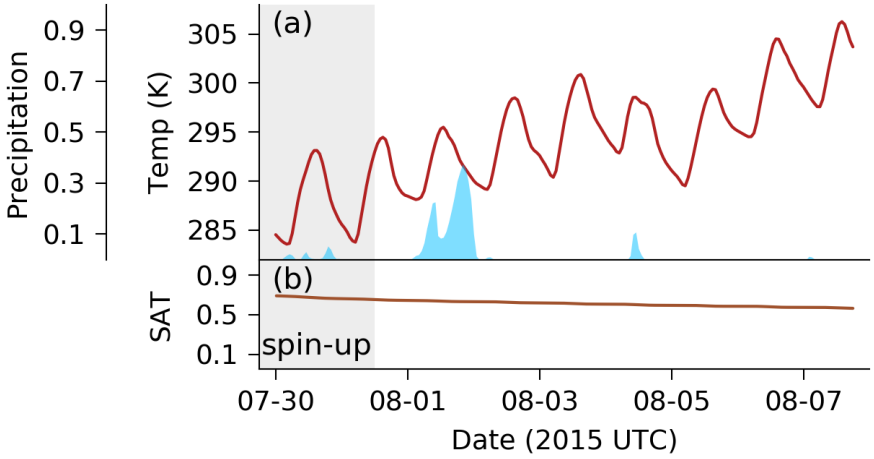


Figure 2.3: Mean weather overview over the simulated time period, extracted from the nature run. The grey-shaded area indicates the spin-up phase, which is not used for the experiments. (a) The hourly development of the 10 m temperature as area mean, while the blue bars show the fraction of grid points with precipitation ($> 0\,\mathrm{kg\,m^{-2}\,h^{-1}}$) in the previous hour. (b) Hourly soil moisture saturation, defined as soil moisture divided by the saturation point, in root-dept h (0.21 m depth) as area mean.

In my simulated nine-day period, the daily mean 2-metre-temperature increases with time, whereas the soil moisture in root-depth decreases. The amplitude of the diurnal cycle in the 2-metre-temperature is further quite large. Together with the soil moisture drying and the temperature increase, this amplitude indicates a period with strong incoming solar radiation at the land surface and without large precipitation events. The only larger precipitation event is on 2015-08-01, whereas a smaller event on 2015-08-04 has the largest impact on the 2-metre-temperature. In addition, the soil moisture is in a mixed regime, as indicated by saturation values around 0.5. This mixed regime indicates a strong sensitivity of the 2-metre-temperature to perturbations in the soil moisture. Based on this weather overview only, I would expect a coupling between atmospheric boundary layer temperature and soil moisture in root-depth during day-time.

2.4.2 Synthetic 2-metre-temperature observations

The nature run with all its weather features is the basis for my synthetic 2-metre-temperature observations that are assimilated into my data assimilation experi-

ments. In the following, I explain how I synthesize these observations out of my nature run.

The observational positions should be as near as possible to true measurement sites to estimate a realistic assimilation impact from 2-metre-temperature observations on the soil moisture. Thus, I use positions from the surface measurement network of the DWD. As a result, I obtain 99 measurement sites in my selected area (marked as black dots in Fig. 2.1). On this basis, I synthesize hourly 2-metre-temperature observations by using the diagnostic 2-metre-temperature fields, estimated in COSMO. I hereby select the 2-metre-temperature at the nearest horizontal neighbor grid point to the real measurement sites. To avoid the need for vertical interpolation in my observation operator, I use the orographic height from COSMO also as measurement height. As a consequence, the scheme perfectly describes the observation operator that I use in my data assimilation experiments to translate the model state into observational space (for more information about this observation operator see also Section 3.5.1).

The only perturbations arise from initial soil conditions in my experiments. This limits perturbations in the atmospheric boundary layer significantly. By using realistic observational errors ($\mathcal{O}(1\,\mathrm{K})$), these errors would overshadow the signal coming from the soil moisture on the 2-metre-temperature observations. I therefore chose observational errors that are one order of magnitude smaller than realistic observational errors in the 2-metre-temperature. I generate them as additive Gaussian white noise with a standard deviation of $\sigma^o = 0.1\,\mathrm{K}$. Because I have a constant standard deviation, my resulting observational covariance matrix $\mathbf{R}$ for the data assimilation is a static diagonal matrix with $(\sigma^o)^2$ as diagonal elements $\mathbf{R} = (\sigma^o)^2\mathbf{I}$.

This page intentionally left blank

3

A localized ensemble Kalman filter for the atmosphere-land interface

In this Chapter, I elaborate about the principles of ensemble Kalman filtering for the atmosphere-land interface, setting the general data assimilation method for the following Chapters. I start with a general derivation of data assimilation from Bayesian and probabilistic principles in Section 3.1. In this thesis, the data assimilation methods are mainly based on the ensemble transform Kalman filter. I therefore derive only this schema in Section 3.2 and Section 3.3, whereas I describe specific deviations from the schema in Chapter 4 and Chapter 5. To get the derived ensemble Kalman filter working, I present inflation and localization in Section 3.4. In addition, an in-deep explanation for my 2-metre-temperature observation operator and the post-processing of the analysis is shown in Section 3.5. There, I present my implementation of the data assimilation in my own-developed software package. In the last section of this chapter, I introduce the concept of offline data assimilation experiments, which I use in the following Chapters. A general overview of the here-used notation is given in Chapter A.1.

The idea of an ensemble Kalman filter is not new. I use here an implementation of the localized ensemble transform Kalman filter (LETKF, Hunt et al., 2007). The derivation therefore mainly resembles the steps in Hunt et al., 2007, where more details about the LETKF are provided. Many scientific advances in data assimilation were previously needed for this development towards a stable localized ensemble Kalman filter.

The ensemble Kalman filter is based on the Kalman filter equations (Kalman, 1960). This Kalman filter gives the optimal solution in the Gaussian-linear case. Geophysical systems have often non-linear dependencies such that states are non-linearly propagated in forecast models. As a consequence, the Kalman filter equations are not directly applicable in geophysical forecast models. Hence, two notably extension of the Kalman filter are the extended Kalman filter (Jazwinski, 1970; Anderson et al., 1979) and the square-root Kalman filter (Battin, 1964; Bierman, 1977; Anderson et al., 1979; Maybeck, 1982). On the one hand, the extended Kalman filter linearizes the non-linearities around a non-linearly propagated Gaussian mean. With this linearization the Kalman filter equations can be applied. On the other hand, the square-root Kalman filter decomposes the involved covariances into their square-roots and propagates these square-roots. The use of square-roots guarantees that the propagated covariance matrices are always positive-definite

and increases the conditioning of these matrices. States in geophysical systems are not only non-linear dependent on each other but also often very high-dimensional.

Hence, the linearization in the extended Kalman filter has high computational costs (Evensen, 1992, 1993). The ensemble Kalman filter (EnKF, Evensen, 1994) approximates the linearization by a Monte-Carlo procedure with a limited number of samples; all samples are propagated with the non-linear forecast model. Based on the approximated linearizations, the Kalman filter equations are applied to update the mean and covariance. By updating single ensemble members with the Kalman filter equations, the ensemble members would be use two times, once for the estimation of the so-called Kalman gain and once for their update. This would lead to too small analysis covariances within the ensemble (Burgers et al., 1998; Houtekamer and Mitchell, 1998). Therefore, the stochastic EnKF (Burgers et al., 1998) independently draws observations from the observational probability function for every ensemble member. Nevertheless, randomly drawn observations causes additional sampling errors (Whitaker and Hamill, 2002). It can be therefore advantageous to estimate the ensemble Kalman filter based on a deterministic schema. Notably other deterministic schemata are the ensemble adjustment Kalman filter (Anderson, 2001) and the ensemble square-root Kalman filter with serial processing of observations (Whitaker and Hamill, 2002), whereas the LETKF is based on the ensemble transform Kalman filter (ETKF, (Bishop et al., 2001)). All of these deterministic methods use a different ensemble transformation to estimate the ensemble members. Hence, they can be unified under one common roof, namely, ensemble square-root filters (Tippett et al., 2003).

The ensemble transformation in the ETKF is not uniquely defined. Thus, it was shown that a symmetric square-root results into better results for high-dimensional systems with a low number of ensemble members than transformations with a positive-negative paired centering (Wang et al., 2004). Furthermore, it was shown that the data assimilation problem can be efficiently split up into smaller and local sub-problems (Ott et al., 2002, 2003). This allows a parallel evaluation of the analysis for every grid point in a forecast model independently. These developments finally evolved into the here-used LETKF.

3.1 Data assimilation from Bayesian principles

In data assimilation, I want to estimate the model-state $\mathbf{x}_t$ at time t based on observations $\mathbf{y}^o_{1:t}$ from time 1 to time t. The initial state $\mathbf{x}_0$ is unknown and has to be specified in a first step. I specify this initial state as Gaussian distributed $\mathbf{x}_0 \sim \mathcal{N}(\overline{\mathbf{x}}_0, \mathbf{P}_0)$ with $\overline{\mathbf{x}}_0$ as initial mean and $\mathbf{P}_0$ as initial covariance. The perfectly-known model $M_{t-1\mapsto t}(\mathbf{x}_{t-1})$ maps then the state from time $t-1$ to time t,

$$\mathbf{x}_t = M_{t-1\mapsto t}(\mathbf{x}_{t-1}). \tag{3.1}$$

I additionally define an observation operator $H_t(\mathbf{x}_t)$ that translates a model-state at time t to an observation at the same time. I assume an additive error $\boldsymbol{\epsilon}_t$, drawn from a Gaussian probability distribution with $\mathbf{0}$ as mean and $\mathbf{R}$ as time-

independent covariance,

$$\mathbf{y}_t^o = H_t(\mathbf{x}_t) + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \tag{3.2}$$

Based on the propagated model-state (3.1) and observation operator (3.2), I define a transition probability density $p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ and an observational likelihood $p(\mathbf{y}_t^o \mid \mathbf{x}_t)$, $\delta(\cdot)$ is hereby a Dirac delta function,

$$p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \delta(\mathbf{x}_t - M_{t-1\mapsto t}(\mathbf{x}_{t-1})), \tag{3.3}$$
$$p(\mathbf{y}_t^o \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t^o - H_t(\mathbf{x}_t), \mathbf{R}). \tag{3.4}$$

With these and the defined initial state $p(\mathbf{x}_0)$, the filtering solution of state $\mathbf{x}_t$, called posterior or analysis, results then from Bayes' rule in a two-step procedure:

1. I **propagate** the distribution of the previous state estimate $\mathbf{x}_{t-1}^a \sim p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}^o)$ at time $t-1$ to the current step t,

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o) = \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}^o) d\mathbf{x}_{t-1}.$$

   In the following I will use $\mathbf{x}_t^b = M_{t-1\mapsto t}(\mathbf{x}_{t-1}^a)$ as short form for the propagated model-state or call it prior forecast.
   The uncertainties of the prior distribution $p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o)$ are only a product of the propagated uncertainties from the previous state estimate. Here, I assume that these propagated uncertainties are additive and independent of any observational error. In addition, the uncertainties are presumably represented by a Gaussian distribution with $\mathbf{0}$ as mean and $\mathbf{P}_t^b$ as time-dependent covariance, leading to

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o) = \mathcal{N}(\mathbf{x}_t - \mathbf{x}_t^b, \mathbf{P}_t^b). \tag{3.5}$$

2. I **update** the current state $\mathbf{x}_t$ based on the observation $\mathbf{y}_t^o$ and the prior distribution $p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o)$,

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t}^o) = \frac{p(\mathbf{y}_t^o \mid \mathbf{x}_t) p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o)}{\int p(\mathbf{y}_t^o \mid \mathbf{x}_t') p(\mathbf{x}_t' \mid \mathbf{y}_{1:t-1}^o) d\mathbf{x}_t'}. \tag{3.6}$$

To estimate the best possible current state $\mathbf{x}_t^a$, I employ a maximum-a-posterior (MAP) procedure with the update step (3.6),

$$\begin{aligned} \mathbf{x}_t^a &= \underset{\mathbf{x}_t}{\operatorname{argmax}}\, p(\mathbf{x}_t \mid \mathbf{y}_{1:t}^o), \\ &\propto \underset{\mathbf{x}_t}{\operatorname{argmin}} -\log(p(\mathbf{y}_t^o \mid \mathbf{x}_t)) - \log(p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o)). \end{aligned} \tag{3.7}$$

Together with the Gaussian distribution (3.3) and the Gaussian likelihood (3.4), MAP results in the minimization of the following cost function $\mathcal{L}(\mathbf{x}_t)$, with $[\cdot]^\mathsf{T}$ as transposed and $[\cdot]^{-1}$ as inverse,

$$\begin{aligned}\mathcal{L}(\mathbf{x}_t) =&[\mathbf{x}_t - \mathbf{x}_t^b]^\mathsf{T}(\mathbf{P}_t^b)^{-1}[\mathbf{x}_t - \mathbf{x}_t^b] \\ &+[\mathbf{y}_t^o - \mathrm{H}_t(\mathbf{x}_t)]^\mathsf{T}\mathbf{R}^{-1}[\mathbf{y}_t^o - \mathrm{H}_t(\mathbf{x}_t)].\end{aligned} \tag{3.8}$$

The first term of this cost function constrains the solution to be as close as possible to the propagated state, and hence to the previous observations, whereas the second term nudges the translated solution in observational space to the current observations. This cost function is the basis for variational procedures like 3D-Var (Lorenc, 1981; Parrish and Derber, 1992; Courtier et al., 1998), where I would explicitly minimize (3.8) by gradient descent methods. For this minimization, I would need the tangent linear model of the observations operator $\mathrm{H}_t(\mathbf{x}_t)$, but this tangent linear model is not provided in my TerrSysMP-based data assimilation environment.

I can linearize the observation operator $\mathbf{H}$ around the current state estimate $\mathbf{x}_t$ to find an analytical solution for (3.8) and avoid the need of a tangent linear model (Hunt et al., 2007). The optimal solution of (3.6) and (3.8) is given as Gaussian distribution, because the Gaussian prior distribution $p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o)$ is the conjugate prior for another Gaussian distribution, here the observational likelihood $p(\mathbf{y}_t^o \mid \mathbf{x}_t)$. The resulting Gaussian posterior distribution $\mathcal{N}(\mathbf{x}_t^a, \mathbf{P}_t^a)$ has then $\mathbf{x}_t^a$ as mean and $\mathbf{P}_t^a$ as covariance (Kalman, 1960),

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}(\mathbf{y}_t^o - \mathbf{H}_t\mathbf{x}_t^b), \tag{3.9}$$

$$\begin{aligned}\mathbf{P}_t^a &= \mathbf{P}_t^b - \mathbf{K}\mathbf{H}_t\mathbf{P}_t^b \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H}_t)\mathbf{P}_t^b \\ &= [(\mathbf{P}_t^b)^{-1} + (\mathbf{H}_t)^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}_t]^{-1},\end{aligned} \tag{3.10}$$

$$\begin{aligned}\mathbf{K} &= \mathbf{P}_t^b(\mathbf{H}_t)^\mathsf{T}[\mathbf{H}_t\mathbf{P}_t^b(\mathbf{H}_t)^\mathsf{T} + \mathbf{R}]^{-1} \\ &= [(\mathbf{P}_t^b)^{-1} + (\mathbf{H}_t)^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}_t]^{-1}(\mathbf{H}_t)^\mathsf{T}\mathbf{R}^{-1}.\end{aligned} \tag{3.11}$$

This set of Kalman filter equations completely defines the posterior distribution with its mean (3.9) and covariance (3.10). The Kalman gain $\mathbf{K}$ (3.11) has a special role. On the one hand, it translates the innovation $(\mathbf{y}_t^o - \mathbf{H}_t\mathbf{x}_t^b)$ into a correction term for the estimation of the mean state $\mathbf{x}_t^a$. On the other hand, the gain determines the reduction of the prior covariance $\mathbf{P}_t^b$ because of new information given by the observations.

All data assimilation methods in this dissertation use a form of this set of equations (3.9)-(3.11) and deviate only in the determination of the linearized observation operator $\mathbf{H}_t$, prior state $\mathbf{x}_t^b$, and prior covariance $\mathbf{P}_t^b$. In the following, I will derive the ensemble transform Kalman filter from these general principles based on an ensemble approximation to (3.3) and (3.5).

3.2 Gaussian ensemble data assimilation

In (3.1) and (3.3), I propagate a single solution from time $t-1$ to time t and specify a temporal-dependent covariance $\mathbf{P}_t^b$ to the propagated state $\mathbf{x}_t^b$. A common simplification is to assume a static covariance matrix $\mathbf{P}_t^b \approx \mathbf{B}$, which can be used in 3D-Var or simplified extended Kalman filters. This simplification of a static $\mathbf{B}$-matrix can lead to a significant over- or underestimation of the Kalman gain in (3.11) because of the chaotic nature of the problem in the case of a non-linear dynamical model.

I approximate $p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o)$ in (3.5) by a Monte-Carlo approach to circumvent these problems and create a dynamical covariance, which depends on the current information flow in the model. For the Monte-Carlo approximation, I draw k samples, in the following called ensemble members, from the previously updated probability density $\mathbf{x}_{t-1}^{a(i)} \sim p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}^o)$, here shown for the i-th ensemble member. To get a sample from the prior state, I propagate every ensemble member independently with the dynamical model $\mathbf{x}_t^{b(i)} = M(\mathbf{x}_{t-1}^{a(i)})$. As a consequence, I approximate the prior distribution $p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o)$ from (3.5) and needed in (3.6) with these samples.

I approximate the prior state $\mathbf{x}_t^b$ and its covariance $\mathbf{P}_t^b$ based on the propagated ensemble members with $\overline{\mathbf{x}}_t^b$ as ensemble mean and $\delta\mathbf{x}_t^{b(i)}$ as i-th perturbation,

$$\mathbf{x}_t^b \approx \overline{\mathbf{x}}_t^b = k^{-1} \sum_{i=1}^{k} \mathbf{x}_t^{b(i)}, \tag{3.12}$$

$$\begin{aligned} \mathbf{P}_t^b &\approx (k-1)^{-1} \sum_{i=1}^{k} (\mathbf{x}_t^{b(i)} - \overline{\mathbf{x}}_t^b)(\mathbf{x}_t^{b(i)} - \overline{\mathbf{x}}_t^b)^T \\ &= (k-1)^{-1} \sum_{i=1}^{k} \delta\mathbf{x}_t^{b(i)} (\delta\mathbf{x}_t^{b(i)})^T. \end{aligned} \tag{3.13}$$

With these first two moments of the empirical distribution, I approximate (3.5) and use $\overline{\mathbf{x}}_t^b$ and $\mathbf{P}_t^b$ to update the mean and the covariance with (3.9) and (3.10). To complete the update step, I need a linearized version of the observation operator $\mathbf{H}$. In the next section, I show how to derive this linearized version in form of the ensemble transform Kalman filter.

3.3 Ensemble transform Kalman filter

The ensemble transform Kalman filter (ETKF, Bishop et al., 2001; Hunt et al., 2007) is an efficient form of an ensemble Kalman filter for large-dimensional geophysical systems. The ETKF estimates the analysis in a transformed space spanned by the perturbations of the prior ensemble members. This reduces the computational costs compared to a naively implemented ensemble-version of the Kalman filter equations. In addition, the ETKF is a square-root ensemble Kalman filter (Tippett

et al., 2003) and deterministically estimates the analysis ensemble based on the ensemble statistics and the Kalman equations.

With the ensemble approximations (3.12) and (3.13), the assimilation increment of the analyzed model-state $\Delta \boldsymbol{x}_t^a = \boldsymbol{x}_t^a - \overline{\boldsymbol{x}}_t^b$ lies in the space spanned by $\delta \boldsymbol{X}_t^b$ (Lorenc, 2003; Hunt et al., 2007) as column-wise matrix of the prior ensemble perturbations with $\delta \boldsymbol{x}_t^{b(i)}$ as i-th column. This motivates a linear variable transformation where the solution $\boldsymbol{x}_t$ is explicitly represented by a vector of weights $\boldsymbol{w}$ with $\mathbf{I}$ as $k \times k$-dimensional identity matrix,

$$\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, (k-1)^{-1}\mathbf{I}), \tag{3.14}$$

$$\boldsymbol{x}_t = \overline{\boldsymbol{x}}_t^b + \delta \boldsymbol{X}_t^b \boldsymbol{w}. \tag{3.15}$$

The specified prior distribution (3.14) ensures that $\boldsymbol{x}_t$ equals to the prior ensemble statistics if we assimilate no observations.

This variable transformation (3.15) affects the variational cost function from (3.8),

$$\begin{aligned}\mathcal{L}(\boldsymbol{w}) =& (k-1)(\boldsymbol{w})^T \boldsymbol{w} \\ &+ [\boldsymbol{y}_t^o - H_t(\overline{\boldsymbol{x}}_t^b + \delta \boldsymbol{X}_t^b \boldsymbol{w})]^T \mathbf{R}^{-1} [\boldsymbol{y}_t^o - H_t(\overline{\boldsymbol{x}}_t^b + \delta \boldsymbol{X}_t^b \boldsymbol{w})].\end{aligned} \tag{3.16}$$

Thus, the variable transformation changes the problem from searching the optimal solution in model space to an optimization in weight space.

To derive an analytically tractable form of the posterior in weight space, I have to linearize the observation operator around the prior ensemble mean. For this linearization, I transform every ensemble member independently from model space into observational space $\boldsymbol{y}_t^{b(i)} = H(\boldsymbol{x}_t^{b(i)})$. Based on the transformed ensemble members, I approximate the prior ensemble mean in observational space $H_t(\overline{\boldsymbol{x}}_t^b) \approx \overline{\boldsymbol{y}}_t^b = k^{-1} \sum_{i=1}^{k} H(\boldsymbol{x}_t^{b(i)})$. With this ensemble mean in observational space, I define a column-wise matrix $\delta \boldsymbol{Y}_t^b$ of ensemble perturbations in observational space with $\delta \boldsymbol{y}_t^{b(i)} = H(\boldsymbol{x}_t^{b(i)}) - \overline{\boldsymbol{y}}_t^b$ as i-th column. This matrix acts as approximated tangent linear, translating from weight space to observational space. As linearization, I obtain the following equation,

$$H_t(\overline{\boldsymbol{x}}_t^b + \delta \boldsymbol{X}_t^b \boldsymbol{w}) \approx \overline{\boldsymbol{y}}_t^b + \delta \boldsymbol{Y}_t^b \boldsymbol{w}. \tag{3.17}$$

With this linearized observation operator, I simplify the cost function in weight space (3.16),

$$\begin{aligned}\mathcal{L}(\boldsymbol{w}) =& (k-1)(\boldsymbol{w})^T \boldsymbol{w} \\ &+ [\boldsymbol{y}_t^o - \overline{\boldsymbol{y}}_t^b - \delta \boldsymbol{Y}_t^b \boldsymbol{w}]^T \mathbf{R}^{-1} [\boldsymbol{y}_t^o - \overline{\boldsymbol{y}}_t^b - \delta \boldsymbol{Y}_t^b \boldsymbol{w}].\end{aligned} \tag{3.18}$$

For this cost function, I do not need access to the observation operator any longer if I have access to the transformed ensemble in observational space. This is important

for my implementation, because I can simply use the model output from COSMO without needing an explicit observation operator for the 2-metre-temperature.

Based on this simplified cost function, I derive in analogy to (3.9)-(3.11) a set of equations to analytically estimate the posterior in weight space with $\boldsymbol{w}^a$ as mean and $\widetilde{\mathbf{P}}^a$ as covariance,

$$\boldsymbol{w}^a = [(k-1)\mathbf{I} + (\delta\mathbf{Y}_t^b)^\top \mathbf{R}^{-1} \delta\mathbf{Y}_t^b]^{-1} (\delta\mathbf{Y}_t^b)^\top \mathbf{R}^{-1} (\mathbf{y}_t^o - \overline{\mathbf{y}}_t^b), \tag{3.19}$$

$$\widetilde{\mathbf{P}}^a = [(k-1)\mathbf{I} + (\delta\mathbf{Y}_t^b)^\top \mathbf{R}^{-1} \delta\mathbf{Y}_t^b]^{-1}. \tag{3.20}$$

With this posterior distribution and the variable transformation (3.15), I estimate the posterior in model space in the following way,

$$\boldsymbol{x}_t^a = \overline{\boldsymbol{x}}_t^b + \delta\mathbf{X}_t^b \boldsymbol{w}^a, \tag{3.21}$$

$$\mathbf{P}_t^a = \delta\mathbf{X}_t^b \widetilde{\mathbf{P}}^a (\delta\mathbf{X}_t^b)^\top. \tag{3.22}$$

To get prior ensemble members for the next update step, I have to reconstruct the posterior ensemble members at the current update step. I use a symmetric square-root of $\mathbf{P}_t^a$ to define the ensemble perturbations in weight space. With this symmetric square-root formulation, I guarantee that the perturbations are centered and have $\mathbf{P}_t^a$ from (3.22) as covariance (Wang et al., 2004). For the ensemble perturbations in weight space, this results in

$$\delta\boldsymbol{W}^a = [(k-1)\widetilde{\mathbf{P}}^a]^{\frac{1}{2}}. \tag{3.23}$$

This weight perturbation matrix $\delta\boldsymbol{W}^a$ contains column-wise the transformations needed to construct perturbations for single ensemble members. Furthermore, I add column-wise the mean weight for a more efficient construction of the ensemble,

$$\begin{aligned} \boldsymbol{x}_t^{a(i)} &= \overline{\boldsymbol{x}}_t^b + \delta\mathbf{X}_t^b \boldsymbol{w}^a + \delta\mathbf{X}_t^b \delta\boldsymbol{w}^{a(i)} \\ &= \overline{\boldsymbol{x}}_t^b + \delta\mathbf{X}_t^b (\boldsymbol{w}^a + \delta\boldsymbol{w}^{a(i)}). \end{aligned} \tag{3.24}$$

I can propagate these reconstructed posterior ensemble members to get the prior ensemble members for the next update step. The data assimilation cycle for the ETKF is therefore completely described by this set of equations (3.19)-(3.24).

3.4 Inflation and Localization

When the dimensionality of state variables is much higher than the number of ensemble members within the LETKF, the error can evolve into directions that cannot be represented with the ensemble members. As a consequence, we need additional techniques to stabilize ensemble data assimilation for high-dimensional geophysical systems like TerrSysMP, even in idealized twin experiments. In this section, I present elaborate more on why I need to inflate the prior ensemble covariances, and why I need to localize the assimilation impact for high-dimensional data assimilation. In addition, I specify which type of inflation and localization I use in Chapter 4 and Chapter 5.

3.4.1 Multiplicative prior inflation

As I explain in Chapter 2, I use in my idealized twin experiments the same model configuration for my data assimilation experiments as I use for my nature run. The data assimilation experiments are model-error free compared to the nature run. Nevertheless, the ensemble approximation from (3.12) and (3.13) causes sampling errors. In addition, the observation operator for the 2-metre-temperature is non-linearly dependent on the soil moisture. This non-linear observation operator can lead to a non-Gaussian posterior. If I then approximate the posterior distribution by a Gaussian distribution, I likely underestimate the uncertainties compared to the forecast error.

To counter-act this underestimation, I use prior multiplicative covariance inflation. In this inflation, I multiply the prior covariance by a factor $\rho > 1$. In my experiments, the inflation factor varies between 1.006 and 1.18. This artificially increases the prior ensemble spread such that the ensemble members cover a larger range of possibilities. I also increase with this technique the posterior ensemble spread, but I reduce the trust in the prior distribution. This reduced trust results in an increased observational impact during the update step.

The need of covariance inflation depends on the used data assimilation technique, the assimilated observations, and the updated state variables. For every experiment in Chapter 4 and Chapter 5, I therefore have to re-tune manually the inflation factor so that the ensemble spread matches the root-mean-squared error to my nature run.

3.4.2 Localization

The number of ensemble members is low compared to the state dimensions, especially for coupled data assimilation involving multiple Earth system components. This discrepancy in the dimensionality introduces spurious correlations into the prior covariances (Miyoshi et al., 2014). These spurious correlations are correlations within the ensemble which are not physically explainable and can degrade the analysis. To solve the problem of spurious correlations, localization is applied in ensemble-based data assimilation. In localization, the influence of grid points to each other is weighted based on the physical distance between two points. The localized ETKF (LETKF, Hunt et al., 2007) is a observationally localized variant of the ETKF. In the LETKF, I estimate the ensemble weights for every grid point independently. The observations for a specific grid point are then weighted by their physical distance to a considered grid point.

I localize here with Gaspari-Cohn covariance functions (Gaspari and Cohn, 1999). They are also used for operational data assimilation in the atmosphere (Schraff et al., 2016). These covariance functions resemble Gaussian covariance functions and are completely specified by their localization radius l. As a difference to Gaussian covariance functions, they decay faster and cut observations after $2 \times l$ the radius off.

For all experiments, I employ the same covariance functions in horizontal and vertical dimensions, which are always applied together (see also Table 3.1 for the used radii). I localize in the horizontal with a radius of 15 km, which is quite small in comparison to operationally used values in the atmosphere (between 50 km and 100 km, Schraff et al. (2016)).

Table 3.1: Chosen localization covariance functions and radii.

Dimension	Covariance function	Localization radius
Horizontal	Gaspari-Cohn	15 km
Vertical / atmosphere	Gaspari-Cohn	0.3 ln hPa
Vertical / land surface	Gaspari-Cohn	0.7 m

As simplification for the vertical localization, I assume that my 2-metre-temperature observations are valid at a height of 0 m. The localization radius in the vertical is different for the atmosphere and the land surface. In the atmosphere, I localize vertically in terms of logarithmic pressure and use a typical value of 0.3 ln hPa from operational settings (Schraff et al., 2016). Observations in the atmospheric boundary layer have their largest impact on soil moisture analysis at root-depth (Muñoz-Sabater et al., 2019). Afterwards, their physically explainable impact is negligible. Therefore, I chose my vertical localization radius in soil (0.7 m) so that the innovations for soil levels below the root-depth (fifth soil layer, 0.21 m) are dampened.

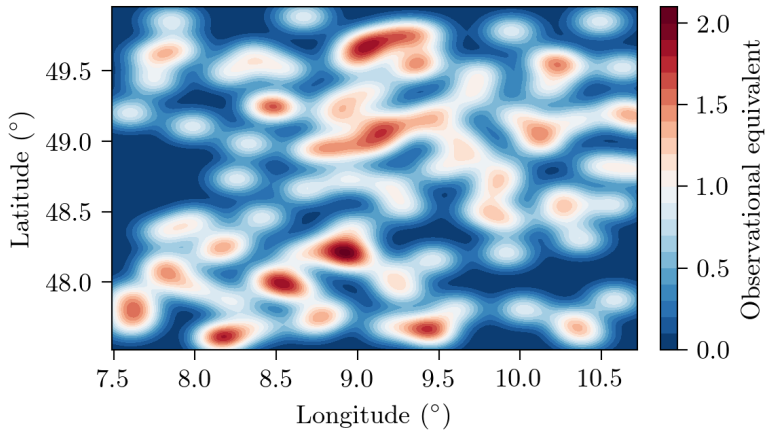


Figure 3.1: Number of potential observational equivalents per grid point estimated based on the chosen horizontal and vertical localization radius in soil (see Table 4.1), representative in 0.21 m depth. Blue colors characterize fewer equivalents per grid point compared to a grid-point-based assimilation (e.g. SEKF), whereas red colors indicate more equivalents.

Based on the chosen localization radii and the position of the 99 measurement sites, I can estimate a potential observational equivalent. The potential observational equivalent would be the number of observations that are used for a specific grid point within the data assimilation if we neglect the influence of the ensemble covariance. I define this potential observational equivalent as the sum of all observational weights. For my ensemble Kalman filter, the number of potential observational equivalents is shown in Fig. 3.1 for the soil moisture in root-depth;

the mean observational equivalent is 0.566. In vertical data assimilation methods, like the simplified extended Kalman filter, observations are assimilated on a grid-point-based level. They assimilate one observation per observational type and grid point, and they would have a potential observational equivalent of 1. As a consequence, I have half the number of potential observations per grid-point relative to a fully-observed field. I therefore have only a limited observability within my ensemble Kalman filter experiments, and in some areas, I would expect no assimilation impact at all.

3.5 Implementation for the atmosphere-land interface

In the previous sections, I discussed the theory behind the localized ensemble transform Kalman filter (LETKF), I outline how I implemented the LETKF for TerrSysMP. In a first step, I explain my observation operator for the 2-metre-temperature. In a second step, I describe the post-processing applied to the posterior gained from the LETKF. I give more technical details about my implementation of the cycled data assimilation in the end.

3.5.1 Observation operator for the 2-metre-temperature

The only observations that I assimilate are 2-metre-temperature observations. I synthesize these observations out of my nature run in twin experiments (see also Section 2.4.2 for more information). For both, the observations and their ensemble equivalent, I define the same 2-metre-temperature observation operator.

The observation operator is based on the diagnostic 2-metre-temperature output from COSMO. The diagnostic 2-metre-temperature is estimated as a weighted average of the ground temperature and temperature at the lowest model level, here in 10 meters height. The weighting results out of the similarity theory (Monin and Obukhov, 1954) and depends on the laminar transfer factor for scalars, which itself is influenced by the surface transfer coefficient for heat. This surface transfer coefficient for heat is a coupled quantity in TerrSysMP and estimated in CLM. Therefore, the soil moisture indirectly modulates the 2-metre-temperature via the surface transfer coefficient for heat and the surface temperature.

I define my own observation operator for the 2-metre-temperature. I simply select the 2-metre-temperature at the nearest grid point to an existing measurement site of the DWD. As measurement height, I use the model orographic height at this selected nearest grid point.

3.5.2 Post-processing of the soil moisture posterior

The LETKF corrects a prior distribution to the posterior distribution with a linear assumption. Thus, the resulting posterior ensemble can contain physically-inconsistent states. This could be especially a problem for the soil moisture, where the quantity is bounded by $0\,\mathrm{m}^3\,\mathrm{m}^{-3}$ and the saturation point. As one

post-processing step, I constrain the soil moisture content to $0\,\mathrm{m}^3\,\mathrm{m}^{-3}$ as a lower bound to avoid negative values.

Data assimilation can add or subtract water to the soil moisture, so the vertically integrated water content is possibly not preserved in the update step. In contrast, the vertically integrated water content in the prior is based on the model propagation with CLM and is in a physical equilibrium. In the update step, I want to retain the same vertically-integrated equilibrium without the need for additional optimization steps. To retain the equilibrium, I estimate the residual of the vertical water content from the posterior to the prior. This residual is afterwards added or subtracted from the unconfined aquifer to balance the data assimilation increment in the upper soil moisture layers. Therefore, I rebalance the vertically integrated water content in the posterior at the cost of water in the unconfined aquifer.

3.5.3 Ensemble-based data assimilation with Torch-Assimilate

The communication between the data assimilation and TerrSysMP is file-based. Hence, TerrSysMP writes its output- and analysis files, which are then modified by the data assimilation. The data assimilation is developed as Python packages (Finn, 2020b,a) and is essentially based on Xarray (Hoyer and Hamman, 2017), Dask (Dask Development Team, 2016; Rocklin, 2015), PyTorch (Paszke et al., 2019), and Prefect (PrefectHQ, 2021).

I have developed these packages under an object-oriented and modular approach in the last two years. The idea behind these packages was to create a generalizable and efficient data assimilation environment. This environment should be in the best case independent from any model- and data assimilation method implementation (the environment is then model- and data assimilation method agnostic). To gather an approach that is independent from any model implementation, the package is currently restricted to ensemble data assimilation methods and based on the variable transformation into ensemble weights as specified in (3.15).

Additionally, I have split the data assimilation into two packages, torch-assimilate (Finn, 2020b) and PyBacy (Finn, 2020a). I designed torch-assimilate to be model-agnostic with a common Xarray-based interface, whereas PyBacy couples torch-assimilate to the models and is therefore method-agnostic. This design approach results in three different layers:

1. The **Core** layer where the data assimilation equations are efficiently implemented in PyTorch code. This layer allows a simple and lightweight implementation of new data assimilation algorithms and concurrently includes core modules for the ensemble transform Kalman filter (see also Section 3.3 for the equations), the iterative ensemble Kalman smoother (IEnKS, Bocquet and Sakov (2014), see also Section A.3), and a novel kernelized ensemble transform Kalman filter (see also Section 6.1 for a derivation) with its kernels. Because these core modules are all implemented in PyTorch, they natively support differentiation. In addition, they could be compiled with the just-in-

time compiler in PyTorch for an additional speed-up. This core layer works on the level of ensemble weights. It consumes tensors in observational space and returns ensemble weights.

2. The **Interface** layer is a convenient higher-level interface to the core modules. This layer is build upon the capacities of Xarray and Dask. It implements localization, observation operators, and pre- and post-processing methods like covariance inflation. I use Dask for an efficient parallelization and taking advantage of high-performance computing. This allows me to scale my methods to high-dimensional coupled models like TerrSysMP. In general, this interface estimates the quantities in Xarray and bridges them to PyTorch tensors, needed in the core modules. The ensemble weights are then collected from the core modules and applied to the given ensemble states. This interface layer therefore works on the level of observational and ensemble states.

3. The **Pipeline** is the outer-most layer and specifies the communication of the data assimilation to the models. By design, this layer is data assimilation method-agnostic, but depends on the utilized models. The pipeline system is based on the principles of directed acyclic graphs (DAG) and is implemented in Prefect. The building blocks for the DAGs are tasks which specify a single operation, like the read-in of output from CLM. These building blocks allow a dynamical composition of graphs based on Python-scripts. The tasks are then pooled together into flows, which specify a composition with a unique purpose, like implementing the update step or propagation step for data assimilation. An outer engine is used to cycle these flows with configurations, specified by YAML-files. This pipeline layer acts as outer-cover and calls the interface layer and model binaries.

The model-specific parts of the data assimilation are contained in the pipeline system, whereas the data assimilation equations are implemented in the core layer. The interface layer acts as transmitter and translates quantities from the model specific parts into quantities needed for the core methods. A generic information flow on the basis of the previously derived (localized) ETKF equations is shown in Appendix A.2. All my experiments in this thesis are based on this data assimilation environment.

3.6 Offline data assimilation experiments

In online data assimilation experiments (see also Fig. 3.2 (a) for a schematic overview), I cycle through the propagation step (3.5) and the update step (3.6) in Section 3.1. The resulting posterior at time t will be then again propagated to time $t+1$ and acts as prior for this next time step. Thus, the posterior of this next time step $t+1$ depends not only on the assimilation at $t+1$, but also on the assimilation at time t. Because of this recursion, I get an accumulated data assimilation impact. Hence, the posteriors in the experiments are also impacted by effects of the propagation. This makes it difficult to compare state trajectories

of online data assimilation experiments. Furthermore, online data assimilation experiments have high computational costs, because the propagation step is the most expensive part of a data assimilation cycle. To circumvent these problems, I utilize, beside online experiments, so-called offline data assimilation experiments.

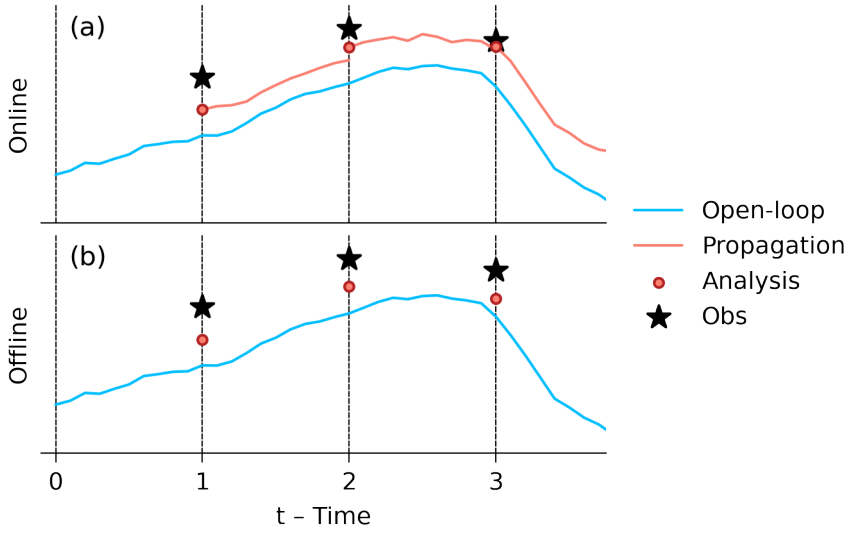


Figure 3.2: A schematic figure showing the differences between (a) online and (b) offline data assimilation experiments. In online data assimilation experiments, the posterior is propagated again to the next time step and acts after the propagation as prior for the next analysis. In offline data assimilation experiments, the trajectory of the open-loop run without data assimilation is the prior for every time step. The posterior of offline experiments is only used for diagnostic purposes without additional propagation runs from the model.

In offline data assimilation experiments (see also Fig. 3.2 (b) for a schematic overview), I create analyses with the update step without an additional propagation. For this procedure, I reuse an existing trajectory, often an open-loop run without any data assimilation. The trajectory at a given time t acts as prior for the update step (3.6), and I assimilate observations into this prior to create a posterior. This posterior is not propagated to gain the prior at the next time step $t+1$, but instead I use again the existing trajectory as prior. Therefore, I create in offline data assimilation experiments a set of independent analyses at different time steps without a recursion.

In these experiments, the analyses are based on the ensemble statistics of the existing trajectory. The statistics in the existing trajectory determine the observational impact in the data assimilation. If I use an open-loop run without any data assimilation, I do not get the propagated effect on the ensemble covariances. Since ensemble Kalman filters can only reduce the covariances in an update step, the ensemble spread in the prior of an offline experiment is likely larger than in the prior of an online experiment. This would mean that I overestimate the observational impact compared to online experiments.

Based on the same prior trajectory, I can generate different analyses with different data assimilation methods. The differences in these analyses are only influenced by

the update step of the data assimilation methods. This simplifies the comparison between data assimilation methods, because I do not need to disentangle the effects of the propagation step from the effects of the update step. In addition, I skip the propagation step in the data assimilation cycle and only the update step remains. This makes offline experiments much cheaper than online experiments.

4

Cross-compartmental ensemble data assimilation for the atmosphere-land interface

In this Chapter, I revise the potential of assimilating atmospheric boundary layer observations into the soil moisture. The evapotranspiration and sensible heat flux couple the atmospheric boundary layer to the land surface. Nevertheless, previous studies often stated a negative assimilation impact of boundary layer observations on the soil moisture analysis. A newly potential of boundary layer observations for land surface data assimilation emerges from recent developments in physically-consistent hydrological model systems, like TerrSysMP (see also Section 2.2), together with ensemble-based data assimilation.

Here, I perform idealized twin experiments for a seven-day period with TerrSysMP, as they are described in Chapter 2. Based on my nature run, I synthesize sparse 2-metre-temperature observations (see also Section 2.4.2). These observations are subsequently assimilated into different experiments. I conduct these experiments with the same model configuration as I use for my nature run (see also Section 2.2). In these experiments, I compare a simplified extended Kalman filter (SEKF) to a localized ensemble transform Kalman filter (LETKF). Since I use the LETKF throughout this thesis, I describe its implementation and components in Chapter 3. As I explain in Section 4.2.1, the SEKF can be derived from the Kalman filter equations (3.9)-(3.11). I describe the comparison experiments in Section 4.3. As a result in Section 4.4, I show that directly assimilating 2-metre-temperature observations hourly across the atmosphere-land interface with a localized ensemble Kalman filter has a positive assimilation impact on the soil moisture analysis. This proves that observations from the atmospheric boundary layer can be assimilated across interfaces in the Earth system with an ensemble Kalman filter in a physical-consistent way. This cross-compartmental data assimilation not only decreases the forecast error of the observed compartment, but also improves the analysis of the other compartment that is updated. On the basis of the results in this Chapter, I therefore propose as my first framework to unify and couple the data assimilation in Earth system models with a localized ensemble Kalman filter.

In another form, this chapter is partially submitted and currently in review as: Finn, T. S., Geppert, G., and Ament, F.: "Ensemble-based data assimilation of atmospheric boundary layer observations improves the soil moisture analysis", *Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2020-672, 2021.* As this chapter is intended for publication with multiple authors, I switch in its content to to the first person plural ("we") form.

4.1 Introduction

Assimilation of atmospheric boundary layer observations into land surface models primarily improves the coupled forecast of the atmospheric boundary layer. The sensible heat flux and evapotranspiration couple the land surface to the boundary layer, and we expect that using boundary layer observations in land surface data assimilation has an additional positive impact on the soil moisture analysis. In contrast to this expectation, previous studies often stated a negative impact on the soil moisture analysis (Hess, 2001; Drusch and Viterbo, 2007; Muñoz-Sabater et al., 2019; Draper et al., 2011; Su et al., 2013; Carrera et al., 2019). Recent developments in physically-consistent hydrological models (Fatichi et al., 2016; Prein et al., 2015; Vereecken et al., 2016) and strongly-coupled ensemble-based data assimilation (Sluka et al., 2016; Penny and Hamill, 2017) allow us to challenge this negative assimilation impact. Through the lens of these developments, we specifically concentrate here on the relationship between the atmospheric 2-metre-temperature and soil moisture. By focusing on this relationship only, we show in this study that we can extract information about the soil moisture from boundary layer observations.

Ensemble-based data assimilation methods, like Ensemble Kalman Filters (EnKF), are used in data assimilation for the atmosphere. By using a three-dimensional EnKF, we take horizontal and vertical covariances into account, and observations at their measurement sites can be assimilated without an additional interpolation step. In land-surface-only data assimilation with in-situ soil moisture observations, the additional use of horizontal covariances decreases the soil moisture analysis error compared to one-dimensional methods (Fairbairn et al., 2015; Reichle et al., 2002), resulting in promising applications on reanalysis problems (Draper and Reichle, 2019). As being computationally more demanding than simplified approaches (Reichle and Koster, 2003), EnKFs are nevertheless rarely used for operational data assimilation in land surface models (Carrera et al., 2015; Milbrandt et al., 2016). One-dimensional Simplified Extended Kalman Filters (SEKF) are thus implemented for land surface data assimilation (Hess, 2001; Rosnay et al., 2013; Mahfouf et al., 2009; Dharssi et al., 2011; Bélair et al., 2003; Giard and Bazile, 2000). Moreover, the soil moisture analysis is often estimated in its own daily assimilation cycle in addition to assimilation cycles for the atmosphere on shorter, hourly-like, time-scales. To combine these assimilation cycles into one single cycle, EnKFs are one candidate because of their ensemble-based flow-dependency. We use here a combined three-dimensional EnKF setup, where we assimilate the 2-metre-temperature at 99 measurement sites. Based on these limited observations, we jointly update the soil moisture and atmospheric temperature, and compare this setup to the SEKF. We additionally test with this EnKF setup the hypothesis of hourly updating the soil moisture based on a flow-dependent coupling between land surface and atmosphere.

Land surface models are often less advanced compared to currently used numerical weather prediction models for the atmosphere. Furthermore, the horizontal resolution is often not fine enough to model soil processes appropriately, leading

to biases and model errors within land surface schemes (Dirmeyer et al., 2017; Kauffeldt et al., 2015; Orth et al., 2017; Best et al., 2015). The Terrestrial Systems Modelling Platform (TerrSysMP, Shrestha et al. 2014; Gasper et al. 2014) is a platform focussed on modelling soil and hydrological processes and can thus scale from continental scales (Kollet et al., 2018) up to metre-scale resolution in soil (Gebler et al., 2017). For our experiments, we utilize TerrSysMP to model the coupling between atmosphere and land surface with an advanced hydrology platform. Together with TerrSysMP, we perform idealized twin experiments, using the same system configuration for our nature run and our data assimilation experiments. In addition, we only perturb the initial soil conditions to create an ensemble of forecasts. With this distilled setup, we are able to isolate the effect of perturbations within the soil moisture on the 2-metre-temperature without having model errors.

Strongly-coupled data assimilation reduces inconsistencies across different interfaces (Sawada et al., 2018; Lin and Pu, 2018, 2019) and is thus a natural approach to initialize fully-coupled earth system models, like TerrSysMP. Here, the same observations are assimilated across all compartments within a unified data assimilation environment. To unify the environment, we would need to integrate land surface data assimilation into the assimilation cycle for the atmosphere, with updating frequencies up to an hour. However, the soil moisture analysis is operationally decoupled from the analysis for the atmosphere, and land surface data assimilation relies on weakly-coupled data assimilation, where only the forecast models are coupled. We reflect this weakly-coupled approach in a SEKF experiment, assimilating observations at 12:00 UTC into soil moisture at 00:00 UTC, the night before. We compare this SEKF experiment to a weakly-coupled localized EnKF experiment, where we hourly update the soil moisture based on instantaneous 2-metre-temperature observations. By additionally updating the atmospheric temperature with the same observations, we test one exemplary prototype of a strongly-coupled EnKF against the other, weakly-coupled, approaches. The results from the strongly-coupled EnKF experiment are then further analyzed with regard to the driving factors for the impact of boundary layer observations on the coupled data assimilation.

4.2 Data assimilation environment

The propagation step in our data assimilation environment is based on the Terrestrial System Modeling Platform (TerrSysMP). In TerrSysMP, the atmospheric model COSMO is coupled to the land surface model CLM by the OASIS3 coupler. For more information about TerrSysMP and the model configuration, we refer the reader to Section 2.2. For a description about the theory behind our offline data assimilation experiments that are additionally used in this study, we refer to Section 3.6.

In this study, we use two different types of data assimilation. On the one hand, we use a simplified extended Kalman filter (SEKF) as reference. An implementation of the SEKF is also operationally used at the ECMWF for land surface data assimilation ("IFS Documentation CY47R1 - Part II: Data Assimilation" 2020). On the

other hand, we test a localized ensemble transform Kalman filter (LETKF, Bishop et al. (2001) and Hunt et al. (2007)). This type of ensemble Kalman filter is used throughout this whole thesis. Therefore, we refer for a technical derivation of the LETKF to Section 3.3, for the used localization functions and the multiplicative inflation to Section 3.4, and for the implementation to Section 3.5. In the following, we describe only the technical details for the simplified extended Kalman filter.

4.2.1 Simplified extended Kalman filter

The simplified extended Kalman filter (SEKF) is a simplified form of an extended Kalman filter. Its simplifications are based on a deterministic prior state, a fixed prior covariance matrix, and a finite-differences' approximation to the tangent linear model of the observation operator.

Our SEKF updates only the soil moisture as deterministic prior state $\mathbf{x}_t^b$, To update the prior state $\mathbf{x}_t^b$ and to get a posterior state $\mathbf{x}_t^a$, we apply the update equation for the state mean (3.9) without estimating the posterior state covariance,

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{P}_t^b(\mathbf{H}_t)^T[\mathbf{H}_t\mathbf{P}_t^b(\mathbf{H}_t)^T + \mathbf{R}]^{-1}(\mathbf{y}_t^o - H(\mathbf{x}_t^b)). \tag{3.9}$$

To ensure consistency in the water balance within a single column, we post-process the updated soil moisture states by leveraging the updated increments at the cost of water in the non-updated unconfined aquifer. The updated soil moisture together with the non-updated other states is propagated with the full TerrSysMP model system to the next update time at 00:00 UTC, the following day. We assume a diagonal and static in time prior covariance matrix $\mathbf{P}_t^b \approx \mathbf{B} = (\sigma^b)^2\mathbf{I}$ with $\sigma^b = 0.01\,m^3m^{-3}$ as constant standard deviation, as operationally used at the ECMWF ("IFS Documentation CY47R1 - Part II: Data Assimilation" 2020).

The soil moisture evolves slowly compared to the atmosphere in dry conditions, as shown in Section 2.4.1. In addition, the incoming solar radiation at the surface has their maximum during noon. As a consequence, also the coupling between land surface and atmospheric boundary layer has its maximum at the same time. Hence, a common strategy in the SEKF is to update the soil moisture once a day at midnight based on boundary layer observations at daytime (Hess, 2001; Balsamo et al., 2004). As simplification, we update our prior state at 00:00 UTC based on 2-metre-temperature observations at 12:00 UTC. Our SEKF implementation is therefore a type of extended Kalman smoother and takes advantage of observations ahead of the update time.

The SEKF is a one-dimensional data assimilation method and takes only vertical covariances between 2-metre-temperature and soil moisture into account. Thus, we need 2-metre-temperature observations at the every grid-point of the CLM grid. In operational data assimilation, the 2-metre-temperature observations are interpolated to the land surface grid with optimal interpolation (Rosnay et al., 2013; "IFS Documentation CY47R1 - Part II: Data Assimilation" 2020), which can be seen as Kalman filter with a fixed gain matrix. Compared to this optimal

interpolation procedure, we use a simplified method to generate the 2-metre-temperature observations, leveraging that we know the true 2-metre-temperature field from our nature run. We bilinearly interpolate the true 2-metre-temperature field from the COSMO grid to the CLM grid, as also done in OASIS3. Afterwards, we disturb the 2-metre-temperature at every grid point by independent and identically distributed random noise $\epsilon_t^o \sim \mathcal{N}(0, 0.01\,\mathrm{K}^2)$. The $\mathbf{R}$-matrix for the Kalman filter thus contains only one entry for every grid point and is given by $\sigma^o = 0.1\,\mathrm{K}^2$ as observational error standard deviation.

For the estimation of the Kalman gain (3.11), we have to linearize the observation operator $\mathbf{H}_t$ around the prior state $\mathbf{x}_t^b$. We employ a finite-differences' approximation, as it was operationally used (Hess, 2001; Rosnay et al., 2013). First, we assume that the effect of soil moisture perturbations on the 2-metre-temperature remain locally at the grid-point above the soil moisture, and we neglect advection of these perturbations. Secondly, we update and perturb the soil moisture only in the first seven layers, up to a depth of 0.62 m, because increments because of data assimilation below this depth would not be physically explainable. For every layer, we have to create an additional smoothing run to estimate the finite-differences' approximation for that specific layer. As perturbation, we add one standard deviation of $\delta\mathbf{x}_t^{b(i)} = 0.01\,\mathrm{m}^3\mathrm{m}^{-3}$ to the soil moisture in that i-th layer, as it is done in (Rosnay et al., 2013). The finite-differences' approximation to the linearized observation operator for the i-th layer is then,

$$\mathbf{H}_t^{(i)} = \frac{H(\mathbf{x}_t^b + \delta\mathbf{x}_t^{b(i)}) - H(\mathbf{x}_t^b)}{\delta\mathbf{x}_t^{b(i)}}. \tag{4.1}$$

Because we use the 2-metre-temperature field at another time than the update time for the soil moisture, our observation operator $H(\cdot) = (H_{T2m} \circ M_{t \mapsto t+\frac{1}{2}})(\cdot)$ is a composition of the 2-metre-temperature observation operator, built-in COSMO, and TerrSysMP as dynamical model, propagating the state for half a day from 00:00 UTC to 12:00 UTC.

4.3 Experiments

In this section, we shortly explain our experimental strategy. Our experiments are based on the idealized twin experiment setup, as explained in Chapter 2. In our experiments, we create an ensemble to investigate interactions between temperature in the atmospheric boundary layer and soil moisture, as described in Section 2.3 Our experiments are based on a perfect model assumption such that we use the same model configuration for every run, as depicted in Table 4.1.

We start the model runs for all of our experiments at 2015-07-30 00:00 UTC. The first 36 hours of simulation are used as spin-up such that perturbations can propagate from the soil into the atmosphere. After this spin-up phase, starting at 2015-07-31 12:00 UTC, we start with our six different experiments. We simulate a period of one week (seven days) and finish our experiments at 2017-08-07 18:00 UTC.

Table 4.1: General environment configuration

Variable	Value
Atmospheric model	COSMO 4.21
Horizontal resolution	~ 2.8 km
Grid points	Lat: 109, Lon: 99
Vertical levels	50
Soil model	CLM 3.5
Horizontal resolution	~ 1 km
Grid points	Lat: 302, Lon: 267
Vertical levels	10
Data assimilation	LETKF + SEKF
Inflation	Prior mult. inflation ($\gamma = 1.006$)
Hori. localization	Gaspari-Cohn (15 km)
Vert. localization	Atmosphere: GC (0.3 ln hPa)
	Soil: GC (0.7 m)
Available observations	99
Observational error	0.1 K

We shortly describe the experiments in the following; their abbreviations are given in Table 4.2.

We define a single and deterministic run without data assimilation (NATURE), the so-called nature run, as our reality in this study. From this run, we synthesize 99 2-metre-temperature observations. An open-loop ensemble forecast without data assimilation (ENS) is used as comparison for the scores in section 4.4 and to investigate the evolution of the ensemble spread with regard to initial ensemble perturbations. Starting from this open-loop ensemble at 2015-07-31 12:00 UTC, we conduct our two LETKF experiments. In the LETKF Soil experiment, we assimilate the 2-metre-temperature with a LETKF to update the soil moisture only. This setting can be seen as weakly-coupled data assimilation experiment and is mainly used as comparison to the SEKF. We expect that most of the soil-generated perturbations in the atmospheric boundary layer can be found in the atmospheric boundary layer temperature. Based on this expectation, we additionally update the atmospheric temperature together with the soil moisture in the LETKF Soil+Temp experiment. We cast this experiment as baseline experiment for strongly-coupled data assimilation, and we will extensively evaluate this experiment in the second part of the results.

In another experiment, we run an open-loop deterministic forecast without data assimilation (DET). We initialize this deterministic forecast with the same initial values as the ensemble mean. Hence, we expect that the errors of this deterministic forecast are comparable to the open-loop ensemble mean. The deterministic run is the baseline experiment for the SEKF experiment. In this SEKF experiment, we assimilate grid-point-based the 2-metre-temperature at 12:00 UTC into the soil moisture at 00:00 UTC, the night before, as described in Section 4.2.1. Because the SEKF is a smoothing algorithm, we already make use of the pseudo-observations

Table 4.2: Experiment abbreviations with experiment description about the run type, assimilation scheme and which variables are updated

Experiment name	run type (members)	scheme	updated variable	description
NATURE	deterministic	-	-	Reference run
ENS	ensemble (40)	-	-	Open-loop ensemble
LETKF Soil	ensemble (40)	LETKF	soil moisture	weakly-coupled EnKF
LETKF Soil + Temp	ensemble (40)	LETKF	soil moisture + atmospheric temperature	strongly-coupled EnKF
DET	deterministic	-	-	Open-loop deterministic
SEKF	deterministic	SEKF	soil moisture	weakly-coupled SEKF

at 2015-07-31 12:00 UTC and start our SEKF experiment at 2015-07-31 00:00 UTC.

4.4 Results

We structure this section into two general parts. In the first subsection, we will compare our experiments and show what we can learn from this comparison. Afterwards, we analyse the LETKF Soil+Temp experiment more in detail with regard to driving factors in the assimilation.

We can expect that assimilating the 2-metre-temperature into soil moisture improves the forecast of the atmospheric boundary layer (e.g. Carrera et al. (2019)). We will analyse in a first step the impact of data assimilation into soil moisture on the prognostic boundary layer temperature (Figure 4.1) in 10 m height above ground, the lowest prognostic model level.

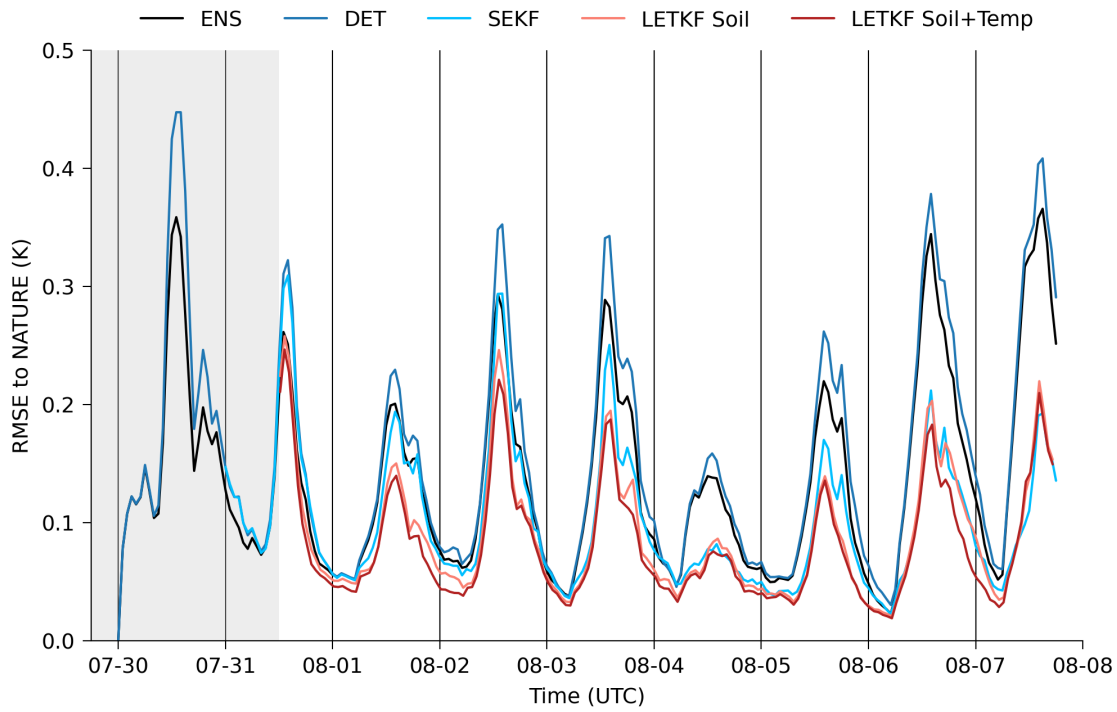


Figure 4.1: RMSE and ensemble spread of different experiments for temperature in 10 metres height as area mean relative to the RMSE of ENS. Different colours denote different experiments. The red colours indicate LETKF experiments with an ensemble, while the bluish colours represent experiments based on a deterministic run. All solid lines show the RMSE, while the red dashed line is the mean ensemble spread over all grid points. The grey-shaded region is the spin-up phase.

Every data assimilation experiment (SEKF; LETKF Soil; LETKF Soil+Temp) has a substantially lower Root-Mean-Squared-Error (RMSE) to NATURE than their counterpart without data assimilation (DET; ENS, Table 4.3). Because this result is found throughout the experiments, this improvement is independent of additional updates in the atmospheric boundary layer. This result confirms previous studies that updating the soil moisture with 2-metre-temperature observations has a

positive assimilation impact on the forecast of the atmospheric boundary layer.

Table 4.3: Spatial and temporal root-mean squared error for depicted experiments compared to NATURE with hourly data from 2015-07-31 13:00 UTC to 2017-08-07 18:00 UTC, representing the background trajectory of the experiments.

Name	T2m (K)	H2O (m^3m^{-3})
ENS	0.158	0.0169
LETKF Soil	0.105	0.0114
LETKF Soil + Temp	0.098	0.0112
DET	0.178	0.0171
SEKF	0.118	0.0145

All experiments have a clearly defined diurnal cycle in the RMSE with the highest errors during day-time. We find the same diurnal cycle in the data assimilation impacts with the highest impacts during day-time and only small impacts during night. Perturbations within the atmosphere are only a result of initial soil perturbations or data assimilation, because our lateral boundary conditions are the same for every run. During day-time, the coupling between land surface and atmospheric boundary layer propagates perturbations into the atmosphere, whereas these two compartments are decoupled during night-time. The collapse of the atmospheric boundary layer in the evening leads to a strong decrease of the atmospheric perturbations. Due to this process, collected information by data assimilation from the day before is also partially lost.

The LETKF Soil+Temp experiment has the smallest error of all experiments, indicating a small positive impact of additionally updating the atmospheric temperature. Nudging the atmospheric temperature to the observations helps us to reduce error components related to a drift of trajectories compared to the NATURE run. By construction of the experiment, the largest part of errors are nevertheless soil-induced, which limits the additional impact of updating the atmospheric temperature. We additionally have a loss of information due to the collapsing boundary layer, as discussed before, and the differences between the LETKF Soil and the LETKF Soil+Temp experiment remain small over the simulation window.

The assimilation impact of the SEKF experiment is similar to the impact of the LETKF Soil experiment, despite the fact that the latter experiment has a smaller absolute magnitude of error. Because the same decreased error can be noticed between the DET and ENS experiment, the smaller errors of the LETKF Soil experiment are mainly accountable to the difference in the type of experiment. Based on this result, both data assimilation methods, the SEKF and LETKF, are similar effective in reducing errors in the atmospheric boundary layer temperature by updating the soil moisture only.

We can expect that the assimilation of 2-metre-temperature observations into soil moisture has also a positive impact on the soil moisture in root-depth, if we improve the forecast of the boundary layer temperature based on the coupling be-

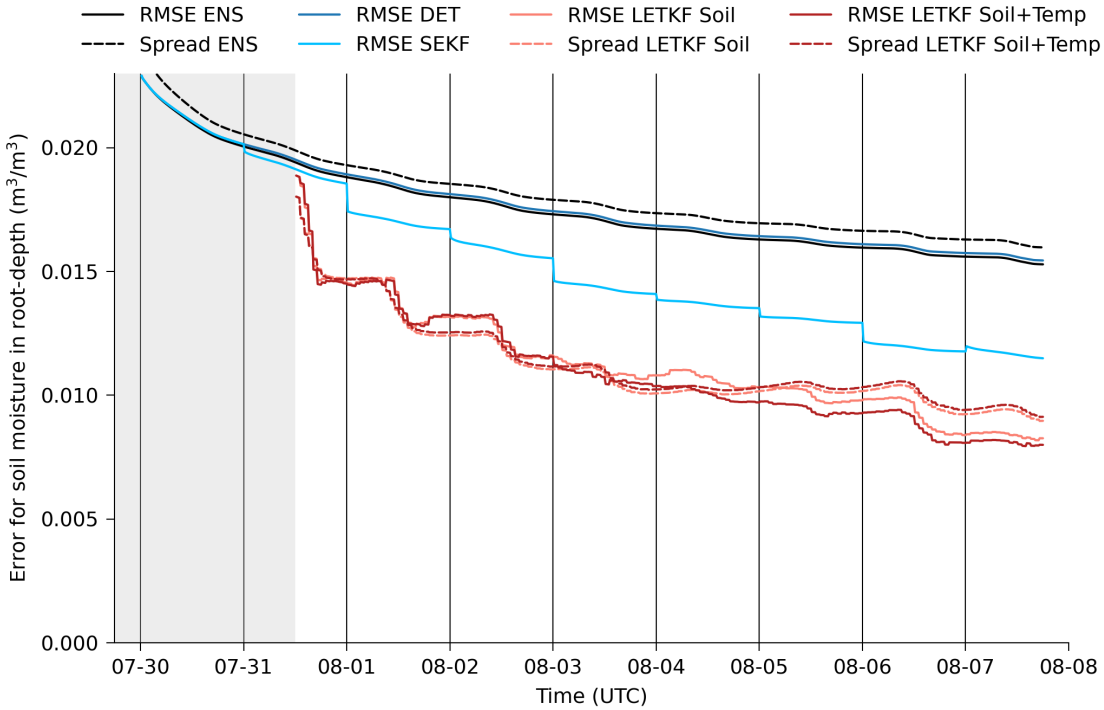


Figure 4.2: RMSE and ensemble spread of different experiments for soil moisture in root-depth as area mean. Different colours denote different experiments. The red colours indicate LETKF experiments with an ensemble, while the bluish colours represent experiments based on a deterministic run. All solid lines show the RMSE, while the red dashed line is the mean ensemble spread over all grid points. The greyish shaded region is the spin-up phase, where no data assimilation experiment was run.

tween atmosphere and land. The error of all experiments reduces with time as the soil dries out in the simulation period (Figure 4.2) Furthermore, data assimilation decreases the error in the SEKF, LETKF Soil, and LETKF Soil+Temp experiment compared to their corresponding reference experiments (DET; ENS). The positive assimilation impact in the LETKF experiment is a result of corrections during daytime, whereas a neutral assimilation impact prevails at night. This diurnal cycle in the impact again reflects the relevance of the coupling strength between land surface and atmosphere, and the flow-dependent background error covariances of the LETKF can represent the situation- and time-dependent coupling strength. The LETKF is therefore able to improve the soil moisture analysis by hourly data assimilation.

The LETKF Soil experiment has a smaller error than the SEKF experiment, but they have similar impacts on the boundary layer temperature. This increased impact in soil moisture is a result of filtering instead smoothing, used in the SEKF experiments. The SEKF can correct foreseeable errors at noon in advance, whereas we only correct instantaneous errors in the filtering framework. Smoothing has thus an advantage compared to filtering for correcting errors in the atmospheric boundary layer based on updates of the soil moisture. For soil moisture, the information content of a single update step is limited by the coupling strength.

Hence, hourly updating the soil moisture with the LETKF is capable to extract more information from limited observations than the SEKF with a fully-observed field and a single update step per day.

Additional nudging of the simulated boundary layer temperature towards the observed temperature results in a positive impact in the LETKF Soil+Temp experiment compared to the LETKF Soil experiment. By updating the boundary layer temperature, we increase the consistency in the analysis errors, which has also a positive assimilation impact on later cycles. In soil, this positive assimilation impact is accumulated over time, and the error of the LETKF Soil+Temp experiment is further reduced in comparison to the LETKF Soil experiment.

Up to this point, we only looked into the error development of either the temperature at the lowest atmosphere layer or the soil moisture in root-depth as spatial mean. In the following, we will analyse how the assimilation impact is spatial distributed (Figure 4.3) in the LETKF Soil and SEKF experiment.

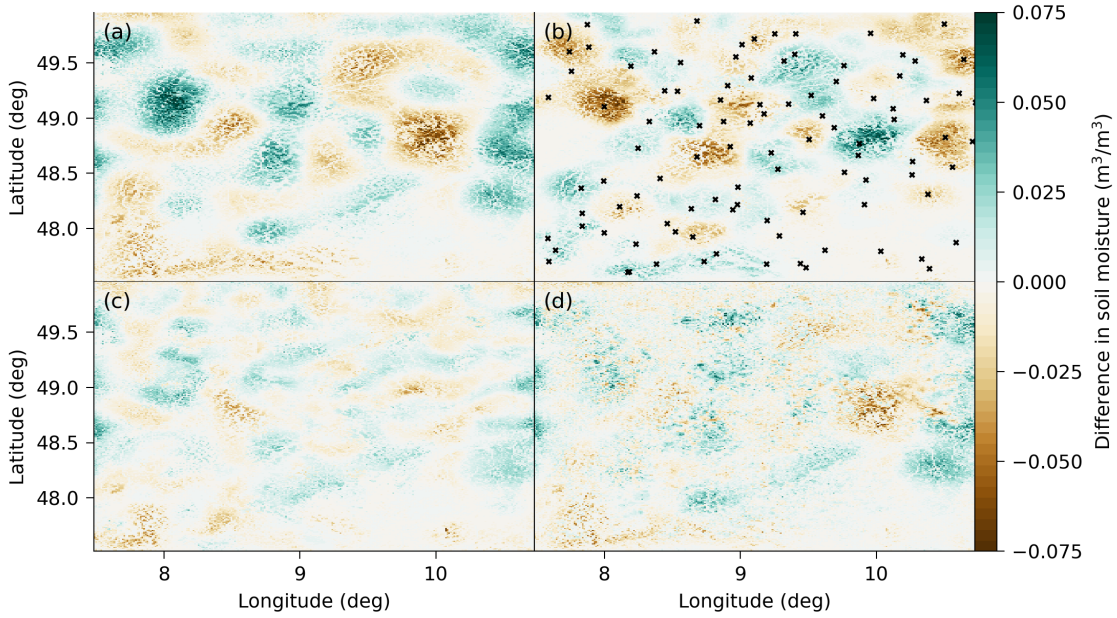


Figure 4.3: Spatial impact of data assimilation in the ensemble and deterministic experiments at the last time step. The upper panels show the error of the ENS experiment (ensemble mean, a) to the NATURE run and the increment of the LETKF Soil (ensemble mean, b) to the ENS (ensemble mean) experiment. The lower panels are the error of the LETKF Soil (ensemble mean, c) and the SEKF (d) experiment to the NATURE run. Blue colours indicate a positive difference, whereas brown colours represent a negative difference. The black crosses in (b) indicate the observational positions as in Figure 2.1.

The spatial distribution of the error in the ENS experiment (Figure 4.3, a) compared to the NATURE run is caused by the initialization of the experiments. Processes leading to a change in patterns within the soil moisture act on longer time-scales than our seven-day simulation time, especially in time periods without large precipitation events. Our initial errors, induced by correlated Gaussian fields, thus

determine the amplitude and patch size of the errors in the experiments without data assimilation.

Data assimilation of the 2-metre-temperature corrects the initial errors, as shown in the accumulated increment of the LETKF Soil experiment compared to the ENS experiment (Figure 4.3, b). This increment depends not only on the error size, but also on the observational positions and chosen localization radius. Nevertheless, the amplitude and patch size of the increments have a similar order of magnitude as the errors, showing that the number of observations and horizontal localization radius are enough and well tuned, respectively.

These increments also influence the remaining error of the LETKF Soil experiment compared to the NATURE run (Figure 4.3, c). Errors are especially dampened in this experiment, if observational position and initial condition errors match. The construction of the ensemble perturbations (Hunt et al., 2007) and spatial localization in the LETKF lead to a spatial smoother error field than for the SEKF experiment (Figure 4.3, d). The SEKF experiment has also higher error amplitudes than the LETKF Soil experiments, showing the effectiveness of the LETKF in this case. Furthermore, the one-dimensional approximation in the SEKF results in error fluctuations across a small area, which are not apparent in the errors and LETKF Soil experiment. The LETKF Soil experiment has thus a spatially more balanced and higher impact than the SEKF experiment, especially in the eastern part of the domain.

In the following (Figure 4.4), we will show the RMSE for soil moisture in root-depth of the offline experiments based on the SEKF trajectory (Figure 4.4, a) and on the LETKF Soil+Temp trajectory (Figure 4.4, b).

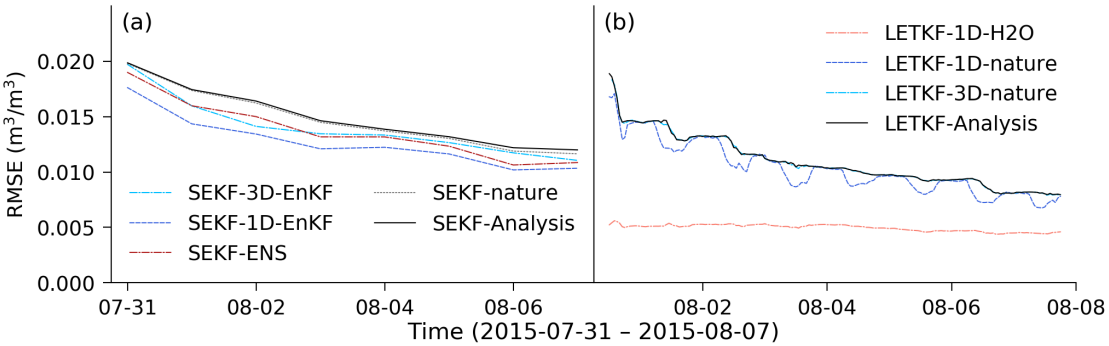


Figure 4.4: RMSE of offline data assimilation experiments for soil moisture in root-depth based on (a) the SEKF and (b) the LETKF Soil+Temp background trajectory. Blueish colours indicate an assimilation of 2-metre-temperature observations from the NATURE run with an ensemble Kalman filter and without observation error, whereas the soil moisture is directly assimilated in the salmon-coloured 1D-H2O experiment in b). The original analyses are black and the SEKF analysis based on perfect NATURE run observations greyish dotted. Only the ensemble mean was updated in column-based 1D EnKF experiments, whereas a full LETKF was used in the 3D EnKF experiments.

The comparison between an experiment with perfect observations, extracted from the NATURE run, and disturbed observations allows us to get an impact of the random observational error. For the SEKF base trajectory (Figure 4.4, a), the difference between an offline experiment with observations from the NATURE run, denoted SEKF-nature, and the original analyses is small. Based on these marginal differences, random observational errors have only a negligible impact on the errors of the SEKF trajectory.

In the SEKF-ENS experiment, we replace the finite-differences' approximation for the Jacobians in Eq. (4.1) by an ensemble approximation from the ENS experiment. We make here the assumption that the ENS experiment is like an external ensemble data assimilation cycle with constrained perturbations in the atmosphere, since we use the same lateral boundary conditions in all experiments. This offline experiment thus resembles the current SEKF implementation at the ECMWF (ECMWF, 2019), except the fact that we do not restart the trajectory within this offline experiment. The error compared to SEKF-nature is reduced, indicating that the ensemble approximation stabilizes the Jacobians in comparison to the finite-differences' approximation.

We take dynamic background covariances from the ENS experiment into account in the SEKF-1D-EnKF experiment, where we use an EnKF instead of a SEKF. In this experiment, we further reduce the error compared to the SEKF-ENS experiment. This error reduction has two reasons: On one hand, we have dynamic covariances, which resemble the flow-dependent uncertainties. On the other hand, the ensemble spread of ENS experiment is larger than the analysis error of the SEKF experiment and the static background covariances for the SEKF. We thus overestimate the assimilation impact in the SEKF-1D-EnKF experiment, which is then a lower bound for the SEKF error.

We replace the column-based data assimilation with a LETKF-based assimilation of 99 discrete observation points in the SEKF-3D-EnKF experiment. Here, we assimilate with a LETKF, based on the perturbations from the ENS experiment, observations from the NATURE run at 12:00 UTC into the background trajectory of the SEKF experiment at 00:00 UTC. This increases the analysis error compared to the SEKF-1D-EnKF experiment, because we have only limited observations compared to a fully observed field. Nevertheless, the error of the SEKF-3D-EnKF experiment is smaller than the SEKF-nature, showing that the ensemble-based assimilation is preferable to a finite-differences-based SEKF.

Similar results can be seen in the offline data assimilation experiments based on the LETKF Soil+Temp experiment (Figure 4.4, b). Replacing the disturbed observations with perfect observations in the LETKF-3D-nature analyses has almost no impact on the analyses error. Using a fully-observed field in the LETKF-1D-nature reduces the error compared to the LETKF-3D-nature experiment, similar to the error reduction in the SEKF experiments. Nevertheless, the impact of a fully-observed field is low compared to the accumulated error reduction in the LETKF Soil+Temp experiment. The LETKF thus assimilates effectively limited

boundary layer observations across the atmosphere-land inferface.

We directly assimilate the soil moisture in root-depth in the LETKF-1D-H2O experiment. With this direct assimilation, we deactivate the source of uncertainty within the vertical covariances, translating from 2-metre-temperature to soil moisture in root-depth. The margin between LETKF-1D-H2O and LETKF-1D-nature is thus representative for the assimilation impact associated to the coupling between atmosphere and land. Based on this margin, the coupling between the 2-metre-temperature and soil moisture dominantly controls the assimilation impact on soil moisture, also during day-time.

The sensible heat flux acts as main coupler between soil moisture and 2-metre-temperature, whereas the evapotranspiration has a bigger impact on humidity in the atmosphere. Based on these physical considerations, we will now show the dependency of the sensible heat flux on the soil moisture (Figure 4.5).

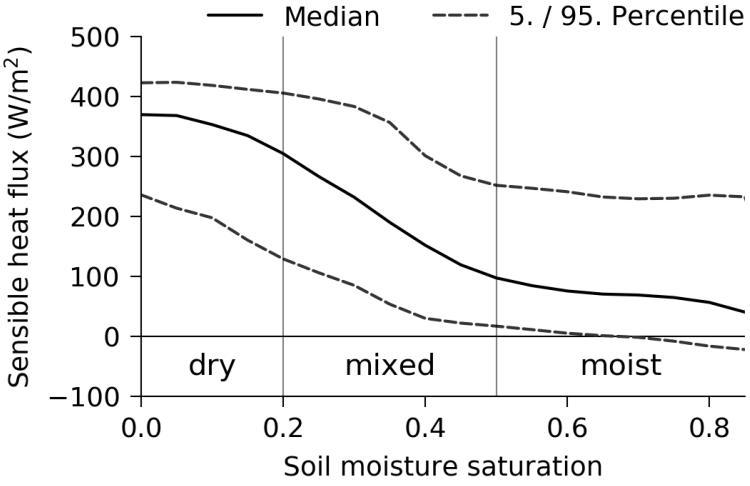


Figure 4.5: The sensible heat flux in dependence on the root-depth soil moisture saturation at 14:00 UTC. All values are estimated based on all ensemble members in the LETKF Soil+Temp experiment, all grid points and all days between 2015-07-31 to 2015-08-07 for 14:00 UTC. The black line is the median over the binarized heat flux (ΔSAT $= 0.05$), whereas the dotted lines shown the 5. and 95. percentile.

Based on the non-linear dependency of the sensible heat flux on the soil moisture (Figure 4.5), we can expect different assimilation impacts for different soil moisture regimes. The sensible heat flux reaches its maximum values in the dry regime, where the ensemble mean soil moisture saturation is below 0.2. Here, near the wilting point, changes in soil moisture have only a small impact on the sensible heat flux, because there is nevertheless to little moisture for plants and their evapotranspiration. The same insensitivity can be found in the moist regime, where the ensemble mean saturation is above 0.5. Plants have in this regime enough water for transpiration and the sensible heat flux is almost insensitive to changes in soil moisture. Hence, the sensible heat flux value is more influenced by other factor, as indicated by higher variances across a soil moisture bin, and

we expect here the smallest assimilation impact. In the mixed regime, where the saturation is between 0.2 and 0.5, plants regulate their transpiration based on the soil moisture, leading to higher sensitivities in the sensible heat flux to changes in the soil moisture. We would therefore expect that most available information from the 2-metre-temperature for the soil moisture is encoded within this mixed regime.

In Figure 4.6, we classify the soil moisture with these three regimes to show its influence on the potential assimilation impact in soil moisture itself. Based on the LETKF-1D-nature experiment from Figure 4.4, we use a potential assimilation impact, which would be the assimilation impact on the soil moisture in root-depth, if we would observe the whole 2-metre-temperature NATURE field. We define here the potential assimilation impact as the difference in the area mean RMSE to the NATURE run from the analysis of the LETKF-1D-nature experiment to the background of the LETKF Soil+Temp experiment for soil moisture.

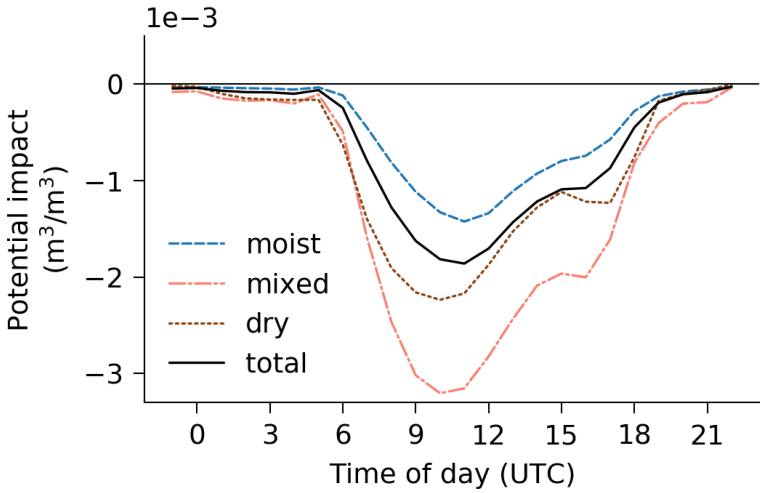


Figure 4.6: Area mean diurnal cycles for the potential assimilation impact, valid from 2015-07-31 19:00 UTC to 2015-08-07 18:00 UTC. We have a positive assimilation impact for negative values and vice versa. The grid point are classified based on the soil moisture saturation classes in Figure 4.5 and the background ensemble mean soil moisture saturation.

The soil moisture saturation clearly determines the potential assimilation impact (Figure 4.6), as previously expected. We find the highest potential impact in grid points with mixed regime, where the sensible heat flux has the highest sensitivity to changes in the soil moisture. The assimilation has its lowest impact in the moist regime, because the sensible heat flux has here its least sensitivity to changes in soil moisture and is mostly influenced by other factors. For our seven-day simulation, we conclude that the soil moisture itself is a main factor to explain variabilities in the assimilation impact across grid points.

In all regimes, we have a positive assimilation impact during day-time, whereas a negligible impact during night. The solar irradiance is the main driver for the

coupling between atmospheric boundary layer and land surface and shapes also the diurnal cycle of the assimilation impact. Nevertheless, in the late afternoon the potential impact deviates from its expected diurnal cycle, which cannot be explained by solar irradiance alone. This potential impact deviation indicates a mechanism, which reinforces the positive assimilation impact in the late afternoon.

In the following, we will reveal that the coupling is additionally controlled by the temporal development of the atmospheric boundary layer (Figure 4.7), leading to the deviation in the late afternoon. We analyse this temporal development within the LETKF Soil+Temp experiment.

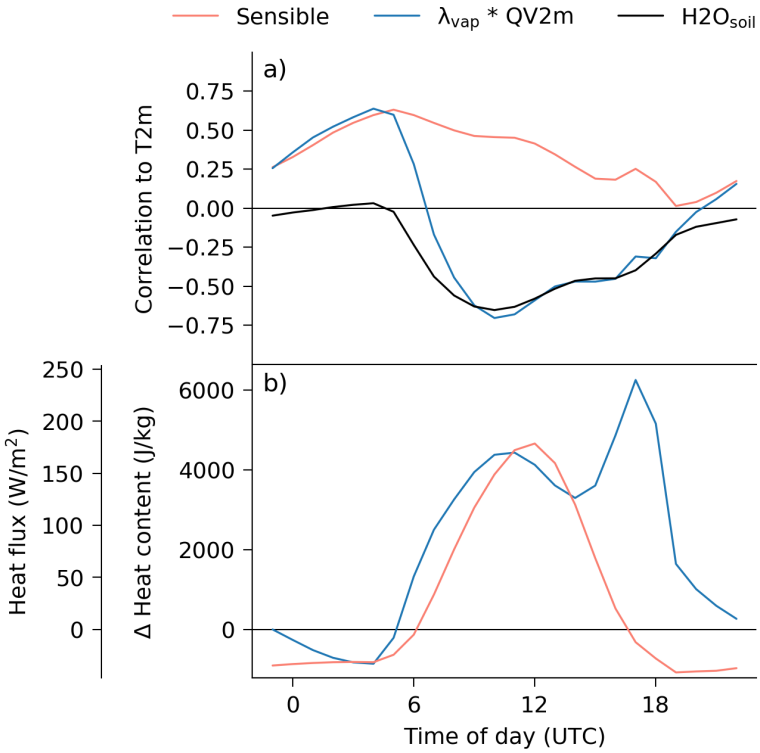


Figure 4.7: Area mean diurnal cycles for: a) the ensemble correlations for different variables (sensible heat flux, water vapour heat content in humidity in 2 metre height, and soil moisture) to the 2-metre-temperature in the LETKF Soil+Temp experiment; b) Change of the heat content relative to 23:00 UTC for the heat content in humidity, and the sensible heat flux, in the LETKF Soil+Temp experiment as average over all ensemble members, valid for 2015-07-31 19:00 UTC to 2015-08-07 18:00 UTC. The water vapour heat content is estimated based on a constant latent heat of vaporization $\lambda_{vap} = 2.501 \times 10^6\,\mathrm{J\,kg^{-1}}$ and the specific humidity in 2 metre height.

As main driver for the assimilation impact, the coupling between land surface and atmosphere correlates the soil moisture to the 2-metre-temperature (Figure 4.7, a). Driven by this coupling, perturbations in the 2-metre-temperature accumulate during day-time. This accumulation decreases the impact of the coupler – the sensible heat flux – on the 2-metre-temperature perturbations, which decorrelates the sensible heat flux and the 2-metre-temperature with time. The water vapour content in the lower boundary layer is mainly controlled by evapotranspiration, and thus, negatively correlated to the 2-metre-temperature during day-time. At

night, an increased water vapour content in the atmosphere decreases the radiative cooling of the land surface (Harrison, 1981), which results in a positive correlation to the 2-metre-temperature. After sunrise, and before perturbations in the boundary layer are accumulated, the sensible heat flux has a direct impact on the 2-metre-temperature.

The same reinforcement mechanism, as in the potential assimilation impact, can be found in the correlations of the sensible heat flux and soil moisture to the 2-metre-temperature. The sensible heat flux follows nevertheless a diurnal cycle without any additional peak (Figure 4.7, b). In contrast to this diurnal cycle, the reinforcement mechanism also heavily influences the diurnal cycle of the water vapour content. Based on this fact, we can trace the reason of the reinforcement mechanism back to the growth and collapse of the boundary layer. The land surface heats up with increasing solar irradiance in the morning. With time, the sensible heat flux and evapotranspiration transport the heat into the boundary layer (Stull, 1988), causing an increase in the heat content of the boundary layer. In the afternoon, the solar irradiance decreases again with time such that also differences between boundary layer and land surface decrease, resulting in lower heat fluxes. Together with a growth of the mixed boundary layer, these lower heat fluxes cause a decrease in the heat content few meters above the surface, as seen in the water vapour content. As a consequence of the strong decrease in solar irradiance, the near-surface boundary layer collapses into a thin strongly-stratified boundary layer. Propagated heat is now stored within this thin layer, leading to a rapid increase in the heat content. This rapidly increased heat content then also strengthens the atmosphere-land coupling above the land surface in the late afternoon.

The atmosphere-land coupling controls the information content encoded within the vertical covariances. In the following, we will also take horizontal covariances and the impact of localization into account and take a deeper look into the dependence of the diagonal covariance on the horizontal distance between 2-metre-temperature and soil moisture (Figure 4.8).

As previously stated, we have negative error covariances during day-time, whereas we have slightly positive error covariances in the evening and night. The ensemble covariances at 2015-08-01 12:00 UTC and 2015-08-03 06:00 UTC resemble the error covariances for local areas. Nevertheless, the ensemble covariances show in both cases too wide horizontal covariances compared to the error covariances, and here, horizontal localization helps to reduce the impact of these spurious correlations. The chosen localization radius of 15 km is too small for 2015-08-01 12:00 UTC and reduces the impact of horizontal covariances too strongly compared to the error covariances, whereas the radius is well-tuned for 2015-08-03 06:00 UTC. At 2015-08-01 19:00 UTC, the ensemble cannot represent the positive error covariances, and we would expect a negative assimilation impact. In this case, the best localization would be 0 km, indicating that a deactivation of the assimilation would be the best choice. The correct localization radius for cross-compartmental data assimilation

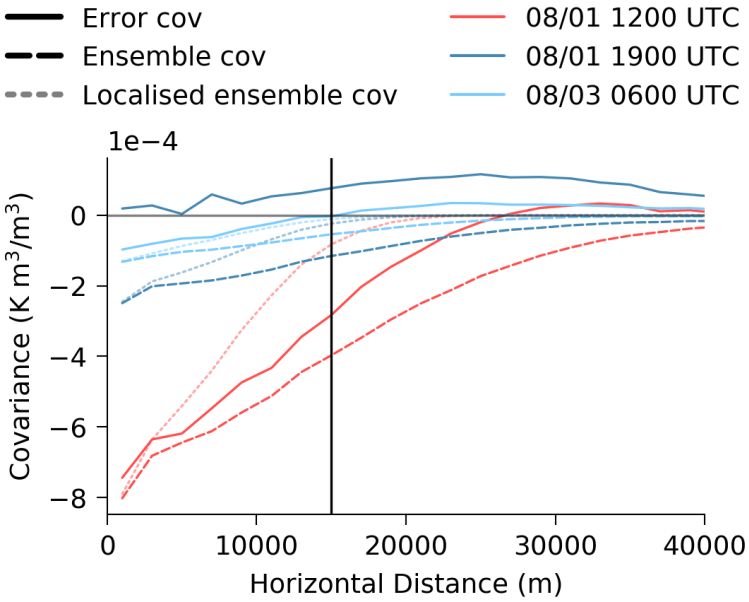


Figure 4.8: Covariances between 2-metre-temperature and soil moisture in root-depth as function of the distance between observation and grid point for the LETKF Soil+Temp experiment. The error covariances are estimated based on the ensemble mean errors with the covariance statistics estimated over the corresponding bin.s The ensemble covariances are a binned mean of the ensemble covariances, whereas the covariances are multiplied by the localization factor for the localized ensemble gain. The covariances are estimated based on randomly sampled 2000000 pairs of grid points.

is therefore highly dependent on the governing processes.

4.5 Discussion and Summary

In this study, we investigate how we can use an ensemble Kalman filter (EnKF) to assimilate sparse 2-metre-temperature observations across the atmosphere-land interface. Because we focus on the coupling between temperature in the atmospheric boundary layer and soil moisture, we perturb only initial soil conditions to generate an ensemble of forecasts. All resulting deviations within the ensemble and between different experiments are therefore only a consequence of these initial soil conditions or due to data assimilation. With this idealized experimentation framework, we are able to prove that the soil moisture analysis can be improved by assimilating boundary layer observations.

The coupling of the land surface to the boundary layer drives this positive assimilation impact during day-time, whereas we have a neutral impact at night. An EnKF with hourly filtering can exploit this coupling, if the ensemble covariances are representative for the error covariances. To shape the ensemble covariances, a well-tuned horizontal localization is crucial for the cross-compartmental assimilation. In the case of representative ensemble covariances, additional updates of the boundary layer temperature increase the consistency of the analysis increments, which has an additional positive assimilation impact on subsequent soil moisture

analyses. This additional assimilation impact hints at a positive consequence of strongly-coupled data assimilation at the atmosphere-land interface.

The EnKF has smaller errors than the simplified extended Kalman filter (SEKF) to our nature run in both, the soil moisture and boundary layer temperature. The EnKF improves hereby the soil moisture analysis by a larger amount than the boundary layer forecast compared to the SEKF. Our offline data assimilation experiments reveal that this is related to the finite-differences' approximation within the SEKF, which can be stabilized by using ensemble-based covariances. We further improve the soil moisture analysis with hourly-based filtering, as it is commonly used for data assimilation in the atmosphere. This improvement by filtering indicates that we can include land surface variables in the ensemble-based analysis cycles of the atmosphere.

With a localized EnKF, we can skip the optimal interpolation step to create a 2-metre-temperature analysis. We find with our offline data assimilation experiments that the additional assimilation impact of a fully-observed 2-metre-temperature field is small compared to the general assimilation impact with coarsely-distributed observations. Furthermore, the additional optimal interpolation step creates uncertainties in the temperature observations, which we have not taken into account in our offline data assimilation experiment. Three-dimensional ensemble-based data assimilation of boundary layer observations for the soil moisture is thus possible with localization.

We have a non-linear coupling between atmospheric boundary layer and land surface, because the strength of the coupling depends on the soil moisture itself. We only make a local linear assumption around the ensemble mean in the ensemble Kalman filter, and these non-linearities do not have a large impact on the results. The global non-linear structure nevertheless constrains the coupling between the atmosphere and land, and above very dry and humid soils, only limited information content is encoded in observations, which is extractable by direct assimilation of the observations.

Beside this dependence of the assimilation on the coupling and on the soil moisture, we also show that the temporal development of the boundary layer has an impact. This impact leads to a peak in information content around noon, whereas we have a decrease in the afternoon. A partial collapse of the boundary layer into a thin layer above the land surface initiates a reinforcement of the atmosphere-land coupling. We can more easily use the temporal development with hourly-filtering, whereas we might have problems with daily-smoothing as done within the SEKF, because we would have to select representative observation times.

We can further exploit the temporal development of the boundary layer with hourly-smoothing instead of hourly-filtering. Because land surface perturbations need some time to propagate into the atmosphere, one possibility would be to assimilate future observations within a given assimilation window and a 4D-LETKF (Harlim and Hunt, 2007; Kalnay et al., 2007b), which would be similar

to an iterative ensemble Kalman smoothing scheme (Kalnay and Yang, 2010; Sakov et al., 2012; Bocquet and Sakov, 2014). Together with smoothing, we could additionally introduce time-dependent localization to tackle problems related to errors by the ensemble approximation of the covariances.

All in all, our results support the view that assimilation of boundary layer observations has a positive impact on the soil moisture, if the model system can adequately represent the governing processes in the boundary layer and land surface. We can therefore see this study as first step towards the goal of assimilating a unified set of observations across the atmosphere-land interface to improve the analysis for both compartments.

4.6 Conclusions

In this study, we assimilate synthetic 2-metre-temperature observations into soil moisture in a fully-coupled limited-area model system for a seven-day period in Summer 2015. Based on our results in idealized twin experiments, we conclude the following:

1. Assimilation of boundary layer observations improves the soil moisture analysis during day-time and has no impact during night; boundary layer observations yield the highest information content for land surface data assimilation above soil moisture saturations between 0.2 and 0.5.

2. Hourly-updating the soil moisture with a Localized Ensemble Transform Kalman filter results in a smaller error for the soil moisture analysis than daily-smoothing with a Simplified Extended Kalman filter, and in addition, we can directly assimilate sparse boundary layer observations across the atmosphere-land interface without an intermediate optimal interpolation step.

3. Ensemble-based approximations of the background covariances and Jacobians stabilizes the analysis increments in a Simplified Extended Kalman filter.

4. Updating the atmospheric temperature together with the soil moisture increases the physical consistency in the analysis for the boundary layer and land surface, which in fact reduces additional errors in the soil moisture analysis.

5. We can merge the decoupled data assimilation cycles – one for the land surface and one for numerical weather prediction – into one strongly-coupled cycle with updates across the atmosphere-land interface and hour-like cycling lengths of the faster atmospheric compartment.

This page intentionally left blank

5

Fingerprint operators to stabilize cross-compartmental data assimilation

In this Chapter, I investigate how I can take advantage of the temporal development of observations in the atmospheric boundary layer to improve the soil moisture analysis. Because Earth system components are temporally dependent on each other, observations from the atmosphere are informed about the state conditions in a neighboring Earth system component with a time lag. During their update step, ensemble Kalman filters cannot take advantage of such an asynchronous information flow, because the analysis is only conditioned on previous and current observations. Contrarily, four-dimensional data assimilation methods are additionally conditioned on future observations. Hence, they are an alternative to take these temporal dependencies into account.

After setting the scene in Section 5.1, I derive an ensemble Kalman smoother based on the ETKF from Chapter 3 in Section 5.2 that creates an analysis at the beginning of an assimilation window, similar to 4DEnVar. To compare this derived smoother with my implementation of a LETKF for the atmosphere-land interface, I conduct similar idealized twin experiments as in Chapter 4, which are described in Section 5.3. As results in Section 5.4, I prove that this smoother with a 24 hour assimilation window can take advantage of cross-compartmental temporal dependencies. To reduce the risk of the smoother to get overconfidence towards the observations, I additionally introduce fingerprint operators using characteristic fingerprints in the 2-metre-temperature that point towards forecast errors in the soil moisture. These fingerprint operators condense the information content from the 2-metre-temperature observations into two observational features. Though, smoothing with my two designed fingerprint operators is more robust against miss-specifications in the localization radius and observational error covariance. On the basis of the results in this Chapter, I therefore propose as my second framework to use fingerprint operators to make cross-compartmental data assimilation

This chapter will be submitted in another form as: Finn, T. S., Geppert, G., and Ament, F.: "Fingerprint operators of atmospheric boundary layer observations to stabilize land surface data assimilation", *to be submitted to Quarterly Journal of Royal Meteorological Society.* As this chapter is intended for publication with multiple authors, I switch in its content to the first person plural ("we") form.

more robust against noise.

5.1 Introduction

The sensible heat flux and evapotranspiration gradually propagate information from the land surface into the atmospheric boundary layer. As the fluxes are mainly driven by the sun, they additionally have a diurnal cycle. Consequently, also the strength of the coupling between the land surface and the atmospheric boundary layer depends on the time of the day. Because the heat fluxes are the main driver for the temporal development of heat and moisture in the atmospheric boundary layer, they constrain in this way the information content of instantaneous atmospheric boundary layer observations about the soil conditions. Despite this constrain, we can instantaneously assimilate boundary layer observations to improve the soil moisture analysis in idealized experiments, as seen in Chapter 4. Because of the existing temporal dependency, we nevertheless expect that boundary layer observations at a different time than the assimilation time have a higher information content than instantaneous observations. In this study, we investigate how we can squeeze more information about the soil conditions out of atmospheric boundary layer observations by taking their temporal development into account.

One way to utilize temporal covariances is to use ensemble Kalman smoothers (EnKS, Leeuwen and Evensen, 1996; Evensen and Leeuwen, 2000; Cosme et al., 2012) instead of ensemble Kalman filters (EnKF, Evensen, 1994; Burgers et al., 1998; Anderson and Anderson, 1999). EnKFs are conditioned on past and current observations only, whereas smoother take also advantage of observations ahead of the update time. In an EnKS, we use observations in an assimilation window ahead of the update time. By using observations in an assimilation window, we model non-instantaneous dependencies between observations and state, which cannot be otherwise used in EnKFs. We use here an Ensemble Transform Kalman filter (ETKF, Bishop et al., 2001; Hunt et al., 2007) and smoother (ETKS), where the analysis is estimated based on ensemble weights. The ensemble weights can be then applied anywhere in the assimilation window (Yang et al., 2009; Kalnay and Yang, 2010) such that also the smoothing solution can be found without the need of any tangent linear model. If we apply the weights at the end of the assimilation window, this procedure leads to a 4D-ETKF (Hunt et al., 2004; Harlim and Hunt, 2007), also operationally used for atmospheric data assimilation (Schraff et al., 2016). By applying the weights at the beginning of the assimilation window, we need to propagate the ensemble a second time through the window to get the filtering solution at the end of the assimilation window. This procedure resembles 4D-Var and specific its ensemble equivalent 4DEnVar (Liu et al., 2008). The application of the ensemble weights at the beginning of the assimilation window is also the linearized variant of the iterative ensemble Kalman smoother (IEnKS, Bocquet and Sakov, 2014) and shows promising results in toy models. However, an open question is if such an ensemble Kalman smoother with weights at the beginning of the assimilation window can improve land surface data assimilation.

The EnKS explicitly makes a linear assumption and models the relationship between 2-metre-temperature and soil moisture based on covariances. As previously seen in Chapter 4, the soil moisture non-linearly influences the sensible heat flux and, hence, also the 2-metre-temperature. Consequently, 2-metre-temperature observations might have a non-linear error fingerprint from the soil moisture. In machine learning, this linear assumption is often relaxed by extracting features out of the observations and regressing these possibly non-linear features to the target variable (Hastie et al., 2009; Rasmussen and Williams, 2006; Murphy, 2012). Until recently in Morzfeld et al., 2018; Rosenthal et al., 2017; Haario et al., 2015, this paradigm of feature engineering was not used in data assimilation. We introduce here this concept for coupled data assimilation across the atmosphere-land interface. The here so-called fingerprint operators explicitly transform the 2-metre-temperature observations into a new feature space. This way, they take advantage of characteristic fingerprints in the 2-metre-temperature that point towards errors in the soil moisture. Since this is a novel methodology for cross-compartmental data assimilation, the effect of these fingerprint operators is unclear.

Coupled land-atmosphere model platforms are often biased compared to the real development of the land-atmosphere system. In this study, we want to show how we can improve the assimilation of 2-metre-temperature observations into land surface models without having to care about model biases and errors. Hence, we use idealized twin experiments together with TerrSysMP (Gasper et al., 2014; Shrestha et al., 2014); a limited-area terrestrial system modelling platform that couples COSMO as model for the atmosphere and CLM as model for the land surface by the OASIS3 coupler. In these experiments, we define a deterministic run, the so-called nature run, as our reality. We conduct this run with the same model configuration as for our data assimilation experiments. On the basis of this reality, we synthesize sparse 2-metre-temperature observations. We assimilate these synthetic 2-metre-temperature observations into the soil moisture. With this idealized modelling setup, we only concentrate on the relationship between 2-metre-temperature and soil moisture.

In our experiments, we compare ensemble Kalman smoother with different assimilation window lengths to ensemble Kalman filters. We elaborate also the question of how an ensemble Kalman smoother might be more successful than an ensemble Kalman filter. A positive assimilation impact would imply that we can take advantage of temporal dependencies with an ensemble Kalman smoother. For possible fingerprint operators, we concentrate on features from the diurnal cycle in the 2-metre-temperature. We start with a feature screening and show which features in the 2-metre-temperature might be suited for cross-compartmental data assimilation. In addition, we do offline data assimilation experiments to compare the effect of data assimilation with fingerprint operators to assimilation without these operators on the update step. In six experiments, we test different assumptions and combinations of fingerprint operators. These last experiments help us to quantify the direct and propagated impact of fingerprint operators on the data assimilation.

We describe the theory of our implemented EnKS and the fingerprint operators in Section 5.2. We elucidate our experimental setup in Section 5.3, where we additionally present the used fingerprint operators for the atmosphere-land interface. In Section 5.4, we show our results for the ensemble Kalman smoothers and fingerprint operators, whereas we summarize and conclude our study in Section 5.5 and 5.6

5.2 Four-dimensional Data Assimilation Environment

In our four-dimensional data assimilation, we take not only vertical and horizontal covariances into account but also temporal covariances. This four-dimensional formulation allows us to assimilate observations in an assimilation window. To assimilate observations in an assimilation window, we have to propagate the state at the beginning of the window throughout the assimilation window. For this propagation step, we use the Terrestrial System Modelling Platform (TerrSysMP, Gasper et al. (2014) and Shrestha et al. (2014)); COSMO (Baldauf et al., 2011) as model for the atmosphere is coupled to the Community Land Model as model for the land surface. Because this model configuration is used throughout the thesis, we refer the reader for more information about the modelling system and setup to Chapter 2.

In this study, the general data assimilation method is based on the localized ensemble transform Kalman filter (LETKF), derived and explained in Chapter 3. In the following, we therefore derive only our ensemble Kalman smoother from a four-dimensional variational cost function. Our derivation closely follows the derivation of the linearized iterative ensemble Kalman smoother in Bocquet and Sakov (2014). Afterwards, we explain the computational costs of the here-considered data assimilation methods. In the end of this section, we introduce the concept of fingerprint operators for cross-compartmental data assimilation and show how their observational covariance can be derived.

5.2.1 ETKS

As explained in Chapter 3, in data assimilation, we are interested in the filtering solution $p(\mathbf{x}_T \mid \mathbf{y}^o_{1:T})$. Instead of sequentially cycling through a propagation step and an update step, we can also update the trajectory once based on all observations within an assimilation window from time 1 to time T. In this study, we specifically investigate the impact of an ensemble Kalman smoother, where we update the initial state at the beginning of the window to get the smoothing solution $p(\mathbf{x}_0 \mid \mathbf{y}^o_{1:T})$. We can propagate this smoothing solution throughout the window to obtain a filtering solution at the end of the assimilation,

$$p(\mathbf{x}_T \mid \mathbf{y}^o_{1:T}) = \int p(\mathbf{x}_T \mid \mathbf{x}_0, \mathbf{y}^o_{1:T}) p(\mathbf{x}_0 \mid \mathbf{y}^o_{1:T}) d\mathbf{x}_0.$$

For this smoothing solution, we want to estimate the state of our model system $\mathbf{x}_0$ at time 0 based on a background forecast $\bar{\mathbf{x}}^b_0$ at the same time and given observations $\mathbf{y}^o_{1:T}$ within an assimilation window from time 1 to time T. Both, the

background forecast and observations, are presumably Gaussian distributed with zero mean and $\mathbf{P}_0^b$ and $\mathbf{R}_t$, respectively, as time-dependent covariances. In addition, the forecast and observations have an associated dynamical model $M_{0\mapsto t}(\mathbf{x}_t)$, mapping a state from time 0 to time t, and an observation operator $H(\mathbf{x}_t)$, translating from state space to observational space. Given the Gaussian distributions of the background and observations and Bayes' theorem, we can formulate a cost function $J(\mathbf{x}_0)$ to optimize our model state, corresponding to the strong-constrained four-dimensional variational data assimilation (4D-Var) cost function (Dimet and Talagrand, 1986; Talagrand and Courtier, 1987),

$$\begin{aligned} J(\mathbf{x}_0) =& (\bar{\mathbf{x}}_0^b - \mathbf{x}_0)^T (\mathbf{P}_0^b)^{-1} (\bar{\mathbf{x}}_0^b - \mathbf{x}_0) \\ &+ \sum_{t=1}^{T} (\mathbf{y}_t^o - H(M_{0\mapsto t}(\mathbf{x}_0)))^T \mathbf{R}_t^{-1} (\mathbf{y}_t^o - H(M_{0\mapsto t}(\mathbf{x}_0))). \end{aligned} \tag{5.1}$$

In ensemble-based data assimilation, we approximate the background forecast and the background covariances by a Monte-Carlo approximation with k ensemble members and $\boldsymbol{\delta}\mathbf{x}_0^{b(i)}$ as ensemble perturbation of the i-th ensemble member at time 0,

$$\bar{\mathbf{x}}_0^b = \frac{1}{n} \sum_{i=1}^{k} \mathbf{x}_0^{b(i)}, \tag{5.2}$$

$$\begin{aligned} \mathbf{P}_0^b &= \frac{1}{n-1} \sum_{i=1}^{k} (\mathbf{x}_0^{b(i)} - \bar{\mathbf{x}}_0^b)(\mathbf{x}_0^{b(i)} - \bar{\mathbf{x}}_0^b)^T \\ &= \frac{1}{n-1} \sum_{i=1}^{n} \boldsymbol{\delta}\mathbf{x}_0^{b(i)} (\boldsymbol{\delta}\mathbf{x}_0^{b(i)})^T. \end{aligned} \tag{5.3}$$

Using these approximations, the assimilation increment of the analyzed model state $\Delta\mathbf{x}_0 = \mathbf{x}_0 - \bar{\mathbf{x}}_0^b$ lies in the space spanned by the ensemble perturbations (Lorenc, 2003; Hunt et al., 2007). Therefore, we can explicitly state the increment as weighted linear combination of the ensemble perturbations with a column-wise matrix of all background perturbations $\boldsymbol{\delta}\mathbf{X}^b$ and weights $\mathbf{w}$,

$$\mathbf{x}_0 = \bar{\mathbf{x}}_0^b + \boldsymbol{\delta}\mathbf{X}_0^b \mathbf{w}, \qquad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, (k-1)^{-1}\mathbf{I}). \tag{5.4}$$

We can express the four-dimensional cost function (5.1) in ensemble space based on this transformation,

$$\begin{aligned} \tilde{J}(\mathbf{w}) =& (k-1)\mathbf{w}^T\mathbf{w} \\ &+ \sum_{t=1}^{T} (\mathbf{y}_t^o - H(M_{0\mapsto t}(\bar{\mathbf{x}}_0^b + \boldsymbol{\delta}\mathbf{X}_0^b\mathbf{w})))^T \mathbf{R}_t^{-1} (\mathbf{y}_t^o - H(M_{0\mapsto t}(\bar{\mathbf{x}}_0^b + \boldsymbol{\delta}\mathbf{X}_0^b\mathbf{w}))). \end{aligned} \tag{5.5}$$

To estimate the minimum of $\tilde{J}(\mathbf{w})$ and find the solution for our model state, we can descent the gradient with respect to the weights,

$$\frac{\partial \tilde{J}(\mathbf{w})}{\partial \mathbf{w}} = (k-1)\mathbf{w} - \sum_{t=1}^{T} \mathbf{Y}_t^T \mathbf{R}_t^{-1}[\mathbf{y}_t^o - H(M_{0\mapsto t}(\overline{\mathbf{x}}_0^b + \boldsymbol{\delta}\mathbf{X}_0^b\mathbf{w}))] \tag{5.6}$$

The first part of the gradient constrains the weights towards 0, whereas the second part is observational-depending and punishes a strong deviation from the observations within the given window between 1 and T. The adjoint $\mathbf{Y}_t^T$ is a short-form for the partial derivative of the propagated state in observational space wrt. to the current weights $\frac{\partial H(M_{0\mapsto t}(\overline{\mathbf{x}}_0^b + \boldsymbol{\delta}\mathbf{X}_0^b\mathbf{w}))}{\partial \mathbf{w}}$ and translates the observational innovations at time t to weights at time 0. We use here a purely ensemble-based approximations to these partial derivatives with a four-dimensional ensemble transform Kalman smoother (ETKS).

The ETKS is based on an approximated linear mapping from weight space into propagated observational space. To estimate the linear mapping, we apply the propagation model and observation operator to every ensemble member independently, here for the i-th ensemble member, $\mathbf{y}_t^{b(i)} = H(M_{0\mapsto t}(\mathbf{x}_0^{b(i)}))$. Afterwards, we linearize the observational operator around the ensemble mean in observational space $\overline{\mathbf{y}}_t^b = \sum_{i=0}^{k} \mathbf{y}_t^{b(i)}$,

$$H(M_{0\mapsto t}(\overline{\mathbf{x}}_0^b + \boldsymbol{\delta}\mathbf{X}_0^b\mathbf{w})) \approx \overline{\mathbf{y}}_t^b + \boldsymbol{\delta}\mathbf{Y}_t^b\mathbf{w}. \tag{5.7}$$

The linearized observation operator $\boldsymbol{\delta}\mathbf{Y}_t^b$ is a column-wise matrix, consisting of all ensemble perturbations in observational space $\boldsymbol{\delta}\mathbf{y}_t^{b(i)} = \mathbf{y}_t^{b(i)} - \overline{\mathbf{y}}_t^b$. Furthermore, this linearized observation operator acts in the ETKS as approximated adjoint $(\boldsymbol{\delta}\mathbf{Y}_t^b)^T \approx \mathbf{Y}_t^T$. Based on this approximated adjoint, (5.5) simplifies to a linear least-squares cost function. We can therefore set the gradient (5.6) to zero and gain an analytical solution for the mean weights $\overline{\mathbf{w}}$ and covariance $\tilde{\mathbf{P}}^a$ in weight space,

$$(\tilde{\mathbf{P}}^a)^{-1} = (k-1)\mathbf{I} + \sum_{t=1}^{T} (\boldsymbol{\delta}\mathbf{Y}_t^b)^T \mathbf{R}_t^{-1} \boldsymbol{\delta}\mathbf{Y}_t^b, \tag{5.8}$$

$$\mathbf{w} = \tilde{\mathbf{P}}^a \sum_{t=1}^{T} (\boldsymbol{\delta}\mathbf{Y}_t^b)^T \mathbf{R}_t^{-1} (\mathbf{y}_t^o - \overline{\mathbf{y}}_t^b). \tag{5.9}$$

To generate a new ensemble based on the found solution $\mathbf{w}$, the ETKS deterministically estimate the i-th ensemble member with an additional weight perturbation $\boldsymbol{\delta}\mathbf{w}^{(i)}$,

$$\mathbf{x}_0^{(i)} = \overline{\mathbf{x}}_0^b + \boldsymbol{\delta}\mathbf{X}_0^b(\mathbf{w} + \boldsymbol{\delta}\mathbf{w}^{(i)}).$$

The weight perturbations are calculated based on the analysis covariance in ensemble space $\tilde{\mathbf{P}}^a$, with $\mathbf{T}$ as column-wise transformation matrix of all ensemble

perturbations in ensemble space,

$$[\boldsymbol{\delta}\mathbf{w}^{(0)}, \boldsymbol{\delta}\mathbf{w}^{(1)}, \ldots, \boldsymbol{\delta}\mathbf{w}^{(k)}] = \mathbf{T} = [(k-1)\tilde{\mathbf{P}}^{a}]^{\frac{1}{2}}.$$

This constructed ensemble represents then the smoothing solution at the beginning of the assimilation window. To get the filtering solution at the end of the assimilation window $p(\mathbf{x}_T \mid \mathbf{y}^o_{1:T})$, we have to propagate the smoothed ensemble throughout the assimilation window with the dynamical model.

This derivation of an ETKS follows closely the derivation of an iterative ensemble Kalman smoother in Bocquet and Sakov, 2014. Hence, the ETKS used-here corresponds to their linearized iterative ensemble Kalman smoother, whereas the adjoint is estimated with the transform variant (Bocquet and Sakov, 2012; Sakov et al., 2012). We derive this ETKS based on the marginal smoothing solution and the variational four-dimensional cost function. This ETKS can be therefore seen as 4DEnVar (Liu et al., 2008; Desroziers et al., 2014; Bannister, 2017) with some modifications. Instead of updating a deterministic run based on ensemble statistics, we update the ensemble mean (5.9) and center the analysis perturbations around this mean. As a consequence, we do not scale the ensemble to estimate the ensemble approximation to the adjoint $\mathbf{Y}^T_t$ and use a simple propagation of the full ensemble. Furthermore, the ensemble is often externally updated in 4DEnVar, whereas we update the ensemble perturbations based on the inverse Hessian (5.8).

Our formulation of the ETKS with its four-dimensional variational cost function can be seen as variant of a marginal fixed-interval ensemble Kalman smoother (Ménard and Daley, 1996; Li and Navon, 2001). In comparison to classical ensemble Kalman smoother (Leeuwen and Evensen, 1996; Evensen and Leeuwen, 2000; Cosme et al., 2012), we do not apply the ensemble weights to update the trajectory in the whole assimilation window, but at the beginning of the window. From there, the ensemble is again propagated throughout the window. This doubled propagation increases the computational costs of our smoother compared to a classical ensemble Kalman smoother, but we expect an advantage in the case of non-linear propagations in the window (Bocquet and Sakov, 2014).

If we restrict our assimilation window to $T = 0$, we assimilate only instantaneous observations at the same time as the estimated state, which equals filtering that was derived in Chapter 3. Such an ensemble transform Kalman filter (ETKF) is one baseline in our experiments, as it was previously used in Chapter 4. Furthermore, we can linearize the first ensemble propagation around the ensemble mean. This allows us to apply the estimated ensemble weights at time $t = 0$ anywhere in the assimilation window (Hunt et al., 2007; Kalnay et al., 2007b; Kalnay and Yang, 2010). Specifically, we consider an application of the ensemble weights at the end of the assimilation window at time $t = T$. In this case, the ETKS equals a four-dimensional ETKF (4D-ETKF, Hunt et al. (2004) and Harlim and Hunt (2007)). The 4D-ETKF smoothes over an assimilation window with past observations and gives a filtering solution.

All three algorithms, the ETKF, the 4D-ETKF, and the ETKS, take the sensitivity of

the model state x_T at time T to observations from a previous time t differently into account. The normal ETKF estimates a filtering solution at time t and propagates this filtering solution to time T with the full dynamical model, as shown in Fig. 5.1, (a). Hence, in the ETKF, the state at T is non-linearly dependent on a previous time. Nevertheless, the ETKF takes temporal dependencies only during the propagation step into account, whereas the update step is restricted to a single time. In contrast, the 4D-ETKF linearizes the sensitivity around the ensemble mean and updates the state at the end of the assimilation window, as shown in Fig. 5.1, (b). The 4D-ETKF takes in this linearized way temporal dependencies during the update step into account, but neglects the non-linear propagation as in the normal ETKF. Our ETKS combines the ETKF and 4D-ETKF in some sense. During the update step of the state at time 0, the ETKS linearizes the sensitivity of the state to the observations ahead of the update time. Again, we can take advantage of temporal dependencies during the update step. Resembling the ETKF, the solution is propagated non-linearly through the assimilation window, as shown in Fig. 5.1, (c). Therefore, we also make use of temporal dependencies during the propagation step.

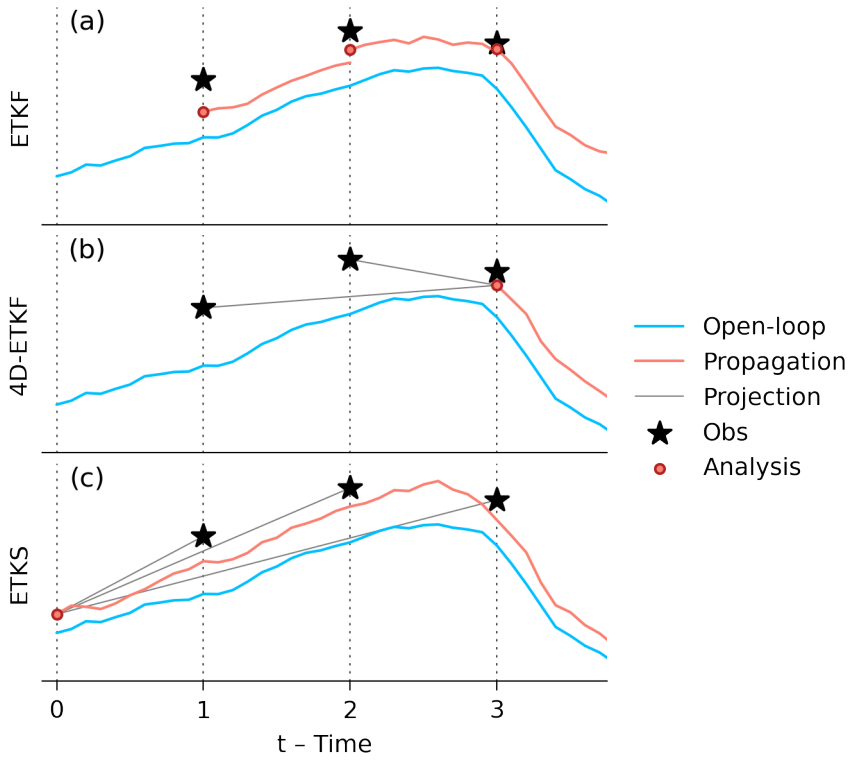


Figure 5.1: An illustrative figure showing the differences between (a) the ETKF, (b) the 4D-ETKF, and (c) the ETKS. The ETKF updates its trajectory at every observation time based on instantaneous observations. The 4D-ETKF collects observations within an assimilation window, linearly projects these observations to the end of the window and updates the trajectory once at this end. The ETKS uses future observations and projects them linearly to the beginning of the assimilation window, where the trajectory is updated once.

From these considerations, we expect that our ETKS combines the strength of the ETKF and 4D-ETKF. In theory, this should lead to an increased assimilation impact in the soil moisture. In the case of a linear model, the solutions of the

normal ETKF, the 4D-ETKF, and the ETKS should be the same, but all with another computational costs, as we show in the following.

5.2.2 Computational Costs

The propagation of the ensemble has the highest computational costs in four-dimensional ensemble-based data assimilation. The costs of weight estimation within the ETKF and ETKS are small compared to the costs for the propagation. Here, we assume that the weight estimation has the same order of magnitude for the computational cost Ω, independent of the number of assimilated observations. We will denote the computational costs for a propagation of one single ensemble member for one single simulation hour as P, whereas k is again the number of ensemble members. In the following, we will derive the computational costs for a T-hours long assimilation window, where we want to get the solution at the end of the assimilation window.

- In the **ETKF**, we hourly update the soil moisture based on instantaneous observations. The costs result into $T \times (P \times k + \Omega)$, because we have T times the costs of a complete ensemble propagation for one hour.
- In the **4D-ETKF**, we assimilate all observations within the window once and apply the weights at the end of the assimilation window. Therefore, the resulting costs $T \times P \times k + \Omega$ are lower than for the ETKF.
- In the **ETKS**, we update the state at the beginning of the assimilation window based on all observations within the window. To get the solution at the end of the window, we have to propagate the ensemble two times, which increases the costs compared to the 4D-ETKF, resulting into $2 \times T \times P \times k + \Omega$.

As a result of linearized trajectories within the assimilation window, the costs for the 4D-ETKF are $(T-1) \times \Omega$ smaller than for the ETKF. The ETKS is roughly two times more expensive than the 4D-ETKF but has the advantage of a non-linearly propagated filtering solution at the end of the assimilation window. To stabilize land surface data assimilation with our smoothing setup, we introduce fingerprint operators for data assimilation across the atmosphere-land interface.

5.2.3 Fingerprint operators

Our fingerprint operators are a form of feature extractor, acting on top of the observations. As a consequence, data assimilation with fingerprint operators are a form of feature-based data assimilation. Therefore, our derived methods are similar to Morzfeld et al., 2018.

In fingerprint operators, we replace the $T \times l$-dimensional observational vector $\mathbf{y}_{1:T}$ by a m-dimensional feature vector $\varphi(\mathbf{y}_{1:T})$, where the feature map ϕ translates from observational space into feature space $\varphi : \mathbb{R}^{T \times l} \mapsto \mathbb{R}^{m}$. To perform data assimilation based on this new feature space, we have also to transform the

observational covariance, which will be denoted as $\tilde{\mathbf{R}}$ for a transformed covariance. The fingerprint operators change thus the likelihood from (5.5), where combined the observation operator and dynamical model into one operator $H_{1:T}(\cdot)$ which maps a state at time 0 to observational states from time 1 to time T,

$$\begin{aligned}\tilde{J}(\mathbf{w}) =&(k-1)\mathbf{w}^T\mathbf{w}+ \\ &+[\varphi(\mathbf{y}^o_{1:T}) - \varphi(H_{1:T}(\bar{\mathbf{x}}^b_0 + \boldsymbol{\delta}\mathbf{X}^b_0\mathbf{w}))]^T\tilde{\mathbf{R}}^{-1}[\varphi(\mathbf{y}^o_{1:T}) - \varphi(H_{1:T}(\bar{\mathbf{x}}^b_0 + \boldsymbol{\delta}\mathbf{X}^b_0\mathbf{w}))].\end{aligned} \tag{5.10}$$

Based on this new cost function, we can derive the equations for a feature-based ETKS, where we linearize the fingerprint operator around the ensemble mean as similarly done in (5.7),

$$\varphi(H_{1:T}(\bar{\mathbf{x}}^b_0 + \boldsymbol{\delta}\mathbf{X}^b_0\mathbf{w}) \approx \overline{\boldsymbol{\varphi}}^b_0 + \boldsymbol{\delta}\boldsymbol{\Phi}^b_0\mathbf{w}. \tag{5.11}$$

The ensemble mean in feature space $\overline{\boldsymbol{\varphi}}^b_0$ and linearized observation operator $\boldsymbol{\delta}\boldsymbol{\Phi}^b_0$ are again constructed based on independently propagated ensemble members. This results into the following solution,

$$(\tilde{\mathbf{P}}^a)^{-1} = (k-1)\mathbf{I} + (\boldsymbol{\delta}\boldsymbol{\Phi}^b_0)^T\tilde{\mathbf{R}}^{-1}\boldsymbol{\delta}\boldsymbol{\Phi}^b_0, \tag{5.12}$$

$$\mathbf{w} = \tilde{\mathbf{P}}^a(\boldsymbol{\delta}\boldsymbol{\Phi}^b_0)^T\tilde{\mathbf{R}}^{-1}(\varphi(\mathbf{y}^o_{1:T}) - \overline{\boldsymbol{\varphi}}^b_0). \tag{5.13}$$

If we compare (5.9) and (5.13), we can see that the form of the ETKS equations remains the same, independent of any feature transformation. This also mean that the computational costs of the fingerprint operators are almost the same as the for ETKS. In the following, we show how the observational covariance in feature space $\tilde{\mathbf{R}}$ can be constructed.

5.2.4 Error covariance for fingerprint operators

The error covariance in feature space should reflect the expected difference between an actual observation in feature space and the unknown truth transformed into feature space. Formally, we define that the error made in feature space is additive and distributed according to an unknown Gaussian distribution around the true state in feature space with an error covariance of $\tilde{\mathbf{R}}$,

$$\varphi(\mathbf{y}^o_t) = \varphi(H(\mathbf{x}_t)) + \epsilon_{\phi,t}, \qquad \epsilon_{\phi,t} \sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{R}}). \tag{5.14}$$

Normally, we do not know the true state $\mathbf{x}_t$, but often we know the error covariance in observational space $\mathbf{R}$ beforehand. Based on this, we can transform the observational covariance from observational space to feature space. We can explicitly estimate the feature covariance for some expressions like linear combinations of features, but often this explicit transformation cannot be used. Thus, we have to approximate the transformation of the observational error covariance. We propose here to use parametric bootstrapping, similar to Morzfeld et al., 2018. If we know the observational covariance and the true observation without any observational

error, then we can draw n observations from this Gaussian distribution and independently transform each observation into feature space $\varphi(H(\mathbf{x}_t) + \epsilon_t^{o(j)})$. In practice, we do not know the true state in observational space. Instead, we can approximate it by the actual observation $H(\mathbf{x}_t) \approx \mathbf{y}_t^o$. The feature covariance is then given as expected squared-error between the mean of the observations $\overline{\boldsymbol{\phi}}^o = n^{-1} \sum_{j=1}^{n} \varphi(H(\mathbf{x}_t) + \epsilon_t^{o(j)})$ and the j-th observation,

$$\begin{aligned}\tilde{\mathbf{R}}_t =& (n-1)^{-1} \sum_{j=1}^{n} (\varphi(H(\mathbf{x}_t)) - \varphi(\mathbf{y}_t^o))(\varphi(H(\mathbf{x}_t)) - \varphi(\mathbf{y}_t^o))^T \qquad (5.15)\\ \approx& (n-1)^{-1} \sum_{j=1}^{n} (\varphi(H(\mathbf{x}_t) + \epsilon_t^{o(j)}) - \overline{\boldsymbol{\phi}}^o)(\varphi(H(\mathbf{x}_t) + \epsilon_t^{o(j)}) - \overline{\boldsymbol{\phi}}^o)^T, \quad \epsilon_t^{o(j)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})\end{aligned}$$

This approximated error covariance converges to the true error covariance in the limit of infinite draws $\lim_{n\to\infty}$ and if the error in observational space is Gaussian distributed. This approximation of the observational covariance in feature space can be used to update the ensemble states with (5.13).

5.3 Experiments

In the following, we elucidate which specific fingerprints for the atmosphere-land interface are used. Afterwards, we explain our experimental setup.

5.3.1 Specific fingerprint operators for land surface data assimilation

In the following, we describe the mean day time temperature and the amplitude of the diurnal cycle as specific fingerprints for data assimilation of 2-metre-temperature observations to update the soil moisture. The 2-metre-temperature is not only shaped by the soil moisture, but also influenced by other factors as the incoming solar radiation, cloudiness, and precipitation (Stull, 1988). We want to filter out these factors and establish fingerprint operators which robustly point towards errors within the soil moisture. These fingerprint operators are related to the diurnal cycle in the 2-metre-temperature within a 24 hour window, because the soil moisture has a major influence on this diurnal cycle, whereas the information content of an instantaneous observations is subject to the coupling strength between boundary layer and land surface.

For our first fingerprint operator, we use the mean daytime temperature. The atmosphere-land interface has its strongest coupling during daytime. At the same time, model errors within the incoming solar radiation especially influence the forecast of instantaneous daytime temperatures. To filter out these variations, we can utilize an averaged quantity over our 24 hour window. We later compare the mean daytime temperature and the median daytime temperature. We see that the mean temperature is higher correlated to the soil moisture than the median daytime temperature. We define here the temperature between 06:00 UTC and 18:00 UTC as daytime temperature.

As second error fingerprint, we use the amplitude of the diurnal cycle fitted with a sine function. The amplitude of the diurnal cycle inthe 2-metre-temperature highly depends on the soil moisture (Idso et al., 1975). Furthermore, the amplitude of the diurnal cycle is a relative fingerprint operator comparing different values within the 2-metre-temperature and thus, robust against model errors which influence the absolute value of the 2-metre-temperature. There are a various number of proxies for the amplitude of diurnal cycle like the standard deviation of the 2-metre-temperature within a 24 hour window or the daily maximum temperature compared to the daily minimum temperature. We find a more robust representation for land surface data assimilation, if we fit a wave function to the 2-metre-temperature measurements and directly use the amplitude of this fitted wave function. We independently fit a wave function to the 2-metre-temperature time series within a 24 hour window and a least-squares approach.

To estimate the covariance of these fingerprints, we use parametric bootstrapping, as explained in Section 5.2.4. For the parametric bootstrapping, we draw 10000 observations from the observational equivalent of the nature run. We additionally utilize prior knowledge. Errors of the amplitude (Breger et al., 1999) and mean daytime temperature directly depend on errors within the 2-metre-temperature. Additionally, we could derive the analytical expression for the covariance between amplitude, mean daytime temperature, and instantaneous 2-metre-temperature, but the covariances related to the amplitude are depending on the phase of the fitted sine wave. Therefore, we use approximated covariances averaged over all 99 observational positions and all simulated days. The approximated covariances for the mean daytime temperature and amplitude are denoted in Table 5.1.

Table 5.1: Estimated fingerprint error covariances based on parametric bootstrapping with the fingerprint equivalent of 10000 generated observations, averaged over 13 days and all 99 observational points.

Name	Variance (K^2)
Mean daytime temperature	0.000833
Sine Amplitude	0.000833
Cross-covariance	0.000412

5.3.2 Experimental description

For our experiments, we use the same idealized twin experiment structure as described in Chapter 2 and Chapter 4. In the following, we therefore only describe shortly the experiments and their goals.

We define a single, deterministic Nature run without data assimilation as our reality. Based on this Nature run, we synthesize 99 observations and an initial ensemble of 40 ensemble members, as described in Chapter 2. We use this initial ensemble to conduct an open-loop run without data assimilation. With this experiment, we want to see the behavior of the generated ensemble members without any data assimilation. This open-loop run is our first baseline experiment; the following data assimilation experiments should have a smaller error to the nature

Table 5.2: Experiments with their accompanied data assimilation schema together with the used data assimilation window, prior multiplicative inflation factor. In additional columns, the used fingerprint operators are indicated and if they are assimilated with a correlated observational error covariance matrix.

Experiment name	Schema	Window	Inf. factor (γ)	Amplitude	Mean	Raw	Correlation
Nature	Ref. run	–	–	✗	✗	✗	✗
Open-loop	No DA	–	–	✗	✗	✗	✗
LETKF	LETKF	Instant.	1.006	✗	✗	✓	✗
4D-LETKF	4D-LETKF	-23– 0 hours	1.18	✗	✗	✓	✗
LETKS (6 h)	LETKS	0– 6 hours	1.05	✗	✗	✓	✗
LETKS (24 h)	LETKS	0–24 hours	1.15	✗	✗	✓	✗
Sine amplitude	LETKS	0–24 hours	1.03	✓	✗	✗	✗
Daytime mean	LETKS	0–24 hours	1.03	✗	✓	✗	✗
Sine+Mean uncorrelated	LETKS	0–24 hours	1.09	✓	✓	✗	✗
Sine+Mean correlated	LETKS	0–24 hours	1.06	✓	✓	✗	✓
Sine+Mean+Raw uncorrelated	LETKS	0–24 hours	1.18	✓	✓	✓	✗
Sine+Mean+Raw correlated	LETKS	0–24 hours	1.15	✓	✓	✓	✓

run to have a positive assimilation impact. Additionally, this open-loop run is used for offline data assimilation experiments (see also Section 3.6 for more information about offline data assimilation experiments) to test different hypotheses regarding the proposed methods and their parameters.

In all of our data assimilation experiments, we use the same horizontal and vertical localization, specified by Gaspari-Cohn covariance functions with a 15 km radius and 0.7 m radius in horizontal and vertical dimensions, respectively. We manually tuned the prior multiplicative inflation for all experiments based on their performance in the first day from 2015-07-31 12:00 UTC to 2015-08-01 12:00 UTC, the tuned inflation factor is shown as additional column in Table 5.2.

As baseline experiment, we assimilate instantaneous 2-metre-temperature observations hourly into the soil moisture with a LETKF (LETKF experiment). We compare this LETKF with a four-dimensional-LETKF (4D-LETKF). There, we assimilate observations once every 24 hours into the soil moisture and use all observations within the previous 24 hours. In this four-dimensional filter, the estimated ensemble weights are applied at the end of the assimilation window.

In the two LETKS experiments, we assimilate observations ahead of the update time and vary the length of the assimilation window and update cycle. In the LETKS (24 h) experiment, we update the soil moisture once a day with observations in a 24 hour window ahead of the update time. This experiment allows us to infer the impact of smoothing trajectories instead of filtering states at single points in time. In the LETKS (6 h) experiment, the soil moisture is corrected every six hours with observations in a 6 hour window ahead of the update time. This shortened LETKS experiment can be seen as compromise between a LETKF and a LETKS with daily updates. These two smoothers are similar to the operationally used methods at the ECMWF ("IFS Documentation CY47R1 - Part II: Data Assimilation" 2020). We replace the simplified extended Kalman filter by a fully localized ensemble Kalman smoother with hourly observations. Our LETKS therefore resembles more ensemble data assimilation methods for the atmosphere.

In all experiments with fingerprint operators, we assimilate features from observations ahead of the update time within a 24 hour window. In addition, we update the soil moisture once a day as in the LETKS (24 h) experiment. To see the impact of single fingerprint operators on the data assimilation, we use the sine amplitude and mean daytime temperature independently in two experiments. We combine the fingerprint operators and assimilate them together in the Sine+Mean experiments. Here, we differentiate between the use of only diagonal observational error covariances and the full covariance matrix with cross-covariances. In the two Sine+Mean+Raw experiments, we assimilate the raw 2-metre-temperature together with the two fingerprint operators to check if we gain more information about the soil moisture by using additionally the fingerprint operators. As in the Sine+Mean experiments, we conduct two different experiments, either with correlated or uncorrelated observational errors.

In contrast to Chapter 4, our experiments were conducted at the super-computing facilities of the DKRZ (German Climate Computing Center) in Hamburg. In preliminary experiments, we saw that there are differences if we either perform the experiments in Hamburg or in Juelich, as in Chapter 4. We therefore decided to rerun the nature run, the open-loop experiment, and the ETKF experiment. As a consequence, their results are slightly different compared to these in Chapter 4.

5.4 Results

This results section has two different parts. In the first part, we analyze the performance of our ensemble Kalman smoother compared to ensemble Kalman filtering for cross-compartmental data assimilation. In the second part, we investigate the increased stability of using fingerprint operators for land surface data assimilation.

5.4.1 Smoothing

As first step, we compare the ensemble Kalman smoother experiments with the ensemble Kalman filter experiment and open-loop run. For this comparison, we use the area-averaged root-mean-squared error of the experiments compared to the nature run (Figure 5.2).

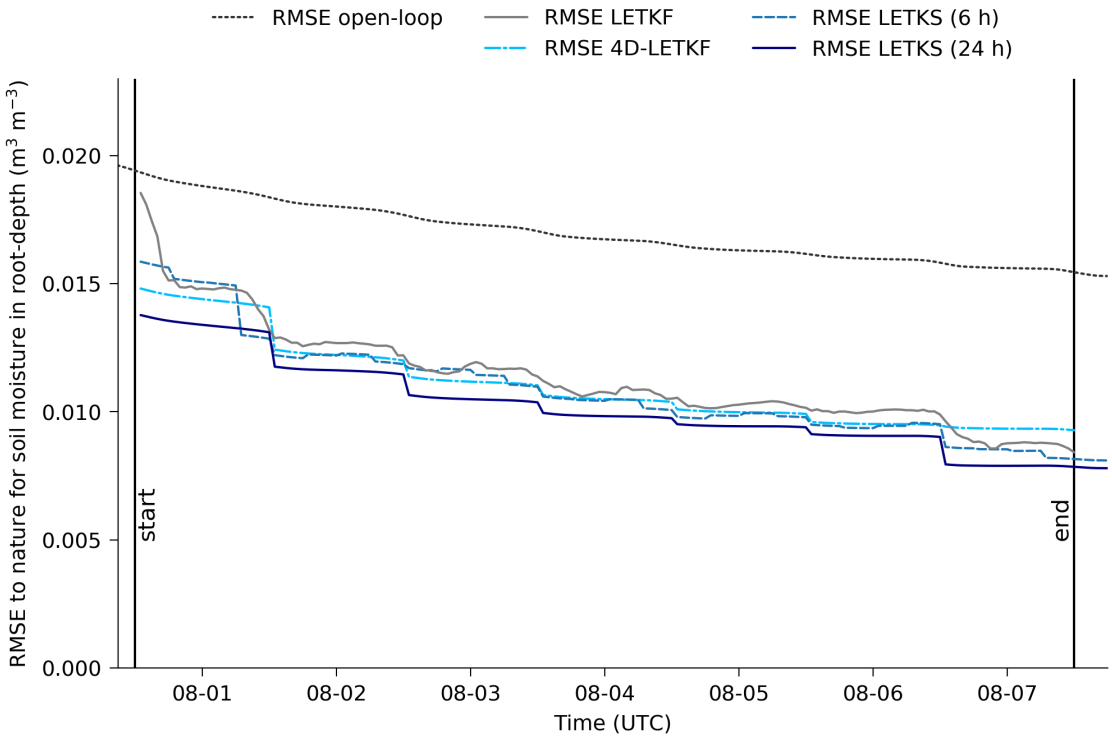


Figure 5.2: Root-mean-squared-error of the smoothing experiments compared to the nature within the simulation window as area average. The light-black dotted line and grey solid line are the baseline experiments with the open-loop run and the LETKF run, respectively. The dashed dotted light-blue line represents the 4D-LETKF experiment with a 24 hour assimilation window, whereas the dashed blue line and solid blue line show the LETKS with a 6 hour and 24 hour assimilation window, respectively.

Data assimilation of the 2-metre-temperature into the soil moisture decreases the analysis error in all experiments compared to the open-loop run (see also Table 5.3). Hence, updating the soil moisture based on the 2-metre-temperature has a positive assimilation impact. Among all data assimilation experiments, the LETKF experiment has the highest RMSE. Hence, ensemble Kalman smoothing reduces the analysis error in the soil moisture.

Table 5.3: Temporal and area-averaged root-mean-squared of the LEKTF and LETKS experiments to the nature run for hourly data from 2015-07-31 13:00 UTC to 2015-08-07 12:00 UTC .

Experiment	RMSE ($m^3\,m^{-3}$)
Open-loop	0.0170
LETKF	0.0115
4D-LETKF	0.0111
LETKS (6 h)	0.0111
LETKS (24 h)	0.0104

By hourly updating, the LETKF experiment non-linearly propagates assimilation increments over time, whereas the 4D-LETKF experiment linearly smooths the increments within a 24 hour window. Despite this linear assumption, data assimilation with the 4D-LETKF has a small positive impact compared to the LETKF. Therefore, the linear assumption has only a negligible impact on the analysis result. The LETKF updates its trajectory based on instantaneous observations, whereas the 4D-LETKF has to wait for 24 hours, before it sees again observations from the previous assimilation window, leading to some inertia. The effect of this inertia can be seen at 2015-08-07, where the 4D-LETKF has the highest error among all data assimilation experiments. At this data, observations have a high assimilation impact on the soil moisture, because of an increased coupling between atmospheric boundary layer and land surface. The LETKF and LETKS already start to incorporate these observations, whereas the 4D-LETKF has to wait to 2015-08-07 12:00 to make use of these observations. As a consequence, The 4D-LETKF has slower response times than the LETKS.

In both LETKS experiments, the LETKS (6 h) experiment and the LETKS (24 h) experiment, we use observations ahead of the analysis time. The increased assimilation window of the LETKF (24 h) experiment results in a decreased analysis error. It seems that within a 24 hour window, the non-linearities in the atmospheric boundary layer have almost no negative impact on the land surface data assimilation. This again confirms that data assimilation of 2-metre-temperature observations from a 24 hour window is almost a linear problem.

In contrast to the 4D-LETKF experiment, the LETKS (24 h) experiment uses observations ahead of the analysis time within the 24 hour window. Using observations ahead of the analysis time additionally reduces the analysis error. As a result, the LETKS (24 h) experiment has the lowest analysis error among all experiments. We can therefore take advantage of temporal dependencies between atmospheric boundary layer and land surface with an ensemble Kalman smoother and a 24 hour assimilation window.

To compared the LETKF and LETKS in the following, we show the increments of the LETKF and LETKS (24 h) experiment compared to the open-loop run in Fig. 5.3 (a & b). We analyze the difference of the prior ensemble mean in the LETKF and LETKS (24 h) experiment to the open-loop run at the end of the simulation time at 2015-08-07 12:00, where the same number of observations were assimilated in both experiments.

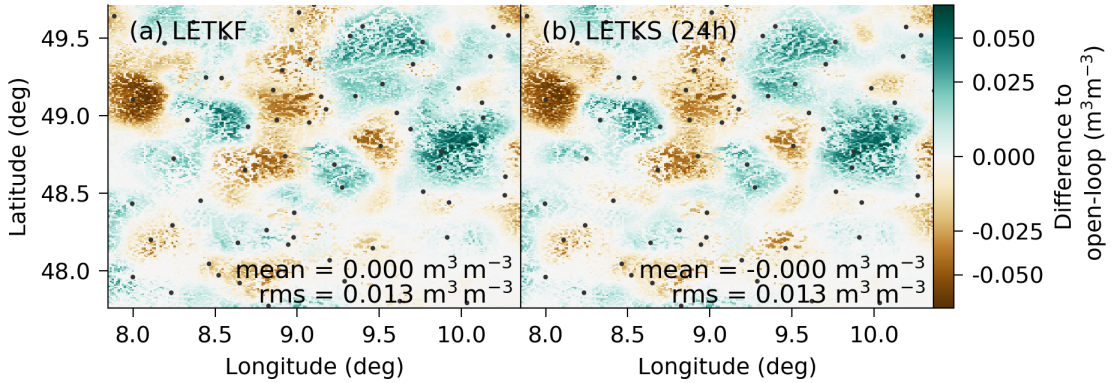


Figure 5.3: Spatial difference in soil moisture of the mean state at 2015-08-07 12:00 UTC compared to the open-loop run for a) the LETKF experiment and b) the LETKS (24 h) experiment.

We find that both data assimilation method have a similar assimilation impact on the soil moisture with a root-mean-squared increment of $\sim 0.013\,\mathrm{m}^3\,\mathrm{m}^{-3}$. In general, both experiments exhibit the same spatial increment structures with only few deviations within the experiments. The assimilation impact is mainly driven by the coupling between the atmospheric boundary layer and the land surface and the position of the 2-metre-temperature observations. In this setting, it might be advantageous to increase the number of observations.

The LETKS (24 h) experiment has nevertheless a lower error than the LETKF experiment. In the following, we analyze why our ensemble Kalman smoother has an advantage compared to the ensemble Kalman filter by showing the area-averaged Kalman gain (Fig. 5.4). This averaged Kalman gain is estimated based on offline data assimilation experiments with a LETKF, where we assimilate 2-metre-temperature observations at the shown observational time into the soil moisture. To see differences in the temporal dependencies, we conduct three different experiments. In these experiments, we shift the soil moisture state by $+24$ hours, by 0 hours, and by -24 hours.

The Kalman gain is mainly driven by the coupling strength between land surface and atmospheric boundary layer. Because the coupling strength has a diurnal cycle, as shown in Section 4.4, also the Kalman gain exhibits a diurnal cycle with its peaked minimum before noon. Around the same time, we find the largest differences in the mean gain between different time shifts of the soil moisture field compared to the 2-metre-temperature field.

Assimilating the 2-metre-temperature into the soil moisture 24 hours before the

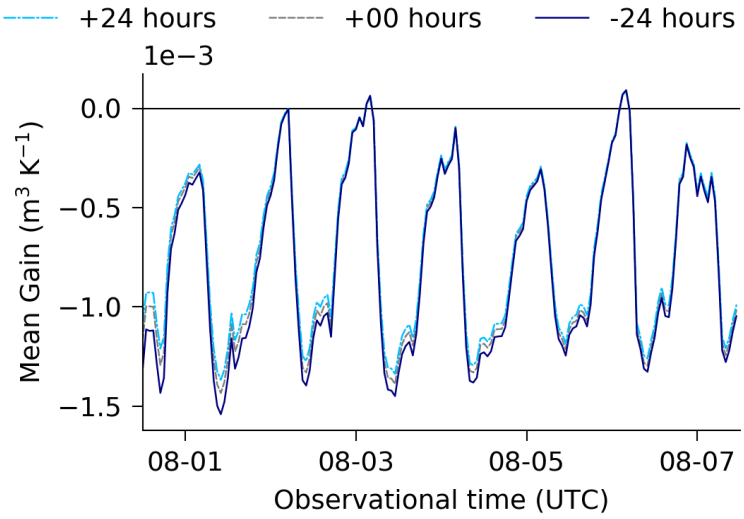


Figure 5.4: The area-averaged Kalman gain from the 2-metre-temperature to the soil moisture for three different shifts in the soil moisture time compared to the shown observational times.

observational time increases the amplitude of the Kalman gain compared to the instantaneous assimilation of the field. On this basis, we would also expect an increased assimilation impact for assimilating observations that are 24 hours ahead of the update time. In contrast, using observations to update the soil moisture 24 hours after the observational time decreases the Kalman gain amplitude. They would hence also decrease the assimilation impact. Because our ensemble Kalman smoother updates the soil moisture based on observations ahead of the update time, we would expect that the Kalman gain of the LETKS (24 h) experiment is increased compared to the LETKF experiment.

Table 5.4: Comparison between the LETKF, the 4D-LETKF, and the LETKS as table for the averaged gains. The gain is averaged over all grid, observational, and time points. Both are averaged over all 168 hours between 2015-07-31 12:00 UTC and 2015-08-07 11:00 UTC.

Experiment	Gain ($m^3\,K^{-1}$)
LETKF	-0.00075
4D-LETKF	-0.00071
LETKS (24 h)	-0.00077

We compare the averaged gain between the LETKF, the 4D-LETKF, and the LETKS (24 h) experiment in Table 5.4. Again, we conduct offline experiments based on the open-loop run, because the data assimilation would otherwise influence the state trajectory so that the gains would be incomparable. The amplitude of the averaged gain in the LETKS (24 h) experiment is increased compared to the LETKF experiment, whereas the gain of the LETKF is larger than for the 4D-LETKF. Hence, smoothing increases the Kalman gain by taking temporal dependencies into account. This explains why the error of the LETKS (24 h) experiment is reduced compared to the other data assimilation experiments (Table 5.3). Furthermore, these results prove again that an assimilation window ahead of the update time improves the soil moisture analysis. Nevertheless, the 4D-LEKTF

has a slightly decreased error compared to the LETKF in online data assimilation experiments (Table 5.3). We do not know if this is by chance, or if there is another cause for this phenomena, which might be related to updating the trajectory only once per day in the 4D-LETKF.

At initialization of the experiments, 2015-07-31 12:00 UTC, all data assimilation experiments have the same initial soil moisture forecast and we would expect the highest assimilation impact. For this single time, we investigate the effect of varying the length of the assimilation window. Specifically, we compare the resulting area-averaged ensemble spread and RMSE of the ensemble mean to the nature run for various assimilation window lengths (Fig. 5.5). On the one hand, an increased window length should result into a decreased analysis error, because we assimilate more observations at once. On the other hand, the relationship between soil moisture and 2-metre-temperature gets more non-linear with an increasing window length, which might have a negative assimilation impact on longer windows.

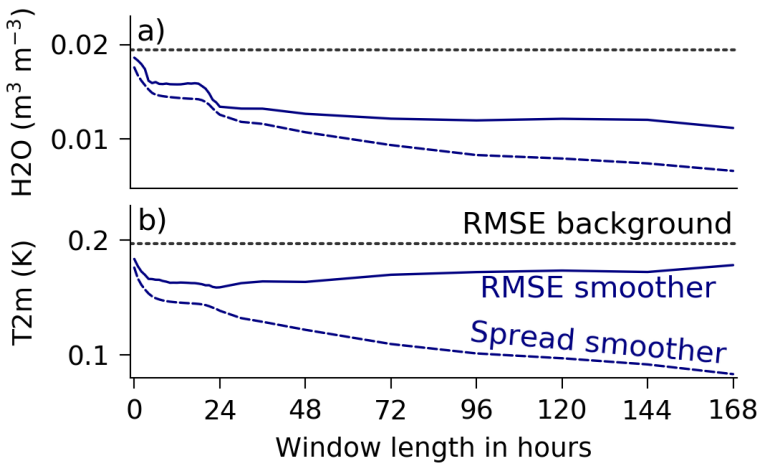


Figure 5.5: Root-mean-squared error as area average at 2015-07-31 12:00 UTC and various assimilation windows compared to the nature run at the same time for a) the soil moisture in root-depth and b) the 2-metre-temperature. The black dotted line represents the static background given by the open-loop ensemble at 2015-07-31 12:00 UTC. The analysis error and spread is shown in dark-blue and light-blue, respectively. The analyses are estimated without any multiplicative inflation.

As we increase the assimilation window , we decrease the analysis error for the soil moisture (Fig. 5.5, a). As a consequence, the analysis with longest assimilation window of 168 hours has the lowest RMSE. Nevertheless, after an assimilation window of 24 hours, the improvement by increasing the assimilation window is small compared to the improvements beforehand. Additionally, the discrepancy between analysis error and analysis spread increases with increasing assimilation window, showing that we would need larger inflation factors for longer assimilation windows. This can be also observed for our online data assimilation experiments, where we needed an increased inflation factor for longer assimilation windows. The inflation factor encounters the effect of non-linearities and, hence,

non-Gaussianities on the analysis. Based on these results, we conclude that an assimilation window of 24 hours seems to be appropriate, if we additionally take computational expenses into account.

If we compare the behavior of the RMSE for the 2-metre-temperature (Fig. 5.5, b) to the soil moisture, we can observe possible problems with strongly-coupled data assimilation for ensemble Kalman smoothers. Whereas the error for the soil moisture decreases with increasing window length, the error for the 2-metre-temperature increases for longer assimilation windows than 24 hours. This is related to the non-linear development of the atmospheric boundary layer. The relationship between 2-metre-temperature and soil moisture seems to have another intrinsic time-scale than the autocorrelative relationship in the 2-metre-temperature. As a result, ensemble Kalman smoothers are more difficult to tune for strongly-coupled data assimilation across the atmosphere-land interface than ensemble Kalman filters.

5.4.2 Fingerprint operators

Hereafter, we step-wise establish that fingerprint operators stabilize the data assimilation of 2-metre-temperature observations for the soil moisture. We start with a feature screening of some possible fingerprint operators based on the ensemble in the open-loop run. For the feature screening, we analyze how much the variance in the soil moisture would be reduced, if we would assimilate 2-metre-temperature observations with the given fingerprint operator into the soil moisture,

$$\mathbf{K}_t \mathbf{H}_t \mathbf{P}_t^b. \tag{5.16}$$

To calculate the Kalman gain $\mathbf{K}_t$, we estimate the observational standard deviation in feature space for every fingerprint operator independently with parametric bootstrapping, as described in Section 5.2.4. The estimated standard deviations are denoted in Table 5.5.

Table 5.5: The estimated feature standard deviations based on 1000 drawn observations from the nature run.

Name	σ (K)
Raw observations	0.100
Daytime-mean	0.029
Daytime-median	0.068
Maximum	0.095
Max-Min-difference	0.132
Standard-deviation	0.021
Sine-amplitude	0.029

To simplify the analysis, we bilinearly interpolate the 2-metre-temperature field from COSMO to the CLM grid, as also done in OASIS3. As an additional simplification, we estimate the variance reduction for every grid point independently. We analyze the grid-point-based variance reduction of different fingerprint operators as temporal and spatial average in Table 5.6. In correspondence to the Kalman filer

and Kalman smoother, we call the instantaneous use of 2-metre-temperature observations *filtering* operator and the use of observations within a 24 hour assimilation window ahead of the update time *smoothing* operator.

The filtering operator has a variance reduction of 0.40, whereas the smoothing operator reduces variances in the soil moisture by 0.78. Hence, the smoothing operator expects that the error in the soil moisture is in average reduced in every assimilation step by 78 % and overestimates very likely the assimilation impact of 2-metre-temperature observations. This overestimation can be seen of some sort of overfitting and is normally counteracted by covariance inflation.

Table 5.6: Variance reduction in soil moisture for grid-point-based covariances from different fingerprint operators to the soil moisture in root-depth averaged over all grid points and times between 2015-07-31 12:00 UTC and 2015-08-07 11:00 UTC in the open-loop run. The amplitude and daytime mean temperature are used in our fingerprint operator experiments. For the filtering operator, we averaged the covariances over the observations at 12:00 UTC. The smoothing operator is the only observational feature where we use more than one observation.

Operator	Variance reduction $(\mathrm{m}^3\,\mathrm{m}^{-3})^2$
Filtering	0.000126
Smoothing	0.000243
Daytime-mean	**0.000207**
Daytime-median	0.000157
Maximum	0.000146
Sine-amplitude	**0.000166**
Standard-deviation	0.000171
Max-Min-difference	0.000116

All fingerprint operators act on the 24 raw observations that are also used in the smoothing operator. They combine these raw observations into a single pseudo-observation. This pseudo-observation is then assimilated instead of the 24 raw observations. Almost all fingerprint operators, except the difference between the maximum and minimum temperature, have an increase variance reduction compared to the filtering operator, where also a single observation is assimilated. The fingerprint operators can therefore condense information from the 24 raw observations into a single pseudo-observations.

Among all fingerprint operators the daytime-mean temperature has the highest variance reduction. This variance reduction is higher than for the daytime-median temperature, which is robust to outliers in the temperature values. Hence, outliers in the 2-metre-temperature might contain information about the soil moisture. To see the effect of outliers, we additionally screen the maximum temperature, which has a slightly increased variance reduction compared to the filtering operator. The maximum temperature very likely influences also the daytime-mean temperature, and we expect that these two fingerprint operators have overlapping information. Thus, we stick to the daytime-mean temperature as our first fingerprint operator.

The amplitude of the diurnal cycle represents differences in the 2-metre-temperature, whereas absolute values are used in the daytime-mean temperature. Hence, we

would expect that the amplitude of the diurnal cycle has additional information about the soil moisture. The difference in the maximum temperature to the minimum temperature acts as proxy of the amplitude but is highly influenced by other processes than the soil moisture, because only two temperature values affect this proxy. These external influences constrain the variance reduction to smaller values than for the filtering operator. The standard deviation of the 2-metre-temperature is a stabilized proxy of the amplitude, because all 24 observations influence the standard deviation. Interestingly, this stabilized proxy is the second moment of the 2-metre-temperature observations. As such, it is only a statistical measure and lacks a physical representation. In contrast, we can fit a sine wave to the diurnal cycle inthe 2-metre-temperature and use the amplitude of this fitted sine wave as direct proxy. This direct proxy is based on physical considerations and has almost the same variance reduction as the standard-deviation operator. We therefore use the sine-amplitude as second fingerprint operator beside the daytime-mean temperature.

In addition to our feature screening, we analyze the impact of the fingerprint operators on the innovations. We use here the normalized innovations in feature space and define them as difference between observation and ensemble mean, normalized by the observational standard deviation in feature space,

$$\Delta\tilde{\boldsymbol{\phi}}^{o} = \tilde{\mathbf{R}}^{-\frac{1}{2}}[\varphi(\mathbf{y}_t^o) - \overline{\boldsymbol{\phi}}_t^b]. \tag{5.17}$$

These normalized innovations are also used within the LETKS to build the product of ensemble perturbations to observational perturbations and tell us something about the impact of the observations on the data assimilation.

Table 5.7: The normalized root-mean-squared innovations over all observational positions between 2015-07-31 12:00 and 2015-08-08 11:00. The innovations are normalized by the fingerprint operator standard deviations from Table 5.5. The amplitude and daytime mean temperature are used in our fingerprint operator experiments.

Name	Innovation (K)
Raw observations	1.940
Daytime-mean	**4.387**
Daytime-median	2.399
Maximum	2.499
Max-Min-difference	1.984
Standard-deviation	3.224
Sine-amplitude	**3.124**

All fingerprint operators have higher root-mean-squared innovations than the raw observations. The normalized innovations of the daytime-mean operator are 2.2 times as large as for the raw observations, showing that observations from the daytime-mean operator have a larger impact. The observational errors of the daytime-mean operator are hereby also smaller than for the daytime-median operator and maximum-temperature operator, which explains their lower normalized innovation values. In terms of observational errors and normalized

innovations, the standard-deviation and sine-amplitude operator dominate the max-min-difference operator with 1.6 as large innovations. These normalized innovations show again that the daytime-mean operator and the sine-amplitude operator are suitable as fingerprint operators.

In the following, we present spatial maps for the variance reduction (5.16) by considering the assimilation of a single observational point at 2015-07-31 12:00 UTC (Fig. 5.6); we use hereby the nearest COSMO and CLM grid point to the selected observational site to estimate the Kalman gain. The variance reduction is normalized by the background variance for every grid point independently to make the values dimensionless. These spatial maps reveal correlation patterns that would be normally suppressed by localization.

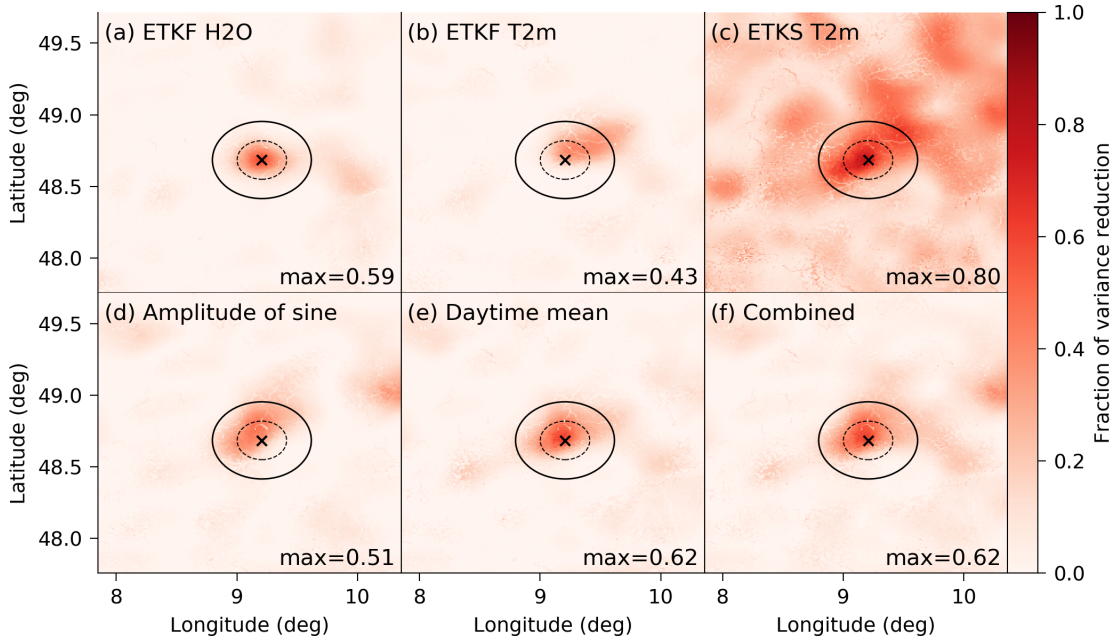


Figure 5.6: Variance reduction from the black crossed point to the soil moisture field at 2015-07-31 12:00 with raw observations (a–c) and fingerprint operators (d–f). Shown are a) a single soil moisture point (observational uncertainty $\sigma^o = 0.01\ \mathrm{m}^3\ \mathrm{m}^{-3}$ as in Fig. 4.4), b) a single 2-metre-temperature point, c) 2-metre-temperature points within the 24 hour window ahead of the soil moisture state (2015-07-31 12:00 – 2015-08-01 11:00), d) the fitted sine amplitude in the same window, e) the daytime mean temperature in the same window, and f) the combined sine amplitude and daytime mean temperature. The ellipsoids symbolize the single and double localization radius, after which an observation shown with the black cross has no influence on grid points.

Assimilating an soil moisture observations from the observational point to the soil moisture field (Fig. 5.6, a) results in an elliptic-shaped pattern of increased variance reduction around the observational point. The pattern corresponds to the selected localization radius in the LETKF, as shown by the black circles, and is shaped by initial soil moisture perturbations in the ensemble. This elliptic pattern is shifted and smeared out to the north-eastern part of the observational site for the ETKF with a single 2-metre-temperature observation at 2015-07-31 12:00 (Fig. 5.6, b). The shifted and smeared out pattern is very likely a result of advected information.

Some of this advected information would be lost in our data assimilation, because of our chosen localization radius of 15 km. For this advected information, the localization radius is too small and constrains the covariance too much (see also Fig. 4.8 for a discussion of this problem).

The assimilation of 2-metre-temperature observations with an ensemble Kalman smoother in a 24 hour window would lead to spurious correlations (Fig. 5.6, c), because of the previously-discussed overfitting problem. This spurious correlation is not physically explainable, and we would need to employ heavy localization. In contrast, the fingerprint operators (Fig. 5.6, d–f) reduce the problem of spurious correlations. The pattern of the combined fingerprint operator (Fig. 5.6, f) within the localization radius hereby resembles the pattern of the smoothing operator, whereas the maximum variance reduction is decreased by 0.18. This shows the efficiency of the fingerprint operators to condense the information within the assimilation window and to reduce the overfitting problem of the smoothing operator.

We show two specific examples of how the daytime-mean temperature and sine-amplitude operators stabilize land surface data assimilation (Fig. 5.7). For this, we use offline experiments (see also Section 3.6 for more information) in the open-loop run, one for the LETKS with a 24 hour window, and one for the combined fingerprint operators. We combine these fingerprint operators neglecting correlations in their errors and using a diagonal error covariance matrix. As measure, we estimate the root-mean-squared error of these experiments to the nature run for the soil moisture in root-depth.

The localization radius shapes and constrains the background covariance matrix $\mathbf{P}_t^b$. In Fig. 5.7, (a), we vary the localization radius. The $\mathbf{R}$-matrix should represent the true error covariance in the observations. In Fig. 5.7, (b), we multiply this $\mathbf{R}$-matrix by a factor, while we keep the observational error constant. In this way, we analyze the robustness of the data assimilation methods in the specification of the $\mathbf{P}_t^b$-matrix and the $\mathbf{R}$-matrix.

For the localization radius, the minimum RMSE is at wider localization radii than the normally used radius (Fig. 5.7, a). On the one hand, this is a result of the offline experiments where we overestimate the ensemble spread, which leads to an increased stability compared to full experiments. On the other hand, we have chosen the localization radius on the basis of the results for the LETKF experiment, and we kept the radius constant across all data assimilation experiments to increase the comparability between the experiments.

We find that the experiment with the fingerprint operators has its minimum in the RMSE at a 4000 m wider radius than the experiment with the LETKS. In addition, the RMSE for the fingerprint operators is more stable for a wider range of localization radii. This illustrates that the smoothing with fingerprint operators is more stable against miss-specified horizontal covariances in the ensemble than smoothing without fingerprint operators.

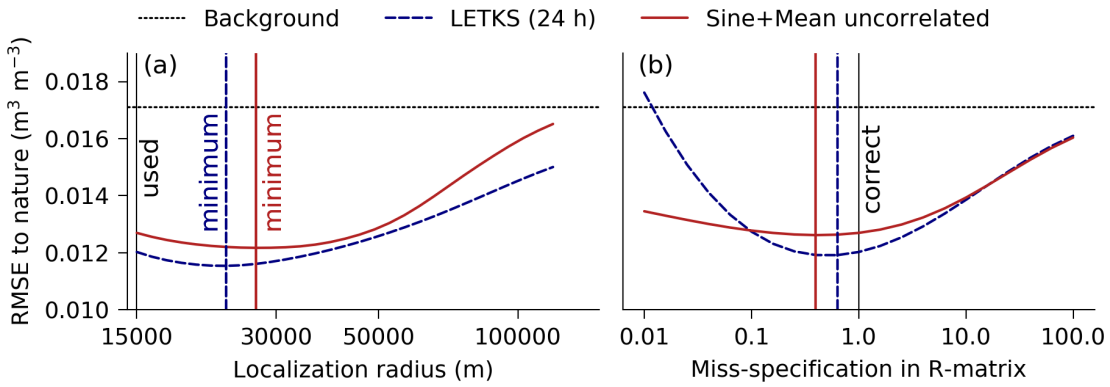


Figure 5.7: The effect of miss-specifications in the covariances on the RMSE of the LETKS and fingerprint operators for offline experiments in the open-loop run with (a) varying the localization radius and b) the miss-specification of the **R**-matrix. The root-mean-squared error is estimated as square-root of the temporally and spatially averaged squared error of the ensemble mean compared to the nature run. In (a), the vertical lines represent the used localization radius in all other experiments, the minimum in the LETKS (24 h) experiment, and the minimum in the Sine+Mean Uncorrelated experiment. In (b), the vertical lines represent a correctly specified magnitude of the **R**-matrix, the minimum in the LETKS (24 h) experiment, and the minimum in the Sine+Mean Uncorrelated experiment.

The amplitude of the **R**-matrix is proportional to the strength of the regularization in the data assimilation. Hence, an increased amplitude corresponds to an increased regularization. This decreases the assimilation impact and increases the RMSE compared to a correctly specified covariance matrix. In contrast, a decreased amplitude corresponds to a decreased regularization. This increases the overfitting to the observations and increases the RMSE compared to a correctly specified covariance matrix.

For all localization radii and correctly specified observational covariances, the LETKS (24 h) experiment has a lower RMSE than the fingerprint operators. The fingerprint operators reduce the information content in favor for an increased robustness against miss-specifications. If the **R**-matrix is more then $10\times$ miss-specified compared to a correctly chosen covariance, then assimilation with the fingerprint operators has a lower error than the direct assimilation of the observations in the LETKS experiment. Therefore, the fingerprint operators make land surface data assimilation more against miss-specification in both, the background covariance matrix and the observational error covariance matrix.

Based on these encouraging results, we perform full data assimilation experiments with the daytime-mean temperature and sine-amplitude operator and compare these experiments with the LETKS (Tab. 5.8 and Fig. 5.8). We use these fingerprint operators independently in two separated experiments. In other experiments, we combine the fingerprint operators. We further differentiate between two observational error covariance settings, where we either assimilate the features with or without their cross-covariances. As last experiment, we assimilate the

features together with the 2-metre-temperature observations.

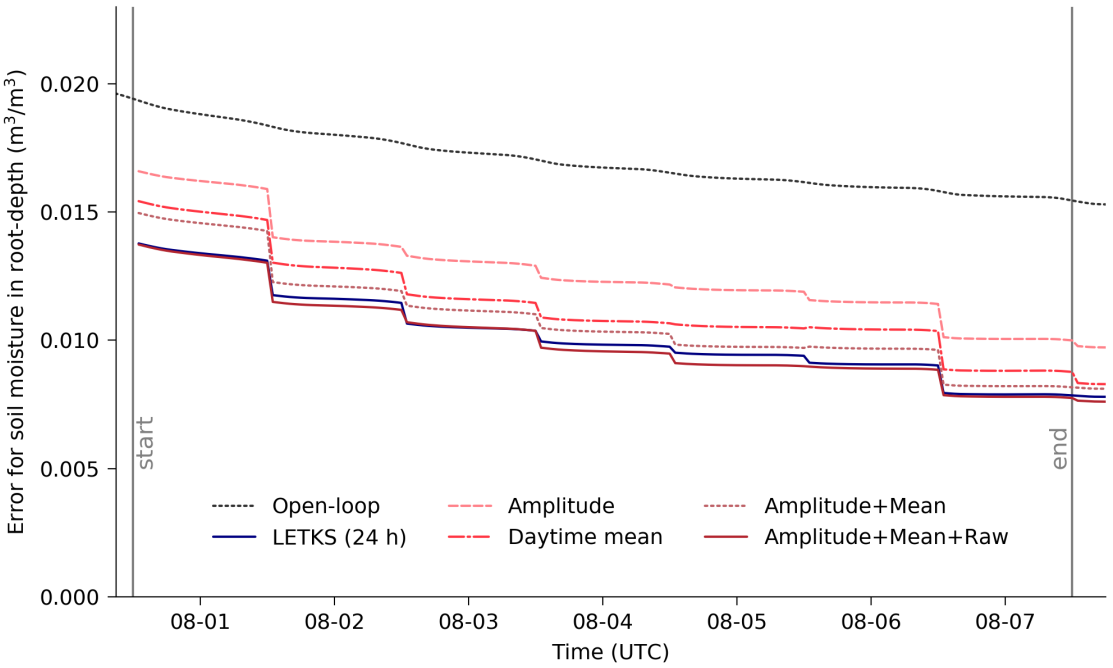


Figure 5.8: The root-mean-squared error of the fingerprint experiments compared to the nature run. The open-loop run and LETKS (24 h) experiment are here shown as baseline experiments. We display here only v the Amplitude+Mean uncorrelated experiment and Amplitude+Mean+Raw uncorrelated experiment, because there is no difference to the correlated experiments in the RMSE (Tab 5.8).

The daytime-mean operator has a higher information content about the soil moisture than the sine-amplitude operator and decreases the error by 9 %. In both experiments with the fingerprint operator and the LETKS experiment, we assimilate observations from a 24 hour assimilation window. The direct use of the observations decreases the assimilation error further by 10 % compared to the assimilation with the daytime-mean operator. Thus, the use of a single fingerprint operator cannot compete with the direct use of the observations in terms of assimilation errors in our case.

By combining the fingerprint operators, we also combine their information content, and we improve the soil moisture analysis by 5 % compared to the experiment with the daytime-mean operator only. We can also combine the observations with the fingerprint operators and assimilate both information together. Together, they decrease the error compared to the LETKS by 2 %. Thus, the fingerprint operators hardly extract more information than the direct use of the 2-metre-temperature observations. But, we can condense the information from the 24 observations into two observational features and retain a similar assimilation impact. In addition, the information extracted by the fingerprint operators is in some way complementary to each other so that they have together a higher assimilation impact.

We find no differences in the RMSE if we either use uncorrelated or correlated

Table 5.8: Temporal and area-averaged root-mean-squared of the experiments to the nature run for hourly data from 2015-07-31 13:00 UTC to 2015-08-07 12:00 UTC.

Experiment	RMSE ($m^3\,m^{-3}$)
Open-loop	0.0170
LETKF	0.0115
LETKS (24 h)	0.0104
Sine amplitude	0.0128
Daytime mean	0.0116
Sine+Mean uncorrelated	0.0110
Sine+Mean correlated	0.0111
Sine+Mean+Raw uncorrelated	0.0102
Sine+Mean+Raw correlated	0.0102

observational error; in theory, the fingerprint operators with correlated errors should give better results. This confirms the result that the combined fingerprint operators make land surface data assimilation more robust against miss-specified observational error covariances, as we have already seen in Fig. 5.7. The fingerprint operators are therefore a possible way to condense information from multiple observations into a few features. At the cost of assimilation impact, this information condensation robustifies land surface data assimilation against miss-specified covariances.

5.5 Discussion and Summary

We investigate how land surface data assimilation can make use of additional information encoded in the temporal development of the 2-metre-temperature. For this, we compare ensemble Kalman filters to ensemble Kalman smoothers in idealized experiments. Our results show that ensemble Kalman smoothers improve the soil moisture analysis based on 2-metre-temperature observations up to 10 % compared to a three-dimensional ensemble Kalman filter with instantaneous observations. We find that this improvement is related to a higher Kalman gain in the smoothing case, and thus related to an enriched representativeness of the 2-metre-temperature observations for the soil moisture. In addition, our analysis of the first analysis step indicates that there is only a small gain in longer assimilation windows than 24 hours.

Our implementation of the ensemble Kalman smoother is based on the Localized Ensemble Transform Kalman filter (LETKF) and equals to the transform variant of the linearized Iterative Ensemble Kalman smoother (Bocquet and Sakov, 2014, 2012); we had no success in using Iterative Ensemble Kalman smoothers with more than one update iterations, as shown in Appendix A.3). As such it belongs to the family of 4DEnVar algorithms, indicating that there might additional gains in idealized experiments by treating cross-compartmental data assimilation as variational problem, as originally proposed by Mahfouf, 1991; Hess, 2001. Our experiments underline that data assimilation of 2-metre-temperature observations into the soil moisture is an almost linear problem in a 24 hour window. As a con-

sequence, we would expect that variational procedures would benefit from more much longer assimilation windows, on a weekly-like time-scale. This would put cross-compartmental data assimilation out of the window needed for numerical weather prediction and move it more into the window for reanalysis and seasonal prediction problems.

In Chapter 4, we show that ensemble Kalman filters for land surface data assimilation benefit from strongly-coupled data assimilation. There, 2-metre-temperature observations are not only assimilated into the soil moisture, but also assimilated into the atmospheric temperature with the same hourly cycle. For ensemble Kalman smoothers, strongly-coupled data assimilation is much more difficult to implement, because the land surface has another intrinsic time-scale than the atmosphere. As a consequence of these different intrinsic time-scales, we would also need different assimilation windows, one for the land surface and one for the atmospheric boundary layer. In addition, ensemble Kalman smoothers with application of the analysis at the beginning of the assimilation window are computationally more demanding than ensemble Kalman filters, because the ensemble has to be propagated a second time through the window. This would especially be a burden in coupled data assimilation with many involved sub-modules.

We find that the ensemble Kalman smoother tends to overfit towards the observations; this causes spurious correlations. Because of these spurious correlations, the ensemble Kalman smoother is much more dependent on correctly specified covariances than the ensemble Kalman filter. In real-world data assimilation, it is much more difficult to tune the localization radius, the inflation factor, and the observational error covariance than in our idealized settings. Therefore, ensemble Kalman smoothers might be too difficult to tune and handle in real-world data assimilation.

As one way forward, we introduce novel fingerprint operators into the land surface data assimilation. Data assimilation with these fingerprint operators are a form of feature-based data assimilation (Morzfeld et al., 2018). As features of 2-metre-temperature observations, we design them to extract characteristic fingerprints that point towards errors in the soil moisture. We show that the daytime-mean-temperature and the amplitude of a fitted sine wave are physically-plausible fingerprint operators to condense the information from 2-metre-temperature observations in a 24 hour window into fewer features. We found that the use of these operators decrease the problem of curious correlations and overfitting to the observations. As a consequence, they make land surface data assimilation more robust against miss-specifications in the localization radius and observational covariance compared to a standard Ensemble Kalman smoother.

We can combine the two fingerprint operators to decrease the error compared to a standard ensemble Kalman filter. Although we assimilate two different features, we have an information loss compared to the raw use of the 2-metre-temperature observations within a 24 hour window, resulting in an increase error of 5 % compared to our ensemble Kalman smoother. Nevertheless, these results

indicate that fingerprint operators are one possible way to increase the stability of ensemble Kalman smoother for land surface data assimilation.

In addition, fingerprint operators open the possibility to use a purely data-driven approach for the discovery of observational features. A generalization of the fingerprint operators would lead us to the so-called kernel and its reproducing kernel Hilbert space (Schölkopf and Smola, 2002). These kernels are a popular trick in machine learning to implicitly embed the observational space into a possibly infinite-dimensional feature space with a specified covariance function. In the case of kernel methods, one could also show that the ensemble transform Kalman filter as dual form definition of an ensemble Kalman filter equals a Gaussian process regression (Rasmussen and Williams, 2006), which is also called Kriging in geostatistics. One of the main problems for this methodology is the high-dimensionality of data assimilation and how to construct kernels specific for data assimilation. Through kernels, we would also introduce other parameter into data assimilation that have to be tuned, which might be tricky in real-world problems. This tuning problem is also evident for localization, which can be seen as one type of specific kernel dealing with high-dimensional data, not acting in data space but in spatial space. Another possibility for a purely data-driven approach would be to use neural networks and deep learning (LeCun et al., 2015; Goodfellow et al., 2016). One could specify the observational features as multi-layered neural network and optimize the variational cost function (5.5) with stochastic gradient descent over a given training dataset. In this construction the ensemble Kalman filter can be seen as last linear layer, performing the regression and implementing a flow-dependency into the network.

All in all, our results support the hypothesis that there is information hidden in the temporal development of the 2-metre-temperature about the soil moisture, which can be extracted by data assimilation. Fingerprint operators can hereby help to decode the information and to stabilize the data assimilation.

5.6 Conclusions

In this study, we investigate how information encoded in the temporal development of the 2-metre-temperature can be used for land surface data assimilation in a limited-area terrestrial system platform and a seven-day period. Based on our results in idealized experiments, we conclude the following:

1. Ensemble Kalman smoothing improves the soil moisture analysis compared to ensemble Kalman filtering. By targeting trajectories instead of single states, we increase the representativeness of 2-metre-temperature observations for the soil moisture.

2. The update of the soil moisture based on 2-metre-temperature observations in a 24 hour window is an almost linear problem.

3. We can apply a LETKS with the analysis weights at the beginning of a 24 hour assimilation window to improve the soil moisture analysis by up to 10 % compared to a 3D-LETKF with hourly updates. Hence, we can take advantage of temporal dependencies between the land surface and the atmospheric boundary layer at the expense of increased computational costs.

4. We find that the daytime mean temperature and amplitude of a fitted sine are fingerprint operators that condense the information content of 2-metre-temperature observations from a 24 hour window into fewer observational features.

5. The fingerprint operators make land surface data assimilation more robust against miss-specifications in the background and observational covariances.

6. We can combine both fingerprint operators time to increase the assimilation impact on the soil moisture analysis by 5 % compared to a 3D-LETKF with hourly updates.

6

Machine learning points of view on the ETKF

In this Chapter, I provide additional theoretical points of view on the ensemble transform Kalman filter (ETKF), beside the standard derivation as done in Chapter 3. These additional points of view are related to machine learning and possibly a way to enable automatic data-driven learning for data assimilation. In the following, I talk more about the principles of these additional points of view.

Let's remind, how I derive the ETKF equations in Chapter 3. Based on Bayes' theorem and an ensemble approximation, I establish a variational cost function with a possibly non-linear observation operator $H_t(\mathbf{x}_t)$. To solve the variational cost function, I translate the ensemble members to observational space. Afterwards, I linearize the observation operator around the ensemble mean to get a sensitivity of the observation operator, as schematically shown in Fig. 6.1, (a). On the basis of this sensitivity, I make one big update step. Although I translate the ensemble with the full observation operator, I estimate an inverse solution only with a linearized version of the operator. This discrepancy between observation operator and its linearized equivalent can lead to a negative assimilation impact on the model-state within the update step. The negative impact can especially occur if assumptions are violated in the data assimilation procedure; the assumptions of the ensemble Kalman filter are violated, if the prior distribution or observational likelihood are non-Gaussian, or if the observation operator is very non-linear. In these cases, the data assimilation problem is very difficult to solve (Bocquet et al., 2010).

An idea is to facilitate the data assimilation with feature-based data assimilation (Morzfeld et al., 2018). In feature-based data assimilation, I transform the observations into a new feature space $\boldsymbol{\phi}_t$ with an additional feature operator or extractor $\varphi_t(\mathbf{y}_t)$, as shown in Fig. 6.1, (b). For feature-based data assimilation, I am flexible in the choice of the feature space. Hence, I can assist data assimilation by selecting the right features for the problem that I am interested in.

For the update step in feature-based data assimilation, I have to find a linearized operator that includes both, the feature operator and the observation operator. In ensemble Kalman filters, I can easily find this linearized operator, because I only need to transform the ensemble members to feature space, and I can linearize the combined operator around the ensemble mean. In such sense, the ensemble approximation makes feature-based data assimilation feasible.

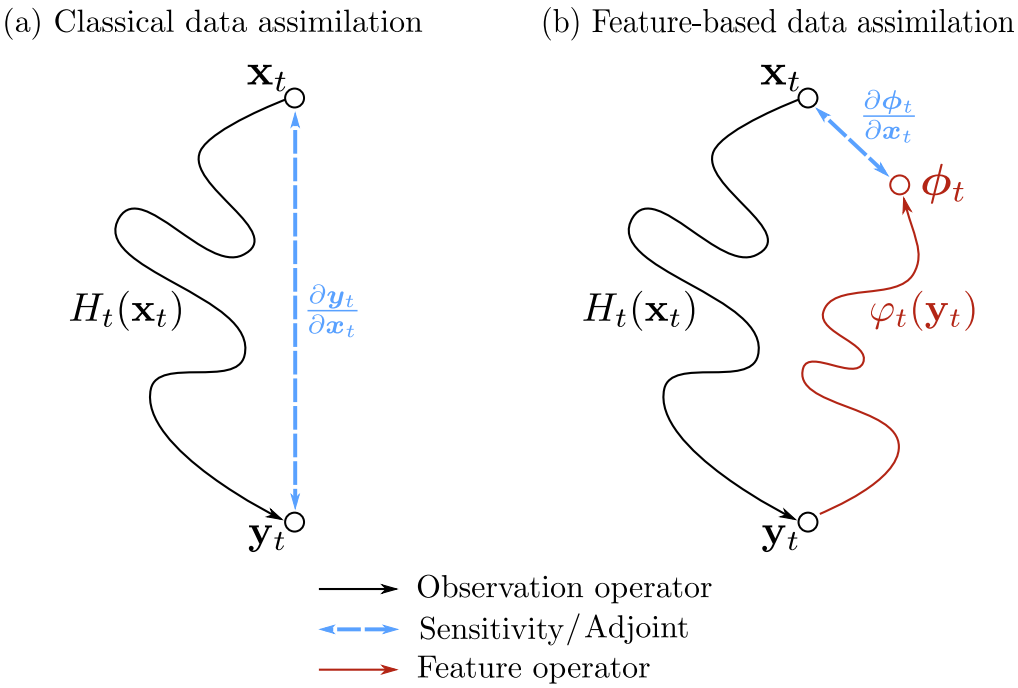


Figure 6.1: Schematic difference between classical data assimilation and feature-based data assimilation. A possibly non-linear and pre-defined observation operator $H_t(\mathbf{x}_t)$ translates a model-state $\mathbf{x}_t$ into an observational equivalent $\mathbf{y}_t$. In (a) classical data assimilation, this operator is linearized around the current model-state $\mathbf{x}_t$ to get the sensitivity $\frac{\partial \mathbf{y}_t}{\partial \mathbf{x}_t}$. In (b) feature-based data assimilation, the observational equivalent is further translated into a feature state $\boldsymbol{\phi}_t$ by another possibly non-linear operator $\varphi_t(\mathbf{y}_t)$. This feature operator is unknown and has to be defined. Then, I only need a sensitivity from feature space into model space $\frac{\partial \boldsymbol{\phi}_t}{\partial \mathbf{x}_t}$, which can be a much easier task in data assimilation.

The feature extractor $\varphi_t(\mathbf{y}_t)$ is still undefined. In Chapter 5, I explicitly define this feature extractor on the basis of physical considerations. There, I take advantage of characteristic fingerprints in 2-metre-temperature observations to correct forecast errors in the soil moisture. Another possibility is to use a Gaussian anamorphosis function in observational space (Bertino et al., 2003; Amezcua and Leeuwen, 2014; Geppert, 2015). In a Gaussian anamorphosis, an explicitly defined feature extractor transforms the observations such that their observational error is more Gaussian distributed.

In the following sections, I show two more principle ways to define this feature extractor. With the kernel trick (Murphy, 2012), I define the feature extractor by an implicitly spanned feature space, as shown in Section 6.1. The kernelized ETKF is related to other methods like Gaussian process regression (Rasmussen and Williams, 2006) or particle filters. As another option, I can also define the feature extractor as neural network, which can be learned by variational Bayes. I introduce variational Bayes as a general way for optimizing parameters of the ETKF together with a small example in Section 6.2.

6.1 A kernel view on feature-based data assimilation

In Chapter 5, I introduce fingerprint operators as features of observations based on a feature extraction function $\varphi_t(\mathbf{y}_t)$. This feature function maps from a l-dimensional observational space to a m-dimensional feature space $\varphi_t : \mathbb{R}^l \mapsto \mathbb{R}^m$. In the following, I show how this introduced feature function is related to kernels in machine learning.

For the fingerprint operators, I derive a variational cost function, working analogous to the linearized cost function for the ETKF. I additionally transform every ensemble member independently into feature space $\boldsymbol{\phi}_t^{b(i)} = \varphi_t(H_t(\mathbf{x}_t^{b(i)}))$. Based on these transformed ensemble members, I create an ensemble mean in feature space $\overline{\boldsymbol{\phi}}_t^b = k^{-1} \sum_{i=0}^{k} \varphi_t(H_t(\mathbf{x}_t^{b(i)}))$ and a column-wise matrix $\delta\boldsymbol{\Phi}_t^b$ of all ensemble perturbations $\delta\boldsymbol{\phi}_t^{b(i)} = \varphi_t(H_t(\mathbf{x}_t^{b(i)})) - \overline{\boldsymbol{\phi}}_t^b$, here for the i-th column. Then, the variational cost function for fingerprint operators (5.10) results into

$$\begin{aligned}\widetilde{\mathcal{L}}(\boldsymbol{w}) =& (k-1)(\boldsymbol{w})^T\boldsymbol{w} \\ &+[\varphi_t(\mathbf{y}_t^o) - \overline{\boldsymbol{\phi}}_t^b - \delta\boldsymbol{\Phi}_t^b\boldsymbol{w}]^T\tilde{\mathbf{R}}^{-1}[\varphi_t(\mathbf{y}_t^o) - \overline{\boldsymbol{\phi}}_t^b - \delta\boldsymbol{\Phi}_t^b\boldsymbol{w}].\end{aligned} \tag{6.1}$$

The feature space term in (6.1) depends on the error covariance in feature space $\widetilde{\mathbf{R}}$. This error covariance is the observational error covariance propagated into feature space. In the following, I incorporate this covariance into the feature function φ_t (Rasmussen and Williams, 2006). Since the covariance is positive definite, I can define a matrix square root of the covariance such that $(\widetilde{\mathbf{R}}^{-\frac{1}{2}})^2 = \widetilde{\mathbf{R}}^{-1}$ holds. By multiplying the result of the original feature function with $\widetilde{\mathbf{R}}^{-\frac{1}{2}}$, I can simply define a new feature function, which incorporates the error covariance. In the following, I drop the dependency on the error covariance and assume that φ_t is a normalized feature function. In addition, I denote $\delta\boldsymbol{\phi}_t^o = \varphi_t(\mathbf{y}_t^o) - \overline{\boldsymbol{\phi}}_t^b$ as the difference between the observations in feature space and the ensemble mean in feature space. With the dropped error covariance and this difference, (6.1) shortens to

$$\widetilde{\mathcal{L}}(\boldsymbol{w}) = (k-1)(\boldsymbol{w})^T\boldsymbol{w} + [\delta\boldsymbol{\phi}_t^o - \delta\boldsymbol{\Phi}_t^b\boldsymbol{w}]^T[\delta\boldsymbol{\phi}_t^o - \delta\boldsymbol{\Phi}_t^b\boldsymbol{w}]. \tag{6.2}$$

This cost function and its solution is solely defined by inner products, either as $(\delta\boldsymbol{\Phi}_t^b)^T\delta\boldsymbol{\Phi}_t^b$, $(\delta\boldsymbol{\Phi}_t^b)^T\delta\boldsymbol{\phi}_t^o$, or $(\delta\boldsymbol{\phi}_t^o)^T\delta\boldsymbol{\phi}_t^o$. I replace these inner products by a positive-definite kernel function $K(\mathbf{y}, \mathbf{y}') = \langle\varphi_t(\mathbf{y}), \varphi_t(\mathbf{y}')\rangle$. This positive-definite kernel function $K : \mathbb{R}^l \times \mathbb{R}^l \longrightarrow \mathbb{R}$ has a corresponding reproducing kernel Hilbert space $\mathcal{H}$ (RKHS, Schölkopf and Smola (2002)); in the following, I will call this positive-definite kernel function simply kernel. As a result, the kernel has a reproducing property such that $f(\mathbf{y}) = \langle f(\cdot), K(\mathbf{y}, \cdot)\rangle_{\mathcal{H}}\ \forall f \in \mathcal{H}$. Especially, this property allows me to replace the feature function by a positive-definite kernel. The kernel is defined for inner products of vectors, whereas the solution of (6.2)

involves also inner products of column-wise matrices. I can replace these inner products by a positive-definite gram matrix $\mathbf{K}$, whose entries are given by $\mathbf{K}^{(i,j)} = K(\mathbf{y}^{(i)}, \mathbf{y}'^{(j)})$. For simplicity, I apply the kernel also on column-wise matrices, e.g. $K(\mathbf{Y}, \mathbf{Y}')$, which then represents the gram matrix. For more information about kernels and their RKHS, I refer to a recent review in Muandet et al. (2017).

Resulting from (6.2), the inner products are centered by the prior ensemble mean in feature space. In contrast, the kernel $K(\mathbf{y}, \mathbf{y}')$ is possibly uncentered wrt. to the ensemble mean in feature space. In this uncentered case, I would not use the difference between observation and ensemble mean to update the state. Thus, I would not recover the ensemble Kalman filter. I can formulate the centering operation solely based on kernels (Schölkopf et al., 1997, 1998); I show a derivation of this centering operation in Appendix A.4. In the following, I simply denote the centered kernel as $\widetilde{K}(\mathbf{y}, \mathbf{y}')$. Neglecting terms that are not dependent on the weight vector $\boldsymbol{w}$, the centered kernel allows me to rephrase (6.2) into

$$\widetilde{\mathcal{L}}(\boldsymbol{w}) \propto (k-1)(\boldsymbol{w})^T\boldsymbol{w} - 2(\boldsymbol{w})^T\widetilde{K}(\mathbf{Y}_t^b, \mathbf{y}_t^o) + (\boldsymbol{w})^T\widetilde{K}(\mathbf{Y}_t^b, \mathbf{Y}_t^b)\boldsymbol{w}. \tag{6.3}$$

As solution to (6.3), I get a kernelized form of the ETKF

$$\boldsymbol{w}^a = [(k-1)\mathbf{I} + \widetilde{K}(\mathbf{Y}_t^b, \mathbf{Y}_t^b)]^{-1}\widetilde{K}(\mathbf{Y}_t^b, \mathbf{y}_t^o), \tag{6.4}$$

$$\tilde{P}^a = [(k-1)\mathbf{I} + \widetilde{K}(\mathbf{Y}_t^b, \mathbf{Y}_t^b)]^{-1}. \tag{6.5}$$

$$\tag{6.6}$$

The inner products in the solution of the ETKF are simply replaced by a kernel function which measures the similarity between the ensemble members in observational space and the observations. In these solutions, the feature space is implicitly spanned by the kernel function $\widetilde{K}(\mathbf{y}, \mathbf{y}')$, instead of explicitly translating the observations into feature space by a feature function $\varphi_t(\mathbf{y}_t)$. Furthermore, I interestingly assimilate absolute values in observational space, instead of normalized differences to the ensemble mean, because the centering operation is defined in feature space. This could be an advantage in cases where the observational error is non-Gaussian distributed. By varying the kernel function, I also vary the feature space. I therefore define the main properties of this kernelized ETKF by the kernel function.

Because of this central importance of the kernel functions, I show some kernel functions in the next section. As example, these kernel functions can recover the original ETKF or an ETKF with an infinite-dimensional feature space. For more examples of kernels, I refer to Rasmussen and Williams (2006) and Murphy (2012).

6.1.1 Kernel functions

The linear kernel is defined as inner product between given two vectors $\mathbf{y}$ and $\mathbf{y}'$. I incorporate the observational error covariance $\mathbf{R}^{-1}$ as modifier of this inner

product,

$$K_{\text{lin}}(\mathbf{y}, \mathbf{y}') = (\mathbf{y})^{\mathsf{T}} \mathbf{R}^{-1} \mathbf{y}'. \tag{6.7}$$

If I plug this kernel together with the centering operation into (6.5) and (6.4), I recover the original ETKF in (3.19) and (3.20). Hence, the ETKF is a linearized operation to update the ensemble by data assimilation.

Another kernel that includes the linear kernel is the polynomial kernel. The polynomial kernel is defined by its polynomial degree p and a constant shifting factor c. Again, I incorporate the observational error covariance into the inner product,

$$K_{\text{poly}}(\mathbf{y}, \mathbf{y}') = ((\mathbf{y})^{\mathsf{T}} \mathbf{R}^{-1} \mathbf{y}' + c)^{p}. \tag{6.8}$$

The polynomial degree and constant shifting factor are additional parameters of this kernel that have to be tuned. The polynomial degree determines the maximum order of products between different observations. Furthermore, the polynomial kernel is one of the kernels, where the feature space can be explicitly build by feature functions with the same order of interactions. If I set the degree to $p = 1$, I recover the linear kernel up to a constant. Similarly to the linear kernel, the kernel is unbounded and has the property that its value increases if the input vectors are dissimilar. The polynomial kernel is additionally non-stationary; its value depends on the absolute value of the input vectors.

In contrast, the Gaussian kernel (also known as radial basis function kernel) is a stationary kernel, defined by the relative difference between the input tensors. The Gaussian kernel resembles a Gaussian function in the following way,

$$K_{\text{gauss}}(\mathbf{y}, \mathbf{y}') = \sigma^2_{\text{gauss}} \exp(-\frac{1}{2}(\mathbf{y} - \mathbf{y}')^{\mathsf{T}} \mathbf{\Gamma} (\mathbf{y} - \mathbf{y}')). \tag{6.9}$$

In this kernel, I parametrize the precision as $\mathbf{\Gamma} = l^{-2}\mathbf{R}^{-1}$ with l as characteristic lengthscale of the kernel, specifying the general smoothness of the kernel, whereas $\mathbf{R}$ non-dimensionalizes the distances. Since the Gaussian kernel is defined by $\mathbf{y} - \mathbf{y}'$, the kernel is translation invariant and bounded by 0 and σ^2_{gauss}, which determines the amplitude of the kernel. Furthermore, the Gaussian kernel specifies an infinite-dimensional feature space and is a characteristic kernel such that all higher moments of the data are captured (Muandet et al., 2017). In theory, this kernel allows me to model any smooth target function, because it is additionally a universal kernel (Micchelli et al., 2006; Sriperumbudur et al., 2011). The Gaussian kernel shows that it is possible to specify an infinite-dimensional feature space with the kernelized ETKF.

The polynomial and Gaussian kernels both specify non-linear interactions between different observations. Therefore, they differ if I apply the kernelized ETKF on all observations simultaneously or independently one after another. In contrast, the linear kernel is independent, because it has an associative property. This associative property is quite important to avoid an overfitting to observations that are independent of each other. For non-linear kernels, I can apply a kernel for

each dimension independently, because the sum of kernels is again a valid kernel. Thus, I can use specific kernels for observations and dimensions, and I can bring structure into my kernels (Duvenaud, 2014) to make use of specific properties of my observations.

The shown reformulation of the ETKF into a kernelized ETKF is only possible because the ETKF is formulated in a dual form of an ensemble Kalman filter. This kernel formulation as theoretical approach allows me to draw similarities of the ensemble Kalman filter to different methods like Gaussian process regression or particle filters.

6.1.2 Relation to other methods

The used kernels in the kernelized ETKF are related to the RKHS. On this basis, I connect the kernelized ETKF to other methods in machine learning and data assimilation.

Instead of specifying the update step of the ETKF from Bayes' theorem (3.6), I can try to fit an inference function $f(\mathbf{y})$. As replacement of the update step, the inference function $f : \mathbb{R}^l \mapsto \mathbb{R}^s$ maps from the l-dimensional observational space to the s-dimensional model-state space. Hence, the inference function predicts based on given observations the model-state. In some sense, the inference function inverts the observation operator and is often also called inverse function. I can condition the inference function on current observations to get one single estimate for the posterior $p(\mathbf{x}_t \mid \mathbf{y}_{1:t})$. Instead of searching for a single best function, I can also specify a distribution of functions $p(f \mid \mathbf{y}_{1:t-1})$, conditioned on previous observations. I can then marginalize over the functions to get the posterior distribution (Murphy, 2012, Chapter 15),

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t}) = \int p(\mathbf{x}_t \mid f, \mathbf{y}_t) p(f \mid \mathbf{y}_{1:t-1}) \mathrm{d}f. \tag{6.10}$$

Based on this posterior, the function should be fitted based on all information from previous observations. In the ensemble Kalman filter, we model the inference function by prior samples in model and observational space (Anderson, 2003; Geppert, 2015). As seen in Chapter 3, the prior samples encode previous observations. They can be therefore used to fit the inference function $f(\mathbf{y})$.

To fit a single best inference function, I consider a regularized least-square loss and a RKHS $\mathcal{H}$ as hypothesis space. This results into the following cost function with the regularization parameter λ,

$$\underset{f \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^{k} \| \delta \mathbf{x}_t^{b(i)} - f(\mathbf{y}_t^{b(i)}) \|^2 + \lambda \|f\|_{\mathcal{H}}^2. \tag{6.11}$$

The regularization $\lambda \|f\|_{\mathcal{H}}^2$ is similar to a Tikhonov regularization in standard data assimilation and smooths here the considered functions. As a consequence of the

representer theorem (Kimeldorf and Wahba, 1970; Schölkopf et al., 2001), the kernel ridge regression (6.11) has a unique solution for $\lambda > 0$. For this unique solution, I can recover the solution of the kernelized ETKF mean (6.4) by setting $\lambda = k - 1$ and using centered kernels, as I show in Appendix A.5. The kernelized ETKF therefore solves in its mean update a regularized kernel least-square loss, mapping from observational space into the space spanned by the ensemble perturbations.

Sætrom and Omre (2011) and Yang (2020) are the only references I am aware of that show a generalization of the ensemble Kalman filter based on kernel ridge regression. In both references, the generalization is solely based on fitting the inference function with kernel ridge regression. In contrast, I derive the kernelized ETKF from feature-based data assimilation and, hence, in a more general way. Because feature-based data assimilation changes the likelihood of the data assimilation (Morzfeld et al., 2018), this also means that I am not solving the original ensemble Kalman filter problem.

The solution of the regularized kernel least-square loss shares the same solution of the mean for Gaussian process regression (Rasmussen and Williams, 2006; Murphy, 2012), and both creates their similarity matrices in data space. Kriging (Cressie, 1993) contrarily acts on spatial distances, but shares otherwise the same solution of Gaussian process regression. Since observational localization in the LETKF can be seen as an additional kernel, using spatial distances between observations and a considered grid point, the LETKF combines information from Gaussian process regression and Kriging. The weighting of the observations is hereby like Kriging, whereas the estimation of the analysis resembles Gaussian process regression.

The (kernelized) ETKF estimates its analysis as linear combination of ensemble member perturbations, which resembles particle filtering. Particle filtering (Doucet et al., 2001) is based on Bayes' theorem and estimates its weights based on the observational likelihood of the ensemble members. Because the weights are afterwards normalized to 1 and the observational likelihood is always positive, the particle filter uses a convex combination of the ensemble member perturbations for the analysis. In contrast, the weights of the kernelized ETKF are often negative and do not necessarily sum up to 1. Nevertheless, the kernelized ETKF with a Gaussian kernel resembles a particle filter with a Gaussian observational likelihood. Contrary to a particle filter, the kernelized ETKF normalizes its weights based on $\mathbf{P}_t^a = [(k-1)\mathbf{I} + \widetilde{K}(\mathbf{Y}_t^b, \mathbf{Y}_t^b)]^{-1}$, which leads to a different behavior in the weights. I exemplify the impact of this normalization in the next subsection.

To make data assimilation more expressive, often the model-state space is transformed with kernels or feature functions, e.g. in Pulido and van Leeuwen, 2019; Luo, 2019; Spantini et al., 2019; Pulido et al., 2019; Stordal et al., 2021. Often the model-state space is higher-dimensional than the observational space. This high-dimensionality can lead to problems with defining the right kernel functions for the problem. It might be therefore advantageous to featurize instead the observational space, as done in the kernelized ETKF.

6.1.3 Wind speed example

I shortly visualize the weights of the kernelized ETKF with a simple example that is inspired by Lorenc (2003). The state vector in this example are the wind direction u and v, $\mathbf{x} = (u, v)$ with a prior distribution of $p(\mathbf{x}) = \mathcal{N}((3,1),(1,1))$, from where I draw $k = 2000$ ensemble members. The observation operator is the wind speed $H(\mathbf{x}) = (\mathbf{u}^2 + \mathbf{v}^2)^{\frac{1}{2}}$ and the observational distribution $p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(1.5, 0.05)$. The observational likelihood and the posterior are non-Gaussian distributed, because the observation operator is non-linear wrt. the model-state space.

For this example, I compare in Fig. 6.2 the linear ETKF, a kernelized ETKF and a simple particle filter (Doucet et al., 2001). The particle filter directly solves the Bayesian inference problem and, hence, is the gold standard in some sense. As kernel in the kernelized ETKF, I utilize the Gaussian kernel (6.9) which has additional parameters in comparison to the linear kernel. For the amplitude factor, I choose $\sigma^2_{\text{gauss}} = n = 1000$, whereas I use the median heuristic (Garreau et al., 2018) for the lengthscale, resulting in $l = 2.93$. Before I evaluate the kernel, I subtract the ensemble mean in observational space such that the kernel acts on observational perturbations.

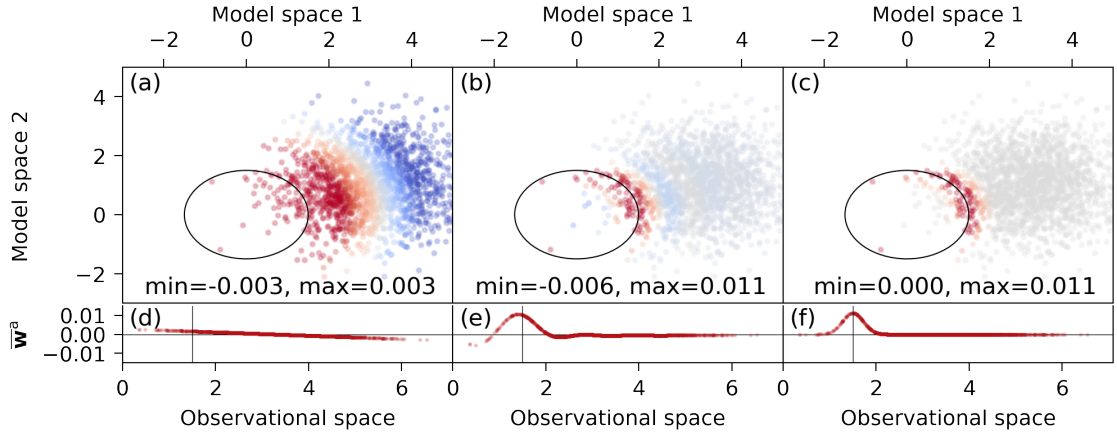


Figure 6.2: The mean ensemble weights in model space (a)–(c), where model space 1 represents the u-wind component and model space 2 the v-wind component, and observational space (d)–(f) for (a) & (d) the linear ETKF, (b) & (e) the kernelized ETKF with a Gaussian kernel, and (c) & (f) the bootstrap particle filter. The dots represent the drawn prior samples either in model space or observational space, whereas the circle and vertical line, respectively, symbolize the observation. The colors in (a)–(c) show the weight for the i-th sample with red colors for positive weights and blue colors for negative weights. Ensemble members with gray colors have almost no weight in the data assimilation.

In model space (Fig. 6.2, (a)–(c)), the mean ensemble weights of the ETKF, kernelized ETKF, and particle filter are all non-linear because of the non-linear observation operator. The kernelized ETKF and particle filter (Fig. 6.2, (b) & (c)) select a few prior members with high ensemble weights. In contrast, the linearity in the ETKF (Fig. 6.2, (a)) causes that whole areas have similar weights. This linear assumption can be especially seen in observational space (Fig. 6.2, (d)–(f)). Here, the ETKF (Fig. 6.2, (d)) fits simply a linear regression so that the ensemble member

with the lowest observational equivalent has the highest weight. In contrast, the weights of the kernelized ETKF (Fig. 6.2, (e)) resembles the Gaussian distribution of the particle filter weights (Fig. 6.2, (f)) such that ensemble members around the observational value have the highest weights. As a consequence of its normalization (Sollich and Williams, 2004; Rasmussen and Williams, 2006), the Gaussian kernel in the kernelized ETKF leads to an oscillatory behavior compared to the weights from the Gaussian observational likelihood in the particle filter.

This example shows potential for the kernelized ETKF. This potential can be especially seen from a theoretical point of view, since the kernelized ETKF a generalization of the fingerprint operators that I have introduced in Chapter 5. These kernels increase the flexibility of the ensemble Kalman filtering toolbox at the cost of additional parameters that have to be tuned within the ensemble Kalman filter; in the case of the Gaussian kernel, the amplitude and the lengthscale. In the next section, I present a possible way to optimize the parameters in the ETKF by variational Bayes.

6.2 Optimizing the ETKF with variational Bayes

In the previously derived kernelized ETKF, I introduced additional parameters into the data assimilation, which have to be optimized in a general way. Here, I show that parameters of the ETKF can be optimized by variational Bayes. I start with a short derivation of variational Bayes (Jordan et al., 1999; Beal, 2003; Hinton and van Camp, 1993) and will logically argument why the ETKF is an optimal solution in the linear-Gaussian case. In addition, I explain how variational Bayes can help us to optimize parameters of the ETKF in the non-linear and non-Gaussian case. As last step, I exemplify the use of variational Bayes in an offline experiment for the atmosphere-land interface, where I optimize the observational error variance based on a fitted inverse gamma distribution. A schematic overview over variational Bayes is shown in Fig. 6.3.

I want to estimate the unknown posterior $p(\mathbf{x}_t \mid \mathbf{y}^o_{1:t})$ in data assimilation, as shown in Section 3.1. Instead of directly optimizing the posterior based on Bayesian principles via maximum-a-posterior as in variational data assimilation, I consider here an approximated posterior distribution $q_\theta(\mathbf{x}_t)$. This distribution has as variational parameters θ, which could represent for example the mean and covariance in model-state space. Ideally, the approximated posterior should be as close as possible to the unknown posterior. I use the reverse Kullback-Leibler divergence (KL-divergence, Kullback and Leibler (1951)) as closeness criterion, where $\mathbb{E}_{q_\theta(\mathbf{x}_t)}$ is the expectation over the approximated posterior,

$$D_{KL}(q_\theta(\mathbf{x}_t) \parallel p(\mathbf{x}_t \mid \mathbf{y}^o_{1:t})) = \mathbb{E}_{q_\theta(\mathbf{x}_t)} \log \frac{q_\theta(\mathbf{x}_t)}{p(\mathbf{x}_t \mid \mathbf{y}^o_{1:t})}. \tag{6.12}$$

The KL-divergence is zero if and only if $q_\theta(\mathbf{x}_t)$ equals $p(\mathbf{x}_t \mid \mathbf{y}^o_{1:t})$ and measures the closeness between two known distributions. Because $p(\mathbf{x}_t \mid \mathbf{y}^o_{1:t})$ is unknown, I have to use a proxy for the criterion. Based on Bayes' theorem, I can reformulate the KL-divergence to get this proxy criterion. As a result, the KL-divergence is

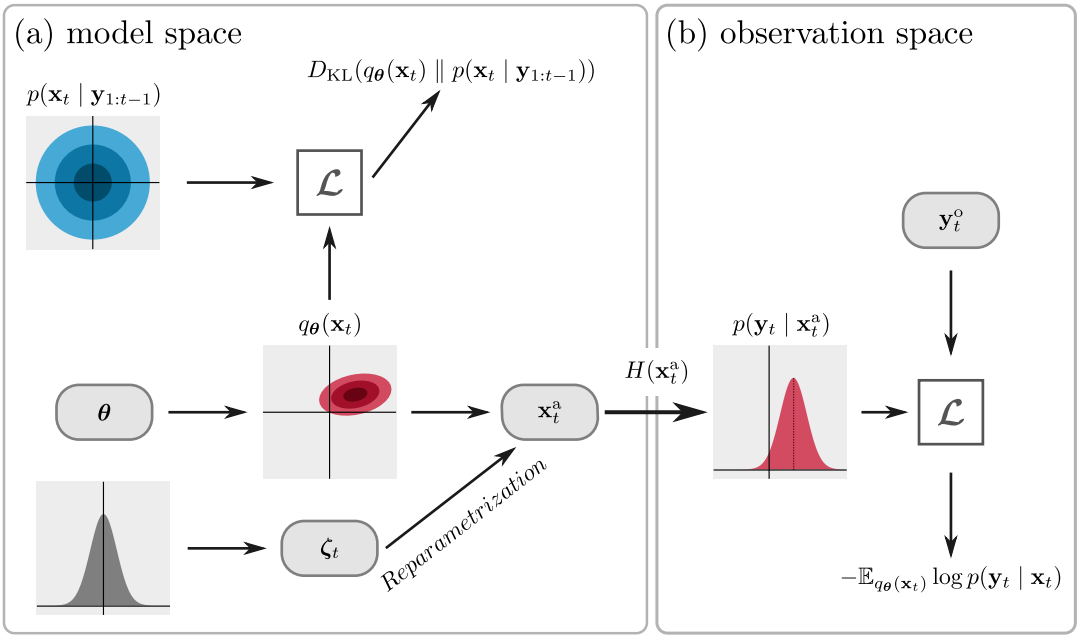


Figure 6.3: A schematic overview over the components of variational Bayes in (a) model space and (b) observational space. In model space, the prior distribution $p(x_t \mid y^o_{1:t-1})$ is compared to the red posterior distribution $q_\theta(x_t)$, generated with its variational parameters θ, by the KL-divergence. The posterior is reparameterized with a drawn random vector ζ_t to a posterior sample x^a_t. This posterior sample is translated with the observation operator $H(x^a_t)$ to the red observational distribution $p(y_t \mid x^a_t)$, conditioned on the posterior sample. The observation operator has a central role and its arrow is therefore thickened. The observation y^o_t is then compared to this conditional distribution, resulting into the negative log-likelihood.

proportional to the variational free energy $J(\theta)$,

$$
\begin{aligned}
(6.12) &= \mathbb{E}_{q_\theta(x_t)} \log \frac{q_\theta(x_t) p(y^o_t \mid y^o_{1:t-1})}{p(y^o_t \mid x_t) p(x_t \mid y^o_{1:t-1})} \\
&= \mathbb{E}_{q_\theta(x_t)} [\log q_\theta(x_t) - \log(p(y^o_t \mid x_t) p(x_t \mid y^o_{1:t-1})) + \mathbf{C}] && (6.13) \\
&\propto -\mathbb{E}_{q_\theta(x_t)} [\log p(y^o_t \mid x_t) + \log q_\theta(x_t) - \log p(x_t \mid y^o_{1:t-1})], && (6.14)
\end{aligned}
$$

$$
J(\theta) = -\mathbb{E}_{q_\theta(x_t)} \log p(y^o_t \mid x_t) + D_{\mathrm{KL}}(q_\theta(x_t) \parallel p(x_t \mid y^o_{1:t-1})). \tag{6.15}
$$

In (6.13), I move $p(y^o_t \mid y^o_{1:t-1})$ into the constant $\mathbf{C}$, because the observational evidence is constant with respect to the parameters θ, which allows me to drop the constant in (6.14). In (6.15), I combine $\mathbb{E}_{q_\theta(x_t)}[\log q_\theta(x_t) - \log p(x_t \mid y^o_{1:t-1})]$ into $D_{\mathrm{KL}}(q_\theta(x_t) \parallel p(x_t \mid y^o_{1:t-1}))$.

The variational free energy $J(\theta)$ in (6.15) depends on the expected negative log-likelihood of the observations $-\mathbb{E}_{q_\theta(x_t)} \log p(y^o_t \mid x_t)$ given the current posterior. This term translates the current posterior into observational space and compares in the observational space the posterior to the observations. As a consequence, this term nudges the posterior to the observations. The second term is the KL-

divergence from the current posterior to the prior $D_{KL}(q_\theta(x_t) \parallel p(x_t \mid y^o_{1:t-1}))$. This term constrain the posterior to stay in the surrounding of the prior. Compared to the maximum-a-posterior derivation in (3.7), I additionally minimize the entropy of the posterior $\mathbb{E}_{q_\theta(x_t)} \log q_\theta(x_t)$ in this term. With the variational free energy, I therefore optimize not only the mode of a distribution but the whole distribution.

Additionally, I lower bound the model evidence (Murphy, 2012) with the negative of the variational free energy $-J(\theta) \leqslant p(y^o_t \mid y^o_{1:t-1})$; the model evidence describes how much information is stored within the posterior. Therefore, the negative of the variational free energy is also called evidence lower bound (ELBO). In the following, I use the term negative evidence lower bound (NELBO) interchangebly for the variational free energy. Thus, minimizing (6.15) do not only optimizes the KL-divergence, but also tightens the evidence lower bound.

To optimize $J(\theta)$ from (6.15), I can simply use Monte-Carlo sampling, draw samples from the current posterior $x_t \sim q_\theta(x_t)$, and use gradient descent. In addition, I can analytically evaluate the KL-divergence between the posterior distribution and prior distribution $p(y^o_t \mid y^o_{1:t-1})$ for certain distributions, like Gaussians, reducing the approximation error for the KL-divergence. Nevertheless, uncertainties caused by the Monte-Carlo sampling of the observational likelihood remain (Gal, 2016, Chapter 3.1). To reduce the uncertainties from the observational likelihood, I can use the reparametrization trick (Kingma and Welling, 2013; Rezende et al., 2014) for the propagation of the signal coming from the observational log-likelihood to the inference. The reparametrization trick further decreases the variance of sampling approximation. Instead of directly sampling from the posterior distribution, I reparametrize the sampling procedure by introducing an auxiliary noise ζ. The randomness of the sampling in the posterior is then moved to this auxiliary noise. I express the sampled random variable x_t as the result of a deterministic function $x_t = g(\theta, \zeta)$, which gets as input the variational parameters θ and the auxiliary noise. For a Gaussian posterior, the variational parameters $\theta = (\bar{x}_t, \boldsymbol{\Sigma}_t)$ are given by the mean $\bar{x}_t$ and the covariance $\boldsymbol{\Sigma}_t$, and the reparametrization could look like the following equation, where $\boldsymbol{\Sigma}_t^{\frac{1}{2}}$ corresponds to the square-root of the covariance,

$$x_t = \bar{x}_t + \boldsymbol{\Sigma}_t^{\frac{1}{2}} \zeta, \qquad \zeta \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{6.16}$$

For certain distributions, I can therefore reduce the approximation error of the Monte-Carlo sampling by using the analytical form of the KL-divergence and the reparametrization trick for the observational likelihood.

In the following, I argue why the ETKF reduces (6.15) for a posterior in weight space. In the ETKF, we optimize the posterior in weight space (3.15) based on a Gaussian assumption on the prior distribution,

$$x_t = \overline{x^b_t} + \delta X^b_t w. \tag{3.15}$$

Thus, it is natural to define the approximated posterior as Gaussian distribution with $\boldsymbol{w}^a$ as mean and $\widetilde{\mathbf{P}}^a$ as covariance, $q_\theta(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}^a, \widetilde{\mathbf{P}}^a)$. It is known that the Kalman filter optimizes the variational free energy (or often called "maximum relative entropy" principle) in the linear-Gaussian case (Mitter and Newton, 2005; Giffin and Urniezius, 2014), whereas also the variational cost function (3.8) can be recovered (Bocquet, 2008) from the free energy. In addition, the ETKF optimizes the linearized variational cost function (3.18), if I linearize the observation operator around the ensemble mean in model space. Therefore, the solution of the ETKF (3.19) and (3.20) is also the maximum relative entropy solution for the approximated posterior. In Appendix A.6, I rigorously prove that the ETKF update equations corresponds to the solution of a full Gaussian posterior distribution in the linearized-Gaussian case. Hence, the ETKF can be used as last layer in an inference chain, giving always the optimal solution in a linearized-Gaussian case.

Since I know that the ETKF optimizes the variational free energy (6.15) in the linearized-Gaussian case and the ETKF is fully differentiable, I can use the variational Bayes to optimize parameters in the ETKF with gradient descent. Compared to expectation maximization (for example used in Pulido et al. (2018)), variational Bayes does not discriminate between state variables and parameters. Hence, I can optimize the full distribution of a parameter and include priors to constrain their solution. In addition, I can optimize the ETKF and parameters in the same update step based on the same cost function. In ensemble Kalman filtering, an approach for parameter estimation is to simply augment the model-state space by the parameters such that they are updated at the same time as the model-state. Therefore, the variational Bayes approach is similar to this augmentation approach, except that I analyze the state with an ETKF, whereas I optimize the parameters on the basis of the variational free energy.

Variational Bayes and the reparametrization trick allows me to optimize data assimilation with the ETKF as core inference method, translating current observations $\mathbf{y}_t^o$ and additional parameters θ into a Gaussian posterior distribution. From the perspective of the parameters, the ETKF belongs to the estimation of the observational log-likelihood, whereas the ETKF sees the parameter estimation as pre-processing step, needed for its weights estimation. By this procedure, I am able to optimize parameters without the need of analytical expectations, and I can utilize tools developed for neural networks like PyTorch or TensorFlow. In the following, I outline an example where I optimize the observational error covariance based on an offline experiment for the atmosphere-land interface from Chapter 5.

Exemplary tuning of observational error covariance with variational Bayes

I show here that I can optimize the observational error covariance with the previously introduced approach, resulting in an approximation of the full error distribution. In this example, I use an offline experiment (see also Section 3.6),

where I assimilate 2-metre-temperature observations into the open-loop run of Chapter 5 with a non-localized ETKF on an hourly basis from 2015-07-31 12:00 UTC to 2015-08-07 11:00 UTC. I assume that the mapping between weights $\boldsymbol{w}_t$ at time t to the observations $\mathbf{y}_t^o$ is linear. Hence, I can use the static approximated sensitivity $\boldsymbol{\delta Y}_t^b$. As a consequence, I rely completely on the imposed covariances of the open-loop run.

I know from Chapter 2 that the observational error is an independent and identically distributed Gaussian $\mathcal{N}(0, 0.01\,\mathrm{K}^2)$. Thus, I simultaneously optimize the same variance posterior distribution for all 99 observations and all 168 time steps, which gives me $l = 16632$ samples. I use for the prior and posterior distribution of the observational error variance the inverse gamma distribution $\mathrm{IG}(\alpha, \beta)$, because the inverse gamma is one of the conjugated priors for the variance of a Gaussian distribution (Bishop, 2006). Since I want to show the ability to recover the correct observational error variance, I use here $\mathrm{IG}_{\mathrm{prior}} = \mathrm{IG}(2.1269, 0.1269)$ as prior, which results in a wrong mode of $(\sigma^o)^2 = 0.0406\,\mathrm{K}^2$.

I optimize α and β for 500 iterations with the Adam optimizer (Kingma and Ba, 2017) and a learning rate of 0.01. To ensure the positiveness of α and β, I specify their latent equivalent. I translate as one step this latent equivalent by the softplus function (Dugas et al., 2001) into the parameters needed for the inverse gamma distribution. For the stochastic optimization, I use $r = 4$ samples from the variance posterior and $s = 64$ samples from the ETKF posterior. I implemented the optimization in PyTorch (Paszke et al., 2019) together with the ETKF implementation from Chapter 3 (Finn, 2020b). The looped steps for the optimization are the following:

1. Sample r variances from the current variance posterior $(\sigma^{o(i)})^2 \sim \mathrm{IG}(\alpha, \beta)$, here for the i-th sample, with the reparametrization trick (6.16).

2. Based on all observations, their ensemble equivalent, and the drawn observational variances, estimate with the ETKF for all r samples the mean weights $\boldsymbol{w}_t^{a(i)}$ and weight covariances $\widetilde{\mathbf{P}}_t^{a(i)}$.

3. Sample s samples from the current ETKF posterior $\boldsymbol{w}_t^{(i,j)} \sim \mathcal{N}(\boldsymbol{w}_t^{a(i)}, \widetilde{\mathbf{P}}_t^{a(i)})$, here i, j describes the i-th sample from the variance posterior and the j sample from the ETKF posterior, with the reparametrization trick (6.16).

4. Propagate the sampled weights with the approximated sensitivity $\boldsymbol{\delta Y}_t^b$ and ensemble mean in observational space $\overline{\mathbf{y}}_t^b$ to observations

$$\mathbf{y}_t^{a(i,j)} = \overline{\mathbf{y}}_t^b + \boldsymbol{\delta Y}_t^b \boldsymbol{w}_t^{(i,j)}. \tag{6.17}$$

5. Evaluate the variational free energy (6.15) based on a factorized posterior for

the variance and ETKF here written as q,

$$J(\alpha,\beta) = -\mathbb{E}_q \log p(\mathbf{y}_t^o \mid \boldsymbol{w}_t, \sigma^o) + D_{KL}(\mathcal{N}(\boldsymbol{w}^a, \widetilde{\mathbf{P}}^a) \parallel \mathcal{N}(\mathbf{0}, (k-1)^{-1}\mathbf{I})) + D_{KL}(IG(\alpha,\beta) \parallel IG_{prior}).$$

6. Make a gradient descent step $\frac{\partial J(\theta)}{\partial \theta}$ with $\theta = (\alpha, \beta)$ to optimize the variance posterior $IG(\alpha, \beta)$.

For additional reference purpose, I optimize the observational error variance based on Desroziers et al. (2005). In this approach, the observational variance is determined by the deviation of the posterior ensemble mean in observational space to the observations $\Delta\mathbf{y}^a$ and the deviation of the prior ensemble mean to the observations $\Delta\mathbf{y}^b$. Since I know that the observational errors are not correlated, I can simply estimate the product of both quantities $(\sigma^o)^2 = l^{-1}\sum_i^l \Delta\mathbf{y}_l^a \Delta\mathbf{y}_l^b$, averaged over all $l = 16632$ samples with all time steps and observations. This estimated observational variance is iteratively used to determine a new analysis in observational space with the same observation operator (6.17) as for the variational Bayes procedure. As initial guess, I start with the mode of the IG_{prior} distribution $(\sigma^o)^2 = 0.0406\,K^2$, and Desroziers's method converges within 5 iterations.

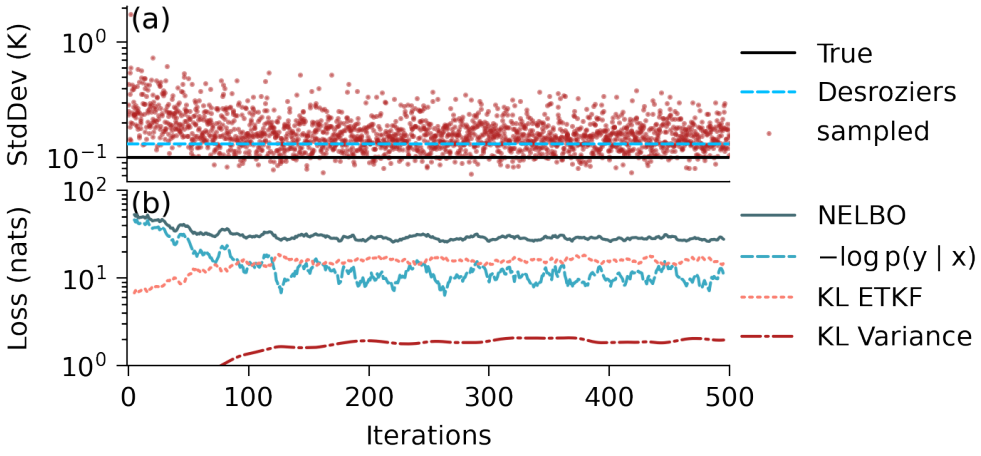


Figure 6.4: The (a) observational error standard deviation and (b) loss functions in dependence on the optimization steps. In (b), KL ETKF is the KL-divergence of the weights posterior to the weights prior from the ETKF inference, whereas KL Variance specifies the KL-divergence of the standard deviation posterior to prior. The negative log-likelihood $-\log\ p(\mathbf{y} \mid \mathbf{x})$ and the variational free energy (NELBO, negative evidence lower bound) are for visualization shifted by 60 nats, whereas all loss values are smoothed over 10 iterations. One nat is the natural unit of information, similar to one bit, but to the basis e.

The 4 samples from the variance posterior for each iteration are shown in Fig. 6.4, (a). As specified by the prior, the sampled standard deviations are miss-placed at early iterations and converge after around 100 iterations to an almost stationary distribution. This distribution is in most cases larger than the true observational error standard deviation of 0.1 K. Most of the samples are also larger than the

estimated standard deviations with Desroziers's method, caused by the shape of the inverse gamma distribution. The losses (Fig. 6.4, (b)) converge at a similar rate as the standard deviation. In the beginning, the loss is dominated by the observational likelihood because of the wrongly-specified standard deviations. In later iterations, the observational likelihood is reduced at the cost of the KL-divergence for the posteriors of the ETKF and of the observational error variance. Variational Bayes is therefore able to optimize the state variables and parameters at the same time.

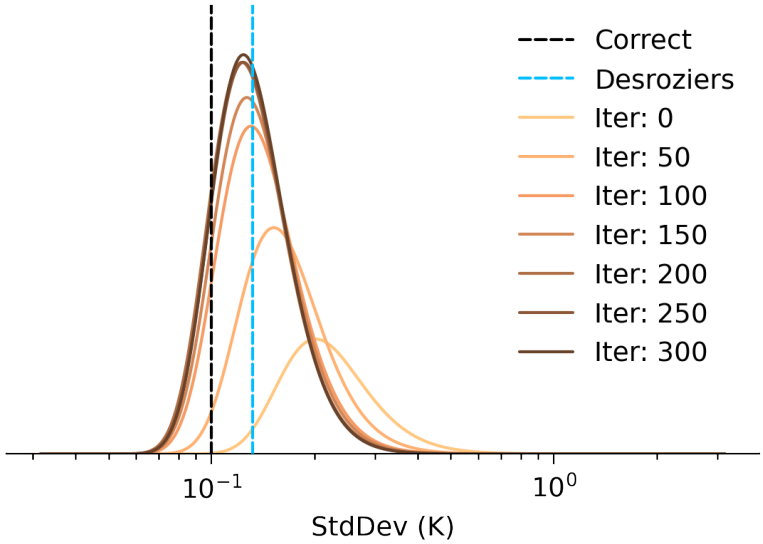


Figure 6.5: The optimized inverse gamma distributions for the observational error standard deviation in dependence on the iteration number. Optimization with $\mathbb{E}[\Delta\mathbf{y}^{a}(\Delta\mathbf{y}^{b})^{T}]$ (Desroziers et al., 2005) is shown in blue as Desroziers.

The posterior of the variance converges within 300 iterations to its almost stationary distribution (Fig. 6.5). For this stationary distribution, the correct observational error standard deviation is within a possible range of values, whereas the initial prior was miss-specified in terms of the probability for the correct value. The mode of the posterior distribution (σ^{o} =0.125 K) almost equals the solution gained with Desroziers et al. (2005) (σ^{o} =0.131 K) and both are larger than the correct standard deviation. In this offline experiment, I use a static and linearized observation operator, which is not influenced by the optimized parameters. Hence, the overestimation of the observational standard deviations is most likely a consequence of this static observation operator.

In Fig. 6.6, I specify the observational error standard deviation instead of α and β in the inverse gamma distribution. The Kullback-Leibler term of the inverse gamma distribution is therefore not shown in Fig. 6.6. The optimization is a convex problem for a linearized observation operator such that there is only one global minimum in the losses. This optimum of the variational free energy is almost the same solution as with Desroziers's method or the full inverse gamma distribution. Based on this result, I would expect that the full variational problem

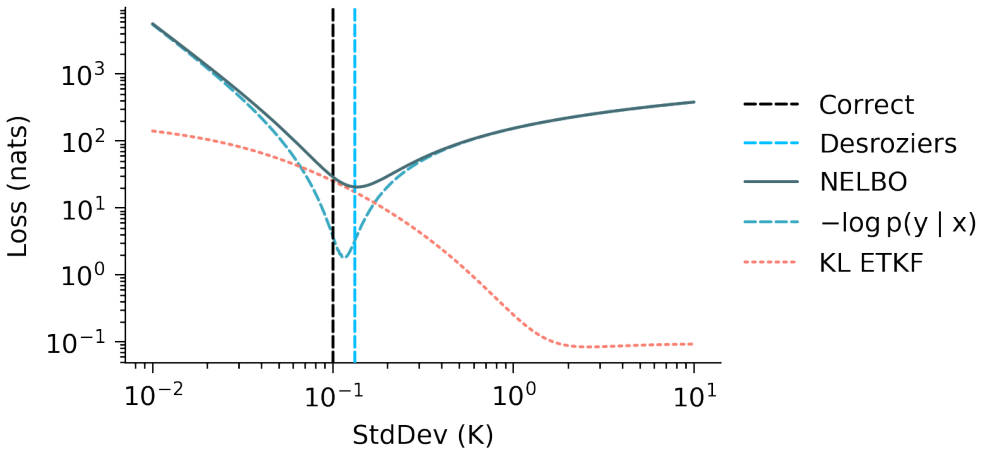


Figure 6.6: The negative evidence lower bound (NELBO) as loss function in dependence on the observational error standard deviation. The negative log-likelihood $-\log\ \mathrm{p}(\mathbf{y} \mid \mathbf{x})$ and the NELBO are for visualization shifted by 60 nats.

is also a convex problem with a global optimum, where the shape α and scale β are constrained to remain similar to their prior values.

This example shows that I can optimize parameters in the ETKF with variational Bayes. By using a prior distribution on the parameters, I can constrain a physical plausible area and nudge the solution towards this area. Furthermore, I get a full probability distribution, which is not possible if I would use expectation-maximization or a degenerated point estimate as posterior.

6.3 Discussion and Outlook

In this Chapter, I introduce two new theoretical approaches into data assimilation, which are especially related to the ETKF. I show that it is possible to use the kernel trick for feature-based data assimilation. By using a linear kernel, I recover the classical linear ETKF, but by choosing other kernels I can increase the flexibility of the ETKF. This increased flexibility can be also seen from a regression point of view. Instead of using a linear regression from observational space to model space, I use a kernel-based non-linear regression. In a simple wind speed example, I show the similarities of the kernelized ETKF to particle filters. The increased flexibility but comes also at a cost, because the kernel and its parameters have to be chosen and tuned.

To solve this optimization problem, I propose a second theoretical framework, the variational Bayes. In the variational Bayes framework, an approximated posterior is optimized based on its Kullback-Leibler divergence to the prior and the negative observational log-likelihood. I reveal that the ETKF is a specific solution to the variational free energy cost function such that parameters of the ETKF can be optimized, even if their analytical expression is not available. Because variational Bayes makes no difference between state and parameter, I can optimize the states and parameters at the same time. In an offline experiment based on the open-

loop run of Chapter 5, I present that this approach can be used to optimize the probability distribution of the observational error variance, here parametrized as inverse gamma distribution.

In my example, I completely rely on linearized ensemble statistics, including the sensitivity, mapping from weight or model space to observational space. This can be seen as if I optimize the observational error variance solely based on inner loops without an updated sensitivity. Hence, the variational Bayes procedure can be improved if I would also use outer loops, where the sensitivity is updated based on updated ensemble statistics. For this, I would need additional ensemble propagations, which would significantly increase the computational costs. Because variational Bayes and the reparametrization allows me to use any gradient descent algorithm, the approach is especially suited for online updates of the parameters, where the posterior at time t could be the next prior at time $t+1$. If I restrict the optimization to one single iteration per time step, this could be a way to reduce the computational costs, but to keep the updated sensitivity of the next time step.

With the same variational Bayes framework, it would be also possible to optimize the kernel parameters of the kernelized ETKF. Because a sum of scaled kernels is again a kernel, it would be also possible to optimize the kernel composition, which would allow a optimization of the importance of single kernels. These two approaches together have the possibility to automatize the finding of kernel functions for the kernelized ETKF.

The observational term in the kernelized cost function of the ETKF (6.2) allows me to draw another point of view on the kernelized ETKF. Since the observations $\mathbf{y}_t^o$ and the observational equivalent of the ensemble members $\mathbf{Y}_t^b$ are both samples from their corresponding probability distributions, the kernel embeds these distributions. Hence, the observational term corresponds to the so-called maximum mean discrepancy (MMD, Gretton et al. (2012, 2006)). In this distributional embedding, the kernel determines how many moments of the distributions are compared to each other. Therefore, the kernelized ETKF is in line with current research on how to use maximum mean discrepancy as an alternative criterion to maximum likelihood for inference procedures (Arbel et al., 2019; Briol et al., 2019; Cherief-Abdellatif and Alquier, 2020).

Since kernel methods belong the state-of-the-art methods for years in machine learning, one can speculate why kernel methods are not more often used in data assimilation. One cause might be the high-dimensionality in geoscience, which can lead to the curse of dimensionality for more flexible schemata than linear regression. The solution to this problem can be the use of additive kernels. They act only on specific dimensions (Duvenaud et al., 2011) and can avoid the curse of dimensionality in this way. Another cause might be the ever-increasing number of observations that have to be assimilated. As a consequence, it can be more important to increase the number of observations than to increase the efficiency of the assimilation for existing observations. In ensemble Kalman filters without B-matrix localization or inflation, the degrees of freedom for signal are bounded

by the number of ensemble members (Hotta and Ota, 2020). This bound can lead to a sub-optimal posterior ensemble if more observations than ensemble members are assimilated. Here, increased robustness of kernel methods and fingerprint operators can help us to use observations more efficiently.

Instead of specifying a kernel for feature extraction, I can also use neural networks (LeCun et al., 2015; Goodfellow et al., 2016). For a dynamical adaption of the neural network to the ensemble members, the ETKF can be used as last layer. In this way, the neural network as static function would then extract features based on their previous performance. The neural networks can be learned based on variational Bayes in a procedure that resembles variational autoencoders (Kingma and Welling, 2013). The neural network together with the ETKF would resemble an encoder, mapping from the observational data space to the latent model space, where as the observation operator acts as decoder, the other way around. This can be especially an advantage in cases, where the assumptions of linear observation operators and Gaussian distributions are violated. Here, variational Bayes could help us to move the non-linear and non-Gaussian parts of the inference problem to the feature extraction network.

All in all, these two additional theoretical approaches open a wide range of possibilities for ensemble Kalman filters. On one hand, they increase the flexibility of ensemble Kalman filters for non-linear data assimilation and provide parameter optimization within the ensemble Kalman filter. On the other hand, they offer another point of view, allowing a shift towards a more data-centric data assimilation, where features of observations are used to increase the efficiency of existing observations. This increased efficiency might be especially important in the case of coupled data assimilation across different Earth system components. Therefore, this chapter proves that recent developments in machine learning can be used to extend data assimilation.

7
Summary and Outlook

7.1 Summary

As I indicate in the introduction, I think that fully-coupled Earth system models together with machine learning are the most likely answer providing us another day of predictability for numerical weather prediction. As numerical weather prediction is an initial value problem, cross-compartmental initialization of Earth system models is therefore crucial but remains an unanswered question. This is a problem for simulations that should be as close as possible to the reality such as the project of a digital twin Earth. In this thesis, I propose two complementary frameworks as my contribution to gain the next day of predictability.

As proof of concept, I investigate the exemplary case of the atmosphere-land interface, where cross-compartmental data assimilation has been used for years. Currently, observations from the atmospheric boundary layer are assimilated across this interface into the soil moisture to primarily improve the forecast of the atmospheric boundary layer. One would expect that the coupling via the sensible heat flux and evapotranspiration between atmospheric boundary layer and land surface allows us to additionally improve the analysis of soil conditions. In contrast to this expectation, previous studies often found a negative assimilation impact such that data assimilation of boundary layer observations increased the analysis error for the soil moisture. It is quite difficult to disentangle the effects of different elements in the numerical weather prediction chain on this negative impact. Here, I take a step back from operational methods and use idealized experiments with a limited-area terrestrial model system. In these idealized experiments, I create a nature run with the same model configuration as for my data assimilation runs. This nature run acts as my reality to which I compare the results of my experiments. I synthesize 99 2-metre-temperature observations from this nature run, which are subsequently assimilated into the soil moisture with different types of data assimilation. Based on these idealized experiments, I am able to show that 2-metre-temperature observations can be used to improve the soil moisture analysis.

As my first framework, I propose to unify and couple the data assimilation with a localized ensemble Kalman filter for the initialization of Earth system models. In observational localization, every grid point is independently updated based on observations in the surrounding, weighted by their distance to the considered grid point. This observational localization allows me to assimilate 2-metre-

temperature observations without the need for an intermediate interpolation step as used in operational data assimilation for the soil moisture. By assimilating 2-metre-temperature observations at their observational sites, I also take horizontal dependencies between the 2-metre-temperature and soil moisture into account.

This framework is additionally based on the ensemble approach, where the model is run multiple times for the same time period, each time with slightly perturbed initial conditions. For a given time, I get an estimate for the uncertainty of the conditions, which can be then used in the data assimilation to weight the forecast against the observations. To update the soil moisture based on 2-metre-temperature observations, I have to estimate the sensitivities of perturbations in the 2-metre-temperature to perturbations in soil moisture, but the ensemble estimates these sensitivities automatically. Hence, I do not need additional methods like a tangent linear model or additional finite-differences' runs to update the soil moisture.

In my experiments, I reveal that the ensemble approach together with the observational localization decreases the analysis error in the soil moisture by up to 50 % compared to an implementation of the simplified extended Kalman filter, used in operational data assimilation. In addition, I discover that it is possible to update the soil moisture based on instantaneous 2-metre-temperature observations with this framework and an hourly cycle, as operationally used for numerical weather prediction with limited-area models in the atmosphere. This result suggests that the updated variables in the data assimilation for numerical weather prediction can be simply augmented by the soil moisture. This therefore proves that I can unify and couple the data assimilation for the atmosphere-land interfaces with a localized ensemble Kalman filter.

Perturbations in one component of the Earth system often have an effect on the forecast of another component with some delay. Especially in data assimilation for Earth system models, it is important to take these temporal dependencies into account. Hence, my second proposed framework is based on the efficient use of temporal dependencies by assimilating observations within a time window. I found that using 2-metre-temperature observations within a 24 hour window ahead of the update time improves the soil moisture analysis by around 10 %. Alternatively assimilating observations in a 24 hour window preceding the update time only decreases the analysis error by around 3 %. This discrepancy reveals that temporal dependencies can be better taken into account for Earth system models if observations ahead of the update time are used.

The sensitivity of 2-metre-temperature perturbations to soil moisture perturbations at update time decreases with increasing time difference. In addition, I introduce noise into the sensitivities because of the ensemble approximation. An increased assimilation window also increases the chance that the additional noise is larger than the additional signal. This would subsequently have a negative assimilation impact. I found in my experiments that using 2-metre-temperature observations at longer time windows than 24 hours have almost no additional gain for the soil

moisture analysis. This confirms an afterwards increased change of this so-called overfitting to the observations. In real-world data assimilation, the noise is much higher than in my idealized experiments, and overfitting would be a much larger problem.

As a novel solution, I introduce fingerprint operators. In fingerprint operators, I assimilate features of observations instead of the raw observations. With these observational features, I exploit specific fingerprints in observations that point towards errors in another component of the Earth system. In a feature screening, I show that the mean daytime temperature and the amplitude of a sine wave, both fitted within a 24 hour window, are physical-plausible fingerprint operators for the atmosphere-land interface. By using only these two features instead of 24 hourly raw observations, I increase the analysis error by only 6 %, but I increase the robustness against miss-specified observational and background covariance matrices. Furthermore, if I consider these two features as either uncorrelated or correlated, as they are in theory, I get almost the same analysis error. This additionally confirms the robustness against miss-specifications in the covariance matrices. As a consequence, the fingerprint operators stabilize coupled data assimilation across the atmosphere-land interface. I therefore propose as my second framework to use fingerprint operators to stabilize the initialization of Earth system models and make more out of the available observations.

I further generalize the approach of feature-based data assimilation with kernel-based data assimilation. In kernel-based data assimilation, I take advantage that the ensemble transform Kalman filter (ETKF) is formulated in weight space. I cast the dot product as similarity measure, which can be replaced by a positive-definite kernel. The kernel then specifies the moments and features of the data that are matched between the observation and their ensemble equivalent in observational space. In addition, I relate the ETKF to kernel least-squares, regressing from observational space to model space. The normal ETKF as a special case corresponds to a linear kernel and a linear regression. Therefore, the kernelized ETKF equals a non-linear regression from observational space to model space.

To tune parameters in the data assimilation, including additional parameters from the kernelized ETKF, I propose variational Bayes as a general approach. In variational Bayes, I optimize an approximated posterior based on an observational likelihood and its divergence to a specified prior. This general approach allows me to find the whole posterior distribution, whereas other methods only estimate a point estimate. Furthermore, variational Bayes does not differentiate between state variables and parameters. As a result, I can optimize both together in the same update step. I additionally prove that the ETKF is the optimal solution for the linearized-Gaussian special case. Hence, I can back-propagate the gradient through the ETKF by treating the ETKF as black box solver. In a simple example, I show that this procedure can be used to find an inverse gamma posterior of the error variance for my 2-metre-temperature observations in the atmosphere-land interface, despite a miss-specified prior.

These two theoretical approaches extend the data assimilation toolbox and increase the flexibility of data assimilation. On the one hand, the kernelized ETKF allows a wide range of kernels, including localization kernels. On the other hand, variational Bayes enables data-driven tuning of parameters and assimilation schemata. Together with the kernelized ETKF, variational Bayes can enable data-driven learning for the assimilation.

All in all, I show that coupled data assimilation across the atmosphere-land interface is possible and improves the soil moisture analysis. I further decrease the analysis error for the soil moisture by taking temporal dependencies up to 24 hours into account. To stabilize the cross-compartmental data assimilation, I take advantage of fingerprints in atmospheric boundary layer observations, pointing towards errors in the land surface.

7.2 Outlook

This thesis is in many ways a proof of concept for coupled data assimilation across the atmosphere-land interface. For these proofs of concept, I simplify my experiments in comparison to operational data assimilation, and coming directly back to operational data assimilation seems to be a too big leap to take at once. Hence, we would need smaller steps towards the big goal of strongly-coupled data assimilation in Earth system models. One of the smaller steps can be the use of a virtual reality (Schalge et al., 2020).

I would replace my nature run with such a virtual reality, which would have a higher resolution and another model configuration than used for my data assimilation experiments. In this way, model errors and biases can be simulated, a step that I have omitted in this thesis. As a consequence of this virtual reality, also observations would be on another grid than used for the experiments, causing possibly errors within the observation operator. Nevertheless, similar to my nature run, the virtual reality provides the possibility to define which observations are available, how densely these observations are distributed, and additionally, what observational error is used. On this basis, a virtual reality can be seen as next logical step between my idealized experiments and operational data assimilation.

Together with a modification of the reality and observations, the ensemble has to be altered as next step. In my ensemble, I concentrate only on the relationship between 2-metre-temperature and soil moisture. As a consequence, my only perturbations within the ensemble stem from initial perturbations in the soil moisture and in the soil temperature. I would need much more perturbations to represent the deviations of the ensemble to nature in more realistic experiments. These deviations can be introduced at the lateral boundary conditions, as operationally done in ICON-D2 (Reinert et al., 2021). This would result in an increased ensemble spread within the atmosphere, which could also propagate into the land surface. Another option is to perturb the model parameters so that, as a consequence, the ensemble members would cover a wider range of possible conditions. This last option would be possible in either the atmosphere or the land surface.

As third pillar, also the data assimilation has to be modified. Because of my idealized experiments, I only use prior multiplicative inflation to artificially increase the ensemble spread. It is likely that this inflation is too weak to counter-act errors and violations of assumptions in operational data assimilation. Therefore, other methods like relaxation-to-prior-perturbations (Zhang et al., 2004; Whitaker and Hamill, 2012) or stochastic perturbations (Palmer et al., 2009; Cardinali et al., 2014) might be needed. In addition, I perfectly know the covariance for the observational likelihood, which would be not the case in more realistic settings. To estimate such parameters more general methods than I use here are needed, e.g. Pulido et al. (2018) and Tandeo et al. (2020). My second approach for data-driven features is variational Bayes that is designed for exactly this problem of parameter estimation. Therefore, this might be a way forward for parameter estimation.

Another step towards a higher realism can be to extend the number of represented processes in the terrestrial system. Such an extension would represent ParFlow (Ashby and Falgout, 1996; Jones and Woodward, 2001; Kollet Stefan J. and Maxwell Reed M., 2008; Kollet and Maxwell, 2006; Maxwell, 2013; Maxwell and Miller, 2005) within the Terrestrial Modelling System Platform (TerrSysMP). ParFlow is a watershed model that allows a three-dimensional water flow and can simulate heterogeneous energy and water transports, which would then partially replace CLM for the representation of the soil moisture. In this way, TerrSysMP could simulate complex interactions and processes below the land surface. With this extension, TerrSysMP could represent hydrological processes within the Earth system more realistically. This might reduce model errors and biases in the land surface schema.

Throughout the thesis, I ignore the big elephant in the room for data assimilation – satellite observations. These satellite observations are especially relevant for data-sparse regions, where conventional observations such as 2-metre-temperature observations are only available to a limited extend (Duan et al., 2019). Furthermore, satellite observations provide a global picture about the soil moisture, whereas 2-metre-temperature observations have only a local picture. Nevertheless, this global picture makes it so difficult to assimilate these observations into a LETKF (Campbell et al., 2010; Miyoshi et al., 2010; Tsyrulnikov, 2013; Bonavita et al., 2015), because it is unclear how to assimilate the vertical correlation structure of satellite observations in an ensemble Kalman filter with observational localization. To avoid such problems, the ensemble is typically modulated in ensemble space (Bishop and Hodyss, 2009a,b; Bocquet, 2016; Bishop et al., 2017; Lei et al., 2018; Huang et al., 2019), but this modulation comes at higher computational costs and the LETKF would loss its simplistic form. It was previously suggested (Tsyrulnikov, 2013; Hotta and Ota, 2020) that the problems of the LETKF with satellite observations are related to the limited number of degrees of freedom within the ensemble. This problem is addressed by feature-based data assimilation. As feature-based data assimilation can condense the information from multiple observations into a few observational features, the problem with the limited number of degrees of freedom is circumvented. Furthermore, observational features make the data assimilation more robust against miss-specifications in the background and

observational covariance, as I show in my experiments. Feature-based ensemble Kalman filtering has therefore the chance to elegantly solve problems with satellite observations in the LETKF.

As I have written earlier, this thesis is first and foremost a proof of concept of initializing Earth system models. Here, I solely concentrate on the relationship between 2-metre-temperature and soil moisture and the atmosphere-land interface. As a consequence, it is difficult to generalize my results to other Earth system components. Nevertheless, I would expect similar result with similarly idealized experiments for other interfaces in the Earth system. This could then provide us with the possibility of coupled data assimilation in Earth system models, as schematically shown in Fig. 7.1.

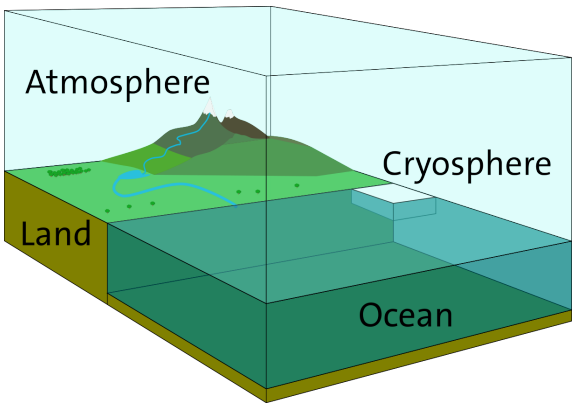


Figure 7.1: A schematic view on the coupled approach of initializing Earth system models. In the future, we might be able to initialize every here-shown component of the Earth together in a unified and coupled approach.

As I find in this thesis, the localized ensemble Kalman filter can be a general basis for the initialization of Earth system models. With their localization, they can assimilate observations across interfaces without the need of additional interpolation steps, and, with their ensemble approach, they can dynamically represent the background covariance, depending on the current conditions in the system. Furthermore, with their condensation of information, I would expect that fingerprint operators are also applicable for other interfaces, because they seem to be a general principle to improve and stabilize cross-compartmental data assimilation. Together, localized ensemble Kalman filters and fingerprint operators can be a cornerstone for future developments of the cross-compartmental initialization in Earth system models.

The aforementioned cyclone Lothar was a so-called extratropical "bomb" cyclone – its central pressure decreased by more than 24 hPa within 24 hours (Sanders and Gyakum, 1980; Black and Pezza, 2013). So, would it be possible to predicted such a cyclone by our forecast systems in 10 years? Studies show that extratropical "bomb" cyclones are also related to the coupling between the atmosphere and the ocean (Gómara et al., 2016; Kuwano-Yoshida and Minobe, 2017; Domingues et al., 2019), similar to tropical cyclones, although other processes are involved in their formation. This means that the forecast of these extratropical cyclones

would benefit from a coupled Earth system forecasting approach. Since I have proven in this thesis that a coupled initialization of the atmosphere-land interface is possible, I would expect that the same holds for the atmosphere-ocean interface. As a consequence, I expect that the forecast of such a extratropical "bomb" cyclones would be much better in this bright future, after the next innovation.

This page intentionally left blank

8
Conclusions

In this thesis, I present two complementary frameworks for the initialization of Earth system models, using the atmosphere-land interface as an exemplary case. As my first framework, I propose to use localized ensemble Kalman filtering for the unification of coupled data assimilation in Earth system models. As my second framework, I propose to use feature-based data assimilation to stabilize cross-compartmental data assimilation. Based on my results with idealized twin experiments, I conclude the following:

1. The soil moisture analysis can be improved by assimilating atmospheric boundary layer observations.
2. Because of its flow-dependency in the background covariance and in the sensitivities from the observations to the soil moisture, a ensemble Kalman filter is better suited for data assimilation across the atmosphere-land interface than a simplified extended Kalman filter.
3. By using observational localization, observations from the atmospheric boundary layer can be directly assimilated into the land surface without needing an intermediate interpolation step.
4. The soil moisture can be hourly updated on the basis of atmospheric boundary layer observations. This implies that the state vector for atmospheric data assimilation can be extended by the soil moisture.
5. Ensemble Kalman smoother with an assimilation window ahead of the update time can take advantage of temporal covariances across the atmosphere-land interface.
6. Fingerprint operators that take advantage of characteristic error fingerprints in observations stabilize data assimilation across the atmosphere-land interface.
7. By fingerprint operators, features of the diurnal cycle in the 2-metre-tempereature can be assimilated to improve the soil moisture analysis.
8. An ensemble transform Kalman filter with observational features can be generalized into a kernelized ensemble transform Kalman filter.
9. Variational Bayes can be used to learn parameters within the ensemble Kalman filter.
10. Developments in machine learning can be used to extend data assimilation.

This page intentionally left blank

Bibliography

Amezcua, Javier and Peter Jan Van Leeuwen (Dec. 2014). "Gaussian Anamorphosis in the Analysis Step of the EnKF: A Joint State-Variable/Observation Approach". In: *Tellus Dyn. Meteorol. Oceanogr.* 66.1, p. 23493. ISSN: null. DOI: 10.3402/tellusa.v66.23493.

Anderson, Brian D. O. et al. (1979). *Optimal Filtering. Information and System Science Series*.

Anderson, Jeffrey L. (Dec. 2001). "An Ensemble Adjustment Kalman Filter for Data Assimilation". en. In: *Mon. Wea. Rev.* 129.12, pp. 2884–2903. ISSN: 0027-0644. DOI: 10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

– (Apr. 2003). "A Local Least Squares Framework for Ensemble Filtering". EN. In: *Mon. Weather Rev.* 131.4, pp. 634–642. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.

Anderson, Jeffrey L. and Stephen L. Anderson (Dec. 1999). "A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts". EN. In: *Mon. Weather Rev.* 127.12, pp. 2741–2758. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.

Arbel, Michael et al. (Dec. 2019). "Maximum Mean Discrepancy Gradient Flow". In: *ArXiv190604370 Cs Stat*. arXiv: 1906.04370 [cs, stat].

Asch, Mark, Marc Bocquet, and Maëlle Nodet (Dec. 2016). *Data Assimilation*. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics. ISBN: 978-1-61197-453-9. DOI: 10.1137/1.9781611974546.

Ashby, Steven F. and Robert D. Falgout (Sept. 1996). "A Parallel Multigrid Preconditioned Conjugate Gradient Algorithm for Groundwater Flow Simulations". In: *NSE* 124.1, pp. 145–159. DOI: dx.doi.org/10.13182/NSE96-A24230.

Baldauf, Michael et al. (Apr. 2011). "Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities". In: *Mon. Wea. Rev.* 139.12, pp. 3887–3905. ISSN: 0027-0644. DOI: 10.1175/MWR-D-10-05013.1.

Balsamo, G., F. Bouyssel, and J. Noilhan (2004). "A Simplified Bi-Dimensional Variational Analysis of Soil Moisture from Screen-Level Observations in a Mesoscale Numerical Weather-Prediction Model". en. In: *Q. J. R. Meteorol. Soc.* 130.598, pp. 895–915. ISSN: 1477-870X. DOI: 10.1256/qj.02.215.

Bannister, R. N. (2017). "A Review of Operational Methods of Variational and Ensemble-Variational Data Assimilation". en. In: *Q. J. R. Meteorol. Soc.* 143.703, pp. 607–633. ISSN: 1477-870X. DOI: 10.1002/qj.2982.

Battin, Richard H. (1964). *Astronautical Guidance*. en. McGraw-Hill.

Bauer, Peter, Alan Thorpe, and Gilbert Brunet (2015). "The Quiet Revolution of Numerical Weather Prediction". en. In: *Nature* 525.7567, pp. 47–55. ISSN: 1476-4687. DOI: 10.1038/nature14956.

Bauer, Peter, Bjorn Stevens, and Wilco Hazeleger (Feb. 2021a). "A Digital Twin of Earth for the Green Transition". en. In: *Nat. Clim. Change* 11.2, pp. 80–83. ISSN: 1758-6798. DOI: 10.1038/s41558-021-00986-y.

Bauer, Peter et al. (Feb. 2021b). "The Digital Revolution of Earth-System Science". en. In: *Nat. Comput. Sci.* 1.2, pp. 104–113. ISSN: 2662-8457. DOI: 10.1038/s43588-021-00023-0.

Beal, Matthew J (2003). "Variational Algorithms for Approximate Bayesian Inference". PhD thesis. UCL (University College London).

Bélair, Stéphane et al. (Apr. 2003). "Operational Implementation of the ISBA Land Surface Scheme in the Canadian Regional Weather Forecast Model. Part I: Warm Season Results". In: *J. Hydrometeor.* 4.2, pp. 352–370. ISSN: 1525-755X. DOI: 10.1175/1525-7541(2003)4<352:OIOTIL>2.0.CO;2.

Bertino, Laurent, Geir Evensen, and Hans Wackernagel (2003). "Sequential Data Assimilation Techniques in Oceanography". en. In: *Int. Stat. Rev.* 71.2, pp. 223–241. ISSN: 1751-5823. DOI: 10.1111/j.1751-5823.2003.tb00194.x.

Best, M. J. et al. (Mar. 2015). "The Plumbing of Land Surface Models: Benchmarking Model Performance". In: *J. Hydrometeor.* 16.3, pp. 1425–1442. ISSN: 1525-755X. DOI: 10.1175/JHM-D-14-0158.1.

Bierman, Gerald J. (May 1977). *Factorization Methods for Discrete Sequential Estimation*. en. Academic Press. ISBN: 978-0-08-095637-4.

Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning*. en. Information Science and Statistics. New York: Springer. ISBN: 978-0-387-31073-2.

Bishop, Craig H. and Daniel Hodyss (2009a). "Ensemble Covariances Adaptively Localized with ECO-RAP. Part 1: Tests on Simple Error Models". en. In: *Tellus A* 61.1, pp. 84–96. ISSN: 1600-0870. DOI: 10.1111/j.1600-0870.2008.00371.x.

– (2009b). "Ensemble Covariances Adaptively Localized with ECO-RAP. Part 2: A Strategy for the Atmosphere". en. In: *Tellus A* 61.1, pp. 97–111. ISSN: 1600-0870. DOI: 10.1111/j.1600-0870.2008.00372.x.

Bishop, Craig H., Brian J. Etherton, and Sharanya J. Majumdar (Mar. 2001). "Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects". In: *Mon. Wea. Rev.* 129.3, pp. 420–436. ISSN: 0027-0644. DOI: 10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Bishop, Craig H., Jeffrey S. Whitaker, and Lili Lei (Nov. 2017). "Gain Form of the Ensemble Transform Kalman Filter and Its Relevance to Satellite Data Assimilation with Model Space Ensemble Covariance Localization". EN. In: *Mon. Weather Rev.* 145.11, pp. 4575–4592. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-17-0102.1.

Black, Mitchell Timothy and Alexandre Bernardes Pezza (2013). "A Universal, Broad-Environment Energy Conversion Signature of Explosive Cyclones". en. In: *Geophys. Res. Lett.* 40.2, pp. 452–457. ISSN: 1944-8007. DOI: 10.1002/grl.50114.

Bocquet, M. (Feb. 2008). "Inverse Modelling of Atmospheric Tracers: Non-Gaussian Methods and Second-Order Sensitivity Analysis". English. In: *Nonlinear Process. Geophys.* 15.1, pp. 127–143. ISSN: 1023-5809. DOI: 10.5194/npg-15-127-2008.

– (2016). "Localization and the Iterative Ensemble Kalman Smoother". en. In: *Q. J. R. Meteorol. Soc.* 142.695, pp. 1075–1089. ISSN: 1477-870X. DOI: 10.1002/qj.2711.

Bocquet, M. and P. Sakov (June 2012). "Combining Inflation-Free and Iterative Ensemble Kalman Filters for Strongly Nonlinear Systems". English. In: *Nonlinear Process. Geophys.* 19.3, pp. 383–399. ISSN: 1023-5809. DOI: 10.5194/npg-19-383-2012.

– (2014). "An Iterative Ensemble Kalman Smoother". en. In: *Q. J. R. Meteorol. Soc.* 140.682, pp. 1521–1535. ISSN: 1477-870X. DOI: 10.1002/qj.2236.

Bocquet, Marc, Carlos A. Pires, and Lin Wu (Aug. 2010). "Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation". EN. In: *Mon. Weather Rev.* 138.8, pp. 2997–3023. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/2010MWR3164.1.

Bonan, Bertrand et al. (Jan. 2020). "An Ensemble Square Root Filter for the Joint Assimilation of Surface Soil Moisture and Leaf Area Index within the Land Data Assimilation System LDAS-Monde: Application over the Euro-Mediterranean Region". English. In: *Hydrol. Earth Syst. Sci.* 24.1, pp. 325–347. ISSN: 1027-5606. DOI: 10.5194/hess-24-325-2020.

Bonavita, Massimo, Mats Hamrud, and Lars Isaksen (Dec. 2015). "EnKF and Hybrid Gain Ensemble Data Assimilation. Part II: EnKF and Hybrid Gain Results". EN. In: *Mon. Weather Rev.* 143.12, pp. 4865–4882. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-15-0071.1.

Breger, M et al. (1999). "30+ Frequencies for the Delta Scuti Variable 4 Canum Venaticorum: Results of the 1996 Multisite Campaign". In: *Astron. Astrophys.* 349, pp. 225–235.

Briol, Francois-Xavier et al. (June 2019). "Statistical Inference for Generative Models with Maximum Mean Discrepancy". In: *ArXiv190605944 Cs Math Stat*. arXiv: 1906.05944 [cs, math, stat].

Brunet, Gilbert et al. (Oct. 2010). "Collaboration of the Weather and Climate Communities to Advance Subseasonal-to-Seasonal Prediction". en. In: *Bull. Am. Meteorol. Soc.* 91.10, pp. 1397–1406. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/2010BAMS3013.1.

Brunet, Gilbert, Sarah Jones, Paolo M Ruti, et al. (2015). *Seamless Prediction of the Earth System: From Minutes to Months*. World Meteorological Organization.

Bundesanstalt fuer Geowissenschaften und Rohstoffe (2016). "Bodenuebersichtskarte". In:

Burgers, Gerrit, Peter Jan van Leeuwen, and Geir Evensen (June 1998). "Analysis Scheme in the Ensemble Kalman Filter". In: *Mon. Wea. Rev.* 126.6, pp. 1719–1724. ISSN: 0027-0644. DOI: 10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Campbell, William F., Craig H. Bishop, and Daniel Hodyss (Jan. 2010). "Vertical Covariance Localization for Satellite Radiances in Ensemble Kalman Filters". EN. In: *Mon. Weather Rev.* 138.1, pp. 282–290. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/2009MWR3017.1.

Cardinali, C. et al. (Sept. 2014). "Representing Model Error in Ensemble Data Assimilation". en. In: *Nonlin. Processes Geophys.* 21.5, pp. 971–985. ISSN: 1607-7946. DOI: 10.5194/npg-21-971-2014.

Carrera, Marco L., Stéphane Bélair, and Bernard Bilodeau (Mar. 2015). "The Canadian Land Data Assimilation System (CaLDAS): Description and Synthetic Evaluation Study". In: *J. Hydrometeor.* 16.3, pp. 1293–1314. ISSN: 1525-755X. DOI: 10.1175/JHM-D-14-0089.1.

Carrera, Marco L. et al. (Apr. 2019). "Assimilation of Passive L-Band Microwave Brightness Temperatures in the Canadian Land Data Assimilation System: Impacts on Short-Range Warm Season Numerical Weather Prediction". In: *J. Hydrometeor.* 20.6, pp. 1053–1079. ISSN: 1525-755X. DOI: 10.1175/JHM-D-18-0133.1.

Cherief-Abdellatif, Badr-Eddine and Pierre Alquier (Feb. 2020). "MMD-Bayes: Robust Bayesian Estimation via Maximum Mean Discrepancy". en. In: *Symposium on Advances in Approximate Bayesian Inference*. PMLR, pp. 1–21.

Cosme, Emmanuel et al. (Feb. 2012). "Smoothing Problems in a Bayesian Framework and Their Linear Gaussian Solutions". en. In: *Mon. Wea. Rev.* 140.2, pp. 683–695. ISSN: 0027-0644, 1520-0493. DOI: 10.1175/MWR-D-10-05025.1.

Courtier, P. et al. (1998). "The ECMWF Implementation of Three-Dimensional Variational Assimilation (3D-Var). I: Formulation". en. In: *Q. J. R. Meteorol. Soc.* 124.550, pp. 1783–1807. ISSN: 1477-870X. DOI: 10.1002/qj.49712455002.

Cressie, Noel A. C. (1993). *Statistics for Spatial Data*. en. Rev. ed. Wiley Series in Probability and Mathematical Statistics. New York: Wiley. ISBN: 978-0-471-00255-0.

Dai, Bin et al. (June 2017). "Hidden Talents of the Variational Autoencoder". In: *ArXiv170605148 Cs*. arXiv: 1706.05148 [cs].

Dask Development Team (2016). *Dask : Library for Dynamic Task Scheduling*.

Desroziers, Gérald et al. (2005). "Diagnosis of Observation, Background and Analysis-Error Statistics in Observation Space". In: *Q. J. R. Meteorol. Soc.* 131.613, pp. 3385–3396.

Desroziers, Gérald, Jean-Thomas Camino, and Loïk Berre (2014). "4DEnVar: Link with 4D State Formulation of Variational Assimilation and Different Possible Implementations". en. In: *Q. J. R. Meteorol. Soc.* 140.684, pp. 2097–2110. ISSN: 1477-870X. DOI: 10.1002/qj.2325.

Dharssi, I. et al. (Aug. 2011). "Operational Assimilation of ASCAT Surface Soil Wetness at the Met Office". English. In: *Hydrol. Earth Syst. Sci.* 15.8, pp. 2729–2746. ISSN: 1027-5606. DOI: 10.5194/hess-15-2729-2011.

Dimet, François-Xavier Le and Olivier Talagrand (1986). "Variational Algorithms for Analysis and Assimilation of Meteorological Observations: Theoretical Aspects". en. In: *Tellus A* 38A.2, pp. 97–110. ISSN: 1600-0870. DOI: 10.1111/j.1600-0870.1986.tb00459.x.

Dirmeyer, Paul A. et al. (Dec. 2017). "Verification of Land–Atmosphere Coupling in Forecast Models, Reanalyses, and Land Surface Models Using Flux Site Observations". In: *J. Hydrometeor.* 19.2, pp. 375–392. ISSN: 1525-755X. DOI: 10.1175/JHM-D-17-0152.1.

Domingues, Ricardo et al. (2019). "Ocean Observations in Support of Studies and Forecasts of Tropical and Extratropical Cyclones". English. In: *Front. Mar. Sci.* 6. ISSN: 2296-7745. DOI: 10.3389/fmars.2019.00446.

Doucet, Arnaud, Nando de Freitas, and Neil Gordon, eds. (2001). *Sequential Monte Carlo Methods in Practice*. en. Information Science and Statistics. New York: Springer-Verlag. ISBN: 978-0-387-95146-1. DOI: 10.1007/978-1-4757-3437-9.

Draper, C. S., J.-F. Mahfouf, and J. P. Walker (2011). "Root Zone Soil Moisture from the Assimilation of Screen-Level Variables and Remotely Sensed Soil

Moisture". en. In: *J. Geophys. Res. Atmospheres* 116.D2. ISSN: 2156-2202. DOI: 10.1029/2010JD013829.

Draper, Clara and Rolf H. Reichle (Mar. 2019). "Assimilation of Satellite Soil Moisture for Improved Atmospheric Reanalyses". In: *Mon. Wea. Rev.* 147.6, pp. 2163–2188. ISSN: 0027-0644. DOI: 10.1175/MWR-D-18-0393.1.

Drusch, Matthias and Pedro Viterbo (Feb. 2007). "Assimilation of Screen-Level Variables in ECMWF's Integrated Forecast System: A Study on the Impact on the Forecast Quality and Analyzed Soil Moisture". In: *Mon. Wea. Rev.* 135.2, pp. 300–314. ISSN: 0027-0644. DOI: 10.1175/MWR3309.1.

Duan, Qingyun et al., eds. (2019). *Handbook of Hydrometeorological Ensemble Forecasting*. en. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-39924-4 978-3-642-39925-1. DOI: 10.1007/978-3-642-39925-1.

Dugas, Charles et al. (2001). "Incorporating Second-Order Functional Knowledge for Better Option Pricing". In: *Adv. Neural Inf. Process. Syst.*, pp. 472–478.

Duvenaud, David (Nov. 2014). "Automatic Model Construction with Gaussian Processes". en. Thesis. University of Cambridge. DOI: 10.17863/CAM.14087.

Duvenaud, David, Hannes Nickisch, and Carl Edward Rasmussen (Dec. 2011). "Additive Gaussian Processes". In: *ArXiv11124394 Cs Stat*. arXiv: 1112.4394 [cs, stat].

ECMWF (2019). *IFS Documentation CY46R1*. IFS Documentation. ECMWF.

"ECMWF Strategy 2021-2030" (2021). In: *ECMWF Strategy 2021-2030*. ECMWF. DOI: 10.21957/s21ec694kd.

Entekhabi, Dara, Ignacio Rodriguez-Iturbe, and Fabio Castelli (Oct. 1996). "Mutual Interaction of Soil Moisture State and Atmospheric Processes". en. In: *Journal of Hydrology* 184.1-2, pp. 3–17. ISSN: 00221694. DOI: 10.1016/0022-1694(95)02965-6.

European Environment Agency (2013). "Digital Elevation Model over Europe (EU-DEM)". In:

Evensen, Geir (1992). "Using the Extended Kalman Filter with a Multilayer Quasi-Geostrophic Ocean Model". en. In: *J. Geophys. Res. Oceans* 97.C11, pp. 17905–17924. ISSN: 2156-2202. DOI: 10.1029/92JC01972.

– (1993). "Open Boundary Conditions for the Extended Kalman Filter with a Quasi-Geostrophic Ocean Model". en. In: *J. Geophys. Res. Oceans* 98.C9, pp. 16529–16546. ISSN: 2156-2202. DOI: 10.1029/93JC01365.

– (1994). "Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics". en. In: *J. Geophys. Res. Oceans* 99.C5, pp. 10143–10162. ISSN: 2156-2202. DOI: 10.1029/94JC00572.

Evensen, Geir and Peter Jan van Leeuwen (Jan. 1996). "Assimilation of Geosat Altimeter Data for the Agulhas Current Using the Ensemble Kalman Filter with a Quasigeostrophic Model". EN. In: *Mon. Weather Rev.* 124.1, pp. 85–96. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(1996)124<0085:AOGADF>2.0.CO;2.

– (June 2000). "An Ensemble Kalman Smoother for Nonlinear Dynamics". EN. In: *Mon. Weather Rev.* 128.6, pp. 1852–1867. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Fairbairn, D. et al. (Dec. 2015). "Comparing the Ensemble and Extended Kalman Filters for in Situ Soil Moisture Assimilation with Contrasting Conditions".

English. In: *Hydrol. Earth Syst. Sci.* 19.12, pp. 4811–4830. ISSN: 1027-5606. DOI: 10.5194/hess-19-4811-2015.

Fairbairn, David, Patricia de Rosnay, and Philip A. Browne (Aug. 2019). "The New Stand-Alone Surface Analysis at ECMWF: Implications for Land–Atmosphere DA Coupling". In: *J. Hydrometeor.* 20.10, pp. 2023–2042. ISSN: 1525-755X. DOI: 10.1175/JHM-D-19-0074.1.

Fatichi, Simone et al. (June 2016). "An Overview of Current Applications, Challenges, and Future Trends in Distributed Process-Based Models in Hydrology". en. In: *Journal of Hydrology* 537, pp. 45–60. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2016.03.026.

Finn, Tobias (Nov. 2020a). *Py_bacy (Version Paper Wind Profile)*. Zenodo. DOI: 10.5281/zenodo.4298499.

– (Aug. 2020b). *Torch-Assimilate (Version Paper_enkf)*. Zenodo. DOI: 10.5281/zenodo.4005995.

Finn, Tobias Sebastian, Gernot Geppert, and Felix Ament (May 2020). "Towards Assimilation of Wind Profile Observations in the Atmospheric Boundary Layer with a Sub-Kilometre-Scale Ensemble Data Assimilation System". In: *Tellus Dyn. Meteorol. Oceanogr.* 72.1, pp. 1–14. ISSN: null. DOI: 10.1080/16000870.2020.1764307.

Frolov, Sergey et al. (Jan. 2016). "Facilitating Strongly Coupled Ocean–Atmosphere Data Assimilation with an Interface Solver". EN. In: *Mon. Weather Rev.* 144.1, pp. 3–20. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-15-0041.1.

Gal, Yarin (2016). "Uncertainty in Deep Learning". PhD thesis.

Garreau, Damien, Wittawat Jitkrittum, and Motonobu Kanagawa (Oct. 2018). "Large Sample Analysis of the Median Heuristic". In: *ArXiv170707269 Math Stat*. arXiv: 1707.07269 [math, stat].

Gaspari, Gregory and Stephen E. Cohn (Jan. 1999). "Construction of Correlation Functions in Two and Three Dimensions". en. In: *Q. J. R. Meteorol. Soc.* 125.554, pp. 723–757. ISSN: 1477-870X. DOI: 10.1002/qj.49712555417.

Gasper, F. et al. (Oct. 2014). "Implementation and Scaling of the Fully Coupled Terrestrial Systems Modeling Platform (TerrSysMP v1.0) in a Massively Parallel Supercomputing Environment – a Case Study on JUQUEEN (IBM Blue Gene/Q)". English. In: *Geosci. Model Dev.* 7.5, pp. 2531–2543. ISSN: 1991-959X. DOI: 10.5194/gmd-7-2531-2014.

Gebler, S. et al. (Apr. 2017). "High Resolution Modelling of Soil Moisture Patterns with TerrSysMP: A Comparison with Sensor Network Data". en. In: *Journal of Hydrology* 547, pp. 309–331. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2017.01.048.

Geppert, Gernot (June 2015). "Analysis and Application of the Ensemble Kalman Filter for the Estimation of Bounded Quantities". eng. PhD thesis. Universität Hamburg Hamburg. DOI: 10.17617/2.2161673.

Giard, D. and E. Bazile (Apr. 2000). "Implementation of a New Assimilation Scheme for Soil and Surface Variables in a Global NWP Model". In: *Mon. Wea. Rev.* 128.4, pp. 997–1015. ISSN: 0027-0644. DOI: 10.1175/1520-0493(2000)128<0997:IOANAS>2.0.CO;2.

Giffin, Adom and Renaldas Urniezius (Feb. 2014). "The Kalman Filter Revisited Using Maximum Relative Entropy". en. In: *Entropy* 16.2, pp. 1047–1069. DOI: 10.3390/e16021047.

Gómara, Iñigo et al. (Nov. 2016). "Abrupt Transitions in the NAO Control of Explosive North Atlantic Cyclone Development". en. In: *Clim Dyn* 47.9, pp. 3091–3111. ISSN: 1432-0894. DOI: 10.1007/s00382-016-3015-9.

Gómez, Breogán et al. (Jan. 2020). "The Met Office Operational Soil Moisture Analysis System". en. In: *Remote Sens.* 12.22, p. 3691. DOI: 10.3390/rs12223691.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. MIT Press.

Gretton, Arthur et al. (2006). "A Kernel Method for the Two-Sample-Problem". In: *Adv. Neural Inf. Process. Syst.* 19, pp. 513–520.

Gretton, Arthur et al. (2012). "A Kernel Two-Sample Test". In: *J. Mach. Learn. Res.* 13.25, pp. 723–773.

Gustafsson, Nils et al. (2018). "Survey of Data Assimilation Methods for Convective-Scale Numerical Weather Prediction at Operational Centres". en. In: *Q. J. R. Meteorol. Soc.* 144.713, pp. 1218–1256. ISSN: 1477-870X. DOI: 10.1002/qj.3179.

Haario, Heikki, Leonid Kalachev, and Janne Hakkarainen (June 2015). "Generalized Correlation Integral Vectors: A Distance Concept for Chaotic Dynamical Systems". In: *Chaos* 25.6, p. 063102. ISSN: 1054-1500. DOI: 10.1063/1.4921939.

Hamill, Thomas M. and Chris Snyder (Aug. 2000). "A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme". EN. In: *Mon. Weather Rev.* 128.8, pp. 2905–2919. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

Hamrud, Mats, Massimo Bonavita, and Lars Isaksen (Dec. 2015). "EnKF and Hybrid Gain Ensemble Data Assimilation. Part I: EnKF Implementation". en. In: *Mon. Wea. Rev.* 143.12, pp. 4847–4864. ISSN: 0027-0644. DOI: 10.1175/MWR-D-14-00333.1.

Harlim, John and Brian R. Hunt (Jan. 2007). "Four-Dimensional Local Ensemble Transform Kalman Filter: Numerical Experiments with a Global Circulation Model". In: *Tellus Dyn. Meteorol. Oceanogr.* 59.5, pp. 731–748. ISSN: null. DOI: 10.1111/j.1600-0870.2007.00255.x.

Harrison, A. W. (Jan. 1981). "Effect of Atmospheric Humidity on Radiation Cooling". en. In: *Solar Energy* 26.3, pp. 243–247. ISSN: 0038-092X. DOI: 10.1016/0038-092X(81)90209-7.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. en. Second. Springer Series in Statistics. New York: Springer-Verlag. ISBN: 978-0-387-84857-0. DOI: 10.1007/978-0-387-84858-7.

Hess, R. (Sept. 2001). "Assimilation of Screen-Level Observations by Variational Soil Moisture Analysis". en. In: *Meteorol. Atmospheric Phys.* 77.1-4, pp. 145–154. ISSN: 0177-7971, 1436-5065. DOI: 10.1007/s007030170023.

Hinton, Geoffrey E. and Drew van Camp (Aug. 1993). "Keeping the Neural Networks Simple by Minimizing the Description Length of the Weights". In: *Proceedings of the Sixth Annual Conference on Computational Learning Theory*. COLT

'93. New York, NY, USA: Association for Computing Machinery, pp. 5–13. ISBN: 978-0-89791-611-0. DOI: 10.1145/168304.168306.

Hotta, Daisuke and Yoichiro Ota (May 2020). "What Limits the Number of Observations That Can Be Effectively Assimilated by EnKF?" In: *ArXiv200600517 Phys. Stat*. arXiv: 2006.00517 [physics, stat].

Houtekamer, P. L. and Herschel L. Mitchell (Mar. 1998). "Data Assimilation Using an Ensemble Kalman Filter Technique". EN. In: *Mon. Weather Rev.* 126.3, pp. 796–811. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

– (Jan. 2001). "A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation". en. In: *Mon. Wea. Rev.* 129.1, pp. 123–137. ISSN: 0027-0644. DOI: 10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Hoyer, Stephan and Joe Hamman (Apr. 2017). "Xarray : N-D Labeled Arrays and Datasets in Python". en. In: *J. Open Res. Softw.* 5.1, p. 10. ISSN: 2049-9647. DOI: 10.5334/jors.148.

Huang, Bo, Xuguang Wang, and Craig H. Bishop (Aug. 2019). "The High-Rank Ensemble Transform Kalman Filter". EN. In: *Mon. Weather Rev.* 147.8, pp. 3025–3043. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-18-0210.1.

Hunt, B. R. et al. (Jan. 2004). "Four-Dimensional Ensemble Kalman Filtering". In: *Tellus Dyn. Meteorol. Oceanogr.* 56.4, pp. 273–277. ISSN: null. DOI: 10.3402/tellusa.v56i4.14424.

Hunt, Brian R., Eric J. Kostelich, and Istvan Szunyogh (June 2007). "Efficient Data Assimilation for Spatiotemporal Chaos: A Local Ensemble Transform Kalman Filter". In: *Physica D: Nonlinear Phenomena*. Data Assimilation 230.1, pp. 112–126. ISSN: 0167-2789. DOI: 10.1016/j.physd.2006.11.008.

Idso, S. B. et al. (1975). "The Utility of Surface Temperature Measurements for the Remote Sensing of Surface Soil Water Status". en. In: *J. Geophys. Res. 1896-1977* 80.21, pp. 3044–3049. ISSN: 2156-2202. DOI: 10.1029/JC080i021p03044.

"IFS Documentation CY47R1 - Part II: Data Assimilation" (2020). In: *IFS Documentation CY47R1*. IFS Documentation 2. ECMWF. DOI: 10.21957/0gtybbwp9.

Jazwinski, Andrew H. (1970). *Stochastic Processes and Filtering Theory*. en. Academic Press. ISBN: 978-0-12-381550-7.

Jones, Jim E. and Carol S. Woodward (July 2001). "Newton–Krylov-Multigrid Solvers for Large-Scale, Highly Heterogeneous, Variably Saturated Flow Problems". In: *Advances in Water Resources* 24.7, pp. 763–774. ISSN: 0309-1708. DOI: 10.1016/S0309-1708(00)00075-0.

Jordan, Michael I. et al. (Nov. 1999). "An Introduction to Variational Methods for Graphical Models". en. In: *Machine Learning* 37.2, pp. 183–233. ISSN: 1573-0565. DOI: 10.1023/A:1007665907178.

Kalman, Rudolph Emil (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Trans. ASME–J. Basic Eng.* 82.Series D, pp. 35–45.

Kalnay, Eugenia (2003). *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge university press.

Kalnay, Eugenia and Shu-Chih Yang (2010). "Accelerating the Spin-up of Ensemble Kalman Filtering". en. In: *Q. J. R. Meteorol. Soc.* 136.651, pp. 1644–1651. ISSN: 1477-870X. DOI: 10.1002/qj.652.

Kalnay, Eugenia et al. (Jan. 2007a). "4-D-Var or Ensemble Kalman Filter?" In: *Tellus Dyn. Meteorol. Oceanogr.* 59.5, pp. 758–773. ISSN: null. DOI: 10.1111/j.1600-0870.2007.00261.x.

– (Jan. 2007b). "Response to the Discussion on "4-D-Var or EnKF?" By Nils Gustafsson". In: *Tellus Dyn. Meteorol. Oceanogr.* 59.5, pp. 778–780. ISSN: null. DOI: 10.1111/j.1600-0870.2007.00263.x.

Kanagawa, Motonobu et al. (July 2018). "Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences". In: *ArXiv180702582 Cs Stat*. arXiv: 1807.02582 [cs, stat].

Kauffeldt, Anna et al. (2015). "Imbalanced Land Surface Water Budgets in a Numerical Weather Prediction System". en. In: *Geophys. Res. Lett.* 42.11, pp. 4411–4417. ISSN: 1944-8007. DOI: 10.1002/2015GL064230.

Keil, Manfred, Michael Bock, and Thomas Esch (2011). *CORINE Land Cover 2006 - Europaweit Harmonisierte Aktualisierung Der Landbedeckungsdaten Für Deutschland*. Tech. rep. Umweltbundesamt.

Khan, Mohammad Emtiyaz et al. (June 2018). "Fast and Scalable Bayesian Deep Learning by Weight-Perturbation in Adam". en. In:

Kimeldorf, George S. and Grace Wahba (Apr. 1970). "A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines". In: *Ann. Math. Stat.* 41.2, pp. 495–502. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177697089.

Kingma, Diederik P. and Jimmy Ba (Jan. 2017). "Adam: A Method for Stochastic Optimization". In: *ArXiv14126980 Cs*. arXiv: 1412.6980 [cs].

Kingma, Diederik P. and Max Welling (Dec. 2013). "Auto-Encoding Variational Bayes". In: *ArXiv13126114 Cs Stat*. arXiv: 1312.6114 [cs, stat].

Kollet, Stefan et al. (Nov. 2018). "Introduction of an Experimental Terrestrial Forecasting/Monitoring System at Regional to Continental Scales Based on the Terrestrial Systems Modeling Platform (v1.1.0)". en. In: *Water* 10.11, p. 1697. DOI: 10.3390/w10111697.

Kollet, Stefan J. and Reed M. Maxwell (July 2006). "Integrated Surface–Groundwater Flow Modeling: A Free-Surface Overland Flow Boundary Condition in a Parallel Groundwater Flow Model". In: *Advances in Water Resources* 29.7, pp. 945–958. ISSN: 0309-1708. DOI: 10.1016/j.advwatres.2005.08.006.

Kollet Stefan J. and Maxwell Reed M. (2008). "Capturing the Influence of Groundwater Dynamics on Land Surface Processes Using an Integrated, Distributed Watershed Model". In: *Water Resources Research* 44.2. ISSN: 0043-1397. DOI: 10.1029/2007WR006004.

Kullback, S. and R. A. Leibler (Mar. 1951). "On Information and Sufficiency". EN. In: *Ann. Math. Statist.* 22.1, pp. 79–86. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177729694.

Kuwano-Yoshida, Akira and Shoshiro Minobe (Feb. 2017). "Storm-Track Response to SST Fronts in the Northwestern Pacific Region in an AGCM". EN. In: *J. Clim.* 30.3, pp. 1081–1102. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/JCLI-D-16-0331.1.

Kwon, In-Hyuk et al. (May 2018). "Assessment of Progress and Status of Data Assimilation in Numerical Weather Prediction". EN. In: *Bull. Am. Meteorol. Soc.* 99.5, ES75–ES79. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-D-17-0266.1.

Laloyaux, Patrick et al. (2018a). "CERA-20C: A Coupled Reanalysis of the Twentieth Century". en. In: *J. Adv. Model. Earth Syst.* 10.5, pp. 1172–1195. ISSN: 1942-2466. DOI: 10.1029/2018MS001273.

Laloyaux, Patrick et al. (2018b). "Implicit and Explicit Cross-Correlations in Coupled Data Assimilation". en. In: *Q. J. R. Meteorol. Soc.* 144.715, pp. 1851–1863. ISSN: 1477-870X. DOI: 10.1002/qj.3373.

Law, Kody, Andrew Stuart, and Konstantinos Zygalakis (2015). *Data Assimilation: A Mathematical Introduction*. en. Vol. 62. Texts in Applied Mathematics. Cham: Springer International Publishing. ISBN: 978-3-319-20324-9 978-3-319-20325-6. DOI: 10.1007/978-3-319-20325-6.

Lawrence, Peter J. and Thomas N. Chase (2007). "Representing a New MODIS Consistent Land Surface in the Community Land Model (CLM 3.0)". en. In: *J. Geophys. Res. Biogeosciences* 112.G1. ISSN: 2156-2202. DOI: 10.1029/2006JG000168.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (May 2015). "Deep Learning". en. In: *Nature* 521.7553, pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539.

Leeuwen, Peter Jan van and Geir Evensen (Dec. 1996). "Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation". EN. In: *Mon. Weather Rev.* 124.12, pp. 2898–2913. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.

Lei, Lili, Jeffrey S. Whitaker, and Craig Bishop (2018). "Improving Assimilation of Radiance Observations by Implementing Model Space Localization in an Ensemble Kalman Filter". en. In: *J. Adv. Model. Earth Syst.* 10.12, pp. 3221–3232. ISSN: 1942-2466. DOI: 10.1029/2018MS001468.

Li, Zhijin and I. M. Navon (2001). "Optimality of Variational Data Assimilation and Its Relationship with the Kalman Filter and Smoother". en. In: *Q. J. R. Meteorol. Soc.* 127.572, pp. 661–683. ISSN: 1477-870X. DOI: 10.1002/qj.49712757220.

Lin, Liao-Fan and Zhaoxia Pu (Sept. 2018). "Characteristics of Background Error Covariance of Soil Moisture and Atmospheric States in Strongly Coupled Land–Atmosphere Data Assimilation". In: *J. Appl. Meteor. Climatol.* 57.11, pp. 2507–2529. ISSN: 1558-8424. DOI: 10.1175/JAMC-D-18-0050.1.

– (Sept. 2019). "Examining the Impact of SMAP Soil Moisture Retrievals on Short-Range Weather Prediction under Weakly and Strongly Coupled Data Assimilation with WRF-Noah". In: *Mon. Wea. Rev.* 147.12, pp. 4345–4366. ISSN: 0027-0644. DOI: 10.1175/MWR-D-19-0017.1.

Lin, Liao-Fan et al. (2017). "Soil Moisture Background Error Covariance and Data Assimilation in a Coupled Land-Atmosphere Model". en. In: *Water Resour. Res.* 53.2, pp. 1309–1335. ISSN: 1944-7973. DOI: 10.1002/2015WR017548.

Liu, Chengsi, Qingnong Xiao, and Bin Wang (Sept. 2008). "An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test". EN. In: *Mon. Weather Rev.* 136.9, pp. 3363–3373. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/2008MWR2312.1.

Liu, Junkai and Zhaoxia Pu (2019). "Does Soil Moisture Have an Influence on Near-Surface Temperature?" en. In: *J. Geophys. Res. Atmospheres* 124.12, pp. 6444–6466. ISSN: 2169-8996. DOI: 10.1029/2018JD029750.

Lorenc, A. C. (Apr. 1981). "A Global Three-Dimensional Multivariate Statistical Interpolation Scheme". EN. In: *Mon. Weather Rev.* 109.4, pp. 701–721. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(1981)109<0701:AGTDMS>2.0.CO;2.

Lorenc, Andrew C. (2003). "The Potential of the Ensemble Kalman Filter for NWP—a Comparison with 4D-Var". en. In: *Q. J. R. Meteorol. Soc.* 129.595, pp. 3183–3203. ISSN: 1477-870X. DOI: 10.1256/qj.02.132.

Lorenc, Andrew C. and Mohamed Jardak (2018). "A Comparison of Hybrid Variational Data Assimilation Methods for Global NWP". en. In: *Q. J. R. Meteorol. Soc.* 144.717, pp. 2748–2760. ISSN: 1477-870X. DOI: 10.1002/qj.3401.

Lorenz, Edward N. (Mar. 1963). "Deterministic Nonperiodic Flow". In: *J. Atmos. Sci.* 20.2, pp. 130–141. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Luo, Xiaodong (July 2019). "Ensemble-Based Kernel Learning for a Class of Data Assimilation Problems with Imperfect Forward Simulators". en. In: *PLOS ONE* 14.7, e0219247. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0219247.

Mahfouf, J.-F. et al. (2009). "A Comparison of Two Off-Line Soil Analysis Schemes for Assimilation of Screen Level Observations". en. In: *J. Geophys. Res. Atmospheres* 114.D8. ISSN: 2156-2202. DOI: 10.1029/2008JD011077.

Mahfouf, Jean-François (Nov. 1991). "Analysis of Soil Moisture from Near-Surface Parameters: A Feasibility Study". In: *J. Appl. Meteor.* 30.11, pp. 1534–1547. ISSN: 0894-8763. DOI: 10.1175/1520-0450(1991)030<1534:AOSMFN>2.0.CO;2.

Maxwell, Reed M. (Mar. 2013). "A Terrain-Following Grid Transform and Preconditioner for Parallel, Large-Scale, Integrated Hydrologic Modeling". In: *Advances in Water Resources* 53, pp. 109–117. ISSN: 0309-1708. DOI: 10.1016/j.advwatres.2012.10.001.

Maxwell, Reed M. and Norman L. Miller (June 2005). "Development of a Coupled Land Surface and Groundwater Model". In: *J. Hydrometeor.* 6.3, pp. 233–247. ISSN: 1525-755X. DOI: 10.1175/JHM422.1.

Maybeck, Peter S. (Aug. 1982). *Stochastic Models, Estimation, and Control*. en. Academic Press. ISBN: 978-0-08-096003-6.

Ménard, Richard and Roger Daley (1996). "The Application of Kalman Smoother Theory to the Estimation of 4DVAR Error Statistics". en. In: *Tellus A* 48.2, pp. 221–237. ISSN: 1600-0870. DOI: 10.1034/j.1600-0870.1996.t01-1-00003.x.

Mercer, James and Andrew Russell Forsyth (Jan. 1909). "XVI. Functions of Positive and Negative Type, and Their Connection the Theory of Integral Equations". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 209.441-458, pp. 415–446. DOI: 10.1098/rsta.1909.0016.

Micchelli, Charles A., Yuesheng Xu, and Haizhang Zhang (2006). "Universal Kernels". In: *J. Mach. Learn. Res.* 7.95, pp. 2651–2667. ISSN: 1533-7928.

Milan, Marco et al. (2020). "Hourly 4D-Var in the Met Office UKV Operational Forecast Model". en. In: *Q. J. R. Meteorol. Soc.* 146.728, pp. 1281–1301. ISSN: 1477-870X. DOI: 10.1002/qj.3737.

Milbrandt, Jason A. et al. (Sept. 2016). "The Pan-Canadian High Resolution (2.5 Km) Deterministic Prediction System". In: *Wea. Forecasting* 31.6, pp. 1791–1816. ISSN: 0882-8156. DOI: 10.1175/WAF-D-16-0035.1.

Mitchell, Herschel L. and P. L. Houtekamer (Feb. 2000). "An Adaptive Ensemble Kalman Filter". EN. In: *Mon. Weather Rev.* 128.2, pp. 416–433. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(2000)128<0416:AAEKF>2.0.CO;2.

Mitter, Sanjoy K. and Nigel J. Newton (Jan. 2005). "Information and Entropy Flow in the Kalman–Bucy Filter". en. In: *J Stat Phys* 118.1, pp. 145–176. ISSN: 1572-9613. DOI: 10.1007/s10955-004-8781-9.

Miyoshi, Takemasa, Yoshiaki Sato, and Takashi Kadowaki (July 2010). "Ensemble Kalman Filter and 4D-Var Intercomparison with the Japanese Operational Global Analysis and Prediction System". EN. In: *Mon. Weather Rev.* 138.7, pp. 2846–2866. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/2010MWR3209.1.

Miyoshi, Takemasa, Keiichi Kondo, and Toshiyuki Imamura (2014). "The 10,240-Member Ensemble Kalman Filtering with an Intermediate AGCM". en. In: *Geophys. Res. Lett.* 41.14, pp. 5264–5271. ISSN: 1944-8007. DOI: 10.1002/2014GL060863.

Monin, Andreı Sergeevich and Aleksandr Mikhaılovich Obukhov (1954). "Basic Laws of Turbulent Mixing in the Surface Layer of the Atmosphere". In: *Contrib Geophys Inst Acad Sci USSR* 151.163, e187.

Morzfeld, Matthias et al. (May 2018). "Feature-Based Data Assimilation in Geophysics". English. In: *Nonlinear Process. Geophys.* 25.2, pp. 355–374. ISSN: 1023-5809. DOI: 10.5194/npg-25-355-2018.

Muandet, Krikamol et al. (June 2017). "Kernel Mean Embedding of Distributions: A Review and Beyond". English. In: *MAL* 10.1-2, pp. 1–141. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000060.

Mucia, Anthony et al. (Jan. 2020). "From Monitoring to Forecasting Land Surface Conditions Using a Land Data Assimilation System: Application over the Contiguous United States". en. In: *Remote Sens.* 12.12, p. 2020. DOI: 10.3390/rs12122020.

Muñoz-Sabater, J. et al. (2019). "Assimilation of SMOS Brightness Temperatures in the ECMWF Integrated Forecasting System". en. In: *Q. J. R. Meteorol. Soc.* 145.723, pp. 2524–2548. ISSN: 1477-870X. DOI: 10.1002/qj.3577.

Murphy, Kevin P. (2012). *Machine Learning: A Probabilistic Perspective*. en. Adaptive Computation and Machine Learning Series. Cambridge, MA: MIT Press. ISBN: 978-0-262-01802-9.

Myneni, R. B. et al. (Jan. 2002). "Global Products of Vegetation Leaf Area and Fraction Absorbed PAR from Year One of MODIS Data". In: *NASA Publ.*

Oleson, Keith et al. (2004). *Technical Description of the Community Land Model (CLM)*. en. Tech. rep. National Center for Atmospheric Research.

Oleson K. W. et al. (2008). "Improvements to the Community Land Model and Their Impact on the Hydrological Cycle". In: *Journal of Geophysical Research: Biogeosciences* 113.G1. ISSN: 0148-0227. DOI: 10.1029/2007JG000563.

Orth, Rene et al. (May 2017). "Advancing Land Surface Model Development with Satellite-Based Earth Observations". en. In: *Hydrol. Earth Syst. Sci.* 21.5, pp. 2483–2495. ISSN: 1607-7938. DOI: 10.5194/hess-21-2483-2017.

Ott, Edward et al. (2002). "Exploiting Local Low Dimensionality of the Atmospheric Dynamics for Efficient Ensemble Kalman Filtering". In: *ArXiv Prepr. Physics0203058*. arXiv: physics/0203058.

Ott, Edward et al. (July 2003). "A Local Ensemble Kalman Filter for Atmospheric Data Assimilation". In: *arXiv:physics/0203058*. arXiv: physics/0203058.

Palmer, Tim and Bjorn Stevens (Dec. 2019). "The Scientific Challenge of Understanding and Estimating Climate Change". en. In: *PNAS* 116.49, pp. 24390–24395. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1906691116.

Palmer, T.N. et al. (Oct. 2009). "Stochastic Parametrization and Model Uncertainty". In: 598, p. 42. DOI: 10.21957/ps8gbwbdv.

Parrish, David F. and John C. Derber (Aug. 1992). "The National Meteorological Center's Spectral Statistical-Interpolation Analysis System". EN. In: *Mon. Weather Rev.* 120.8, pp. 1747–1763. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.

Paszke, Adam et al. (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., pp. 8024–8035.

Penny, Stephen G. and Thomas M. Hamill (Mar. 2017). "Coupled Data Assimilation for Integrated Earth System Analysis and Prediction". In: *Bull. Amer. Meteor. Soc.* 98.7, ES169–ES172. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-17-0036.1.

PrefectHQ (Mar. 2021). *Prefect*. Prefect.

Prein, Andreas F. et al. (2015). "A Review on Regional Convection-Permitting Climate Modeling: Demonstrations, Prospects, and Challenges". en. In: *Rev. Geophys.* 53.2, pp. 323–361. ISSN: 1944-9208. DOI: 10.1002/2014RG000475.

Pulido, Manuel and Peter Jan van Leeuwen (Nov. 2019). "Sequential Monte Carlo with Kernel Embedded Mappings: The Mapping Particle Filter". en. In: *Journal of Computational Physics* 396, pp. 400–415. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2019.06.060.

Pulido, Manuel et al. (Jan. 2018). "Stochastic Parameterization Identification Using Ensemble Kalman Filtering Combined with Maximum Likelihood Methods". In: *Tellus Dyn. Meteorol. Oceanogr.* 70.1, pp. 1–17. ISSN: null. DOI: 10.1080/16000870.2018.1442099.

Pulido, Manuel, Peter Jan vanLeeuwen, and Derek J. Posselt (Jan. 2019). "Kernel Embedded Nonlinear Observational Mappings in the Variational Mapping Particle Filter". In: *ArXiv190110426 Cs Stat*. arXiv: 1901.10426 [cs, stat].

Rasmussen, Carl Edward and Christopher K. I. Williams (2006). *Gaussian Processes for Machine Learning*. en. Adaptive Computation and Machine Learning. Cambridge, Mass: MIT Press. ISBN: 978-0-262-18253-9.

Reichle, Rolf H. and Randal D. Koster (Dec. 2003). "Assessing the Impact of Horizontal Error Correlations in Background Fields on Soil Moisture Estimation". In: *J. Hydrometeor.* 4.6, pp. 1229–1242. ISSN: 1525-755X. DOI: 10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2.

Reichle, Rolf H. et al. (Dec. 2002). "Extended versus Ensemble Kalman Filtering for Land Data Assimilation". In: *J. Hydrometeor.* 3.6, pp. 728–740. ISSN: 1525-755X. DOI: 10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2.

Reinert, D et al. (2021). *DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System*. Tech. rep.

Rezende, Danilo Jimenez, Shakir Mohamed, and Daan Wierstra (Jan. 2014). "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". In: *ArXiv14014082 Cs Stat*. arXiv: 1401.4082 [cs, stat].

Rocklin, Matthew (2015). "Dask : Parallel Computation with Blocked Algorithms and Task Scheduling". en. In: *Python in Science Conference*. Austin, Texas, pp. 126–132. DOI: 10.25080/Majora-7b98e3ed-013.

Rosenthal, W. Steven et al. (Feb. 2017). "Displacement Data Assimilation". en. In: *Journal of Computational Physics* 330, pp. 594–614. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2016.10.025.

Rosnay, Patricia de et al. (2013). "A Simplified Extended Kalman Filter for the Global Operational Soil Moisture Analysis at ECMWF". en. In: *Q. J. R. Meteorol. Soc.* 139.674, pp. 1199–1213. ISSN: 1477-870X. DOI: 10.1002/qj.2023.

Sætrom, Jon and Henning Omre (June 2011). "Ensemble Kalman Filtering for Non-Linear Likelihood Models Using Kernel-Shrinkage Regression Techniques". en. In: *Comput Geosci* 15.3, pp. 529–544. ISSN: 1573-1499. DOI: 10.1007/s10596-010-9222-2.

Sakov, Pavel, Dean S. Oliver, and Laurent Bertino (June 2012). "An Iterative EnKF for Strongly Nonlinear Systems". EN. In: *Mon. Weather Rev.* 140.6, pp. 1988–2004. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-11-00176.1.

Sanders, Frederick and John R. Gyakum (Oct. 1980). "Synoptic-Dynamic Climatology of the "Bomb"". EN. In: *Mon. Weather Rev.* 108.10, pp. 1589–1606. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(1980)108<1589:SDCOT>2.0.CO;2.

Santanello Jr., Joseph A. et al. (Mar. 2019). "Understanding the Impacts of Soil Moisture Initial Conditions on NWP in the Context of Land–Atmosphere Coupling". In: *J. Hydrometeor.* 20.5, pp. 793–819. ISSN: 1525-755X. DOI: 10.1175/JHM-D-18-0186.1.

Sawada, Yohei, Tosiyuki Nakaegawa, and Takemasa Miyoshi (2018). "Hydrometeorology as an Inversion Problem: Can River Discharge Observations Improve the Atmosphere by Ensemble Data Assimilation?" en. In: *J. Geophys. Res. Atmospheres* 123.2, pp. 848–860. ISSN: 2169-8996. DOI: 10.1002/2017JD027531.

Schalge, Bernd et al. (Mar. 2020). "Presentation and Discussion of the High Resolution Atmosphere-Land Surface Subsurface Simulation Dataset of the Virtual Neckar Catchment for the Period 2007–2015". English. In: *Earth Syst. Sci. Data Discuss.*, pp. 1–40. ISSN: 1866-3508. DOI: 10.5194/essd-2020-24.

Schär, Christoph et al. (May 2020). "Kilometer-Scale Climate Models: Prospects and Challenges". EN. In: *Bull. Am. Meteorol. Soc.* 101.5, E567–E587. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-D-18-0167.1.

Schepers, Dinand et al. (2018). "CERA-SAT: A Coupled Satellite-Era Reanalysis". en. In: DOI: 10.21957/SP619DS74G.

Schölkopf, B. and AJ. Smola (Dec. 2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press / Biologische Kybernetik, p. 644.

Schölkopf, Bernhard, Alexander Smola, and Klaus-Robert Müller (1997). "Kernel Principal Component Analysis". en. In: *Artificial Neural Networks — ICANN'97*. Ed. by Wulfram Gerstner et al. Lecture Notes in Computer Science. Berlin,

Heidelberg: Springer, pp. 583–588. ISBN: 978-3-540-69620-9. DOI: 10.1007/BFb0020217.
– (July 1998). "Nonlinear Component Analysis as a Kernel Eigenvalue Problem". In: *Neural Computation* 10.5, pp. 1299–1319. ISSN: 0899-7667. DOI: 10.1162/089976698300017467.
Schölkopf, Bernhard, Ralf Herbrich, and Alex J. Smola (2001). "A Generalized Representer Theorem". en. In: *Computational Learning Theory*. Ed. by David Helmbold and Bob Williamson. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 416–426. ISBN: 978-3-540-44581-4. DOI: 10.1007/3-540-44581-1_27.
Schraff, C. et al. (2016). "Kilometre-Scale Ensemble Data Assimilation for the COSMO Model (KENDA)". In: *Q. J. R. Meteorol. Soc.* 142.696, pp. 1453–1472. ISSN: 1477-870X. DOI: 10.1002/qj.2748.
Shahabadi, Maziar Bani et al. (Sept. 2019). "Impact of Weak Coupling between Land and Atmosphere Data Assimilation Systems on Environment and Climate Change Canada's Global Deterministic Prediction System". In: *Wea. Forecasting* 34.6, pp. 1741–1758. ISSN: 0882-8156. DOI: 10.1175/WAF-D-19-0073.1.
Shrestha, P. et al. (Apr. 2014). "A Scale-Consistent Terrestrial Systems Modeling Platform Based on COSMO, CLM, and ParFlow". In: *Mon. Wea. Rev.* 142.9, pp. 3466–3483. ISSN: 0027-0644. DOI: 10.1175/MWR-D-14-00029.1.
Sluka, Travis C. et al. (2016). "Assimilating Atmospheric Observations into the Ocean Using Strongly Coupled Ensemble Data Assimilation". en. In: *Geophys. Res. Lett.* 43.2, pp. 752–759. ISSN: 1944-8007. DOI: 10.1002/2015GL067238.
Smith, Polly J., Alison M. Fowler, and Amos S. Lawless (Dec. 2015). "Exploring Strategies for Coupled 4D-Var Data Assimilation Using an Idealised Atmosphere–Ocean Model". In: *Tellus Dyn. Meteorol. Oceanogr.* 67.1, p. 27025. ISSN: null. DOI: 10.3402/tellusa.v67.27025.
Smith, Polly J., Amos S. Lawless, and Nancy K. Nichols (Oct. 2017). "Estimating Forecast Error Covariances for Strongly Coupled Atmosphere–Ocean 4D-Var Data Assimilation". EN. In: *Mon. Weather Rev.* 145.10, pp. 4011–4035. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-16-0284.1.
Sollich, Peter and Christopher Williams (2004). "Using the Equivalent Kernel to Understand Gaussian Process Regression". en. In: *Adv. Neural Inf. Process. Syst.* 17.
Spantini, Alessio, Ricardo Baptista, and Youssef Marzouk (June 2019). "Coupling Techniques for Nonlinear Ensemble Filtering". In: *ArXiv190700389 Stat*. arXiv: 1907.00389 [stat].
Sriperumbudur, Bharath K., Kenji Fukumizu, and Gert R. G. Lanckriet (2011). "Universality, Characteristic Kernels and RKHS Embedding of Measures". In: *J. Mach. Learn. Res.* 12.70, pp. 2389–2410. ISSN: 1533-7928.
Steppeler, J et al. (2003). "Meso-Gamma Scale Forecasts Using the Nonhydrostatic Model LM". In: *Meteorol. Atmospheric Phys.* 82.1, pp. 75–96.
Stevens, Bjorn and Sandrine Bony (May 2013). "What Are Climate Models Missing?" en. In: *Science* 340.6136, pp. 1053–1054. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1237554.

Stordal, Andreas S. et al. (Apr. 2021). "P-Kernel Stein Variational Gradient Descent for Data Assimilation and History Matching". en. In: *Math Geosci* 53.3, pp. 375–393. ISSN: 1874-8953. DOI: 10.1007/s11004-021-09937-x.

Storto, Andrea et al. (Apr. 2018). "Strongly Coupled Data Assimilation Experiments with Linearized Ocean–Atmosphere Balance Relationships". EN. In: *Mon. Weather Rev.* 146.4, pp. 1233–1257. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-17-0222.1.

Stull, Roland B. (1988). *An Introduction to Boundary Layer Meteorology*. en. Atmospheric and Oceanographic Sciences Library. Springer Netherlands. ISBN: 978-90-277-2768-8. DOI: 10.1007/978-94-009-3027-8.

Su, Z. et al. (2013). "Evaluation of ECMWF's Soil Moisture Analyses Using Observations on the Tibetan Plateau". en. In: *J. Geophys. Res. Atmospheres* 118.11, pp. 5304–5318. ISSN: 2169-8996. DOI: 10.1002/jgrd.50468.

Talagrand, Olivier and Philippe Courtier (1987). "Variational Assimilation of Meteorological Observations With the Adjoint Vorticity Equation. I: Theory". en. In: *Q. J. R. Meteorol. Soc.* 113.478, pp. 1311–1328. ISSN: 1477-870X. DOI: 10.1002/qj.49711347812.

Tandeo, Pierre et al. (Sept. 2020). "A Review of Innovation-Based Methods to Jointly Estimate Model and Observation Error Covariance Matrices in Ensemble Data Assimilation". EN. In: *Mon. Weather Rev.* 148.10, pp. 3973–3994. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-19-0240.1.

Tippett, Michael K. et al. (July 2003). "Ensemble Square Root Filters". EN. In: *Mon. Weather Rev.* 131.7, pp. 1485–1490. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

Tsyrulnikov, Mikhail (2013). "Is the Local Ensemble Transform Kalman Filter Suitable for Operational Data Assimilation?" In:

Valcke, S. (Mar. 2013). "The OASIS3 Coupler: A European Climate Modelling Community Software". English. In: *Geosci. Model Dev.* 6.2, pp. 373–388. ISSN: 1991-959X. DOI: 10.5194/gmd-6-373-2013.

Vereecken, H. et al. (May 2016). "Modeling Soil Processes: Review, Key Challenges, and New Perspectives". en. In: *Vadose Zone Journal* 15.5. ISSN: 1539-1663. DOI: 10.2136/vzj2015.09.0131.

Wang, Xuguang and Ting Lei (Sept. 2014). "GSI-Based Four-Dimensional Ensemble–Variational (4DEnsVar) Data Assimilation: Formulation and Single-Resolution Experiments with Real Data for NCEP Global Forecast System". EN. In: *Mon. Weather Rev.* 142.9, pp. 3303–3325. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-13-00303.1.

Wang, Xuguang, Craig H. Bishop, and Simon J. Julier (July 2004). "Which Is Better, an Ensemble of Positive–Negative Pairs or a Centered Spherical Simplex Ensemble?" en. In: *Mon. Wea. Rev.* 132.7, pp. 1590–1605. ISSN: 0027-0644. DOI: 10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2.

Wedi, Nils P. et al. (2020). "A Baseline for Global Weather and Climate Simulations at 1 Km Resolution". en. In: *J. Adv. Model. Earth Syst.* 12.11, e2020MS002192. ISSN: 1942-2466. DOI: 10.1029/2020MS002192.

Wergen, Werner and Michael Buchhold (2002). "Datenassimilation Für Das Globalmodell GME". In:

Whitaker, Jeffrey S. and Thomas M. Hamill (July 2002). "Ensemble Data Assimilation without Perturbed Observations". EN. In: *Mon. Weather Rev.* 130.7, pp. 1913–1924. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

– (Sept. 2012). "Evaluating Methods to Account for System Errors in Ensemble Data Assimilation". en. In: *Mon. Weather Rev.* 140.9, pp. 3078–3089. ISSN: 0027-0644, 1520-0493. DOI: 10.1175/MWR-D-11-00276.1.

Xia, Youlong et al. (Apr. 2019). "Regional and Global Land Data Assimilation Systems: Innovations, Challenges, and Prospects". en. In: *J Meteorol Res* 33.2, pp. 159–189. ISSN: 2198-0934. DOI: 10.1007/s13351-019-8172-4.

Yang, Shu-Chih et al. (2009). "Comparison of Ensemble-Based and Variational-Based Data Assimilation Schemes in a Quasi-Geostrophic Model". In: *Mon Weather Rev* 137, pp. 693–709.

Yang, Yin (Mar. 2020). "Ensemble Data Assimilation and Function Approximation: A Kernel View".

Zeng, Xubin et al. (July 2002). "Coupling of the Common Land Model to the NCAR Community Climate Model". In: *J. Climate* 15.14, pp. 1832–1854. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2002)015<1832:COTCLM>2.0.CO;2.

Zhang, Fuqing, Chris Snyder, and Juanzhen Sun (2004). "Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter". In: *Mon. Weather Rev.* 132.5, pp. 1238–1253.

Zhang, Shaoqing et al. (June 2020). "Coupled Data Assimilation and Parameter Estimation in Coupled Ocean–Atmosphere Models: A Review". en. In: *Clim Dyn* 54.11, pp. 5127–5144. ISSN: 1432-0894. DOI: 10.1007/s00382-020-05275-6.

Zupanski, Milija (June 2005). "Maximum Likelihood Ensemble Filter: Theoretical Aspects". EN. In: *Mon. Weather Rev.* 133.6, pp. 1710–1726. ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR2946.1.

This page intentionally left blank

List of Figures

List of Tables

This page intentionally left blank

A
Appendix

A.1 Notation

Vectors and Matrices

Symbol	Meaning
$\boldsymbol{w}$	A vector of weights
$\boldsymbol{x}$	A vector of model states
$\boldsymbol{y}$	A vector of observational states
$\boldsymbol{W}$	A column-wise matrix of weights for all ensemble members
$\boldsymbol{X}$	A column-wise matrix of model states for all ensemble members
$\boldsymbol{Y}$	A column-wise matrix of observational states for all ensemble members
$\delta\boldsymbol{w}$	A vector of weight perturbations compared to the ensemble mean
$\delta\boldsymbol{x}$	A vector of model state perturbations compared to the ensemble mean
$\delta\boldsymbol{y}$	A vector of obsevational state perturbations compared to the ensemble mean
$\delta\boldsymbol{W}$	A column-wise matrix of weight perturbations for all ensemble members
$\delta\boldsymbol{X}$	A column-wise matrix of model state perturbations for all ensemble members
$\delta\boldsymbol{Y}$	A column-wise matrix of obsevational state perturbations for all ensemble members
$\mathbf{C}$	A matrix with constant values
$\mathbf{B}$	An approximated and static covariance matrix, independent from the ensemble
$\mathbf{H}$	The linearized observation operator, mapping from model space to observaitonal space
$\mathbf{I}$	The identity matrix with its shape given by the context

$\mathbf{K}$	The Kalman gain
$\mathbf{M}$	The linearized model for one time step
$\mathbf{P}$	A covariance matrix in model space, often estimated with the ensemble
$\tilde{\mathbf{P}}$	A covariance matrix in weight space, often estimated for the ensemble
$\mathbf{R}$	The observational error covariance matrix
$\tilde{\mathbf{R}}$	The observational error covariance matrix, transformed into feature space
$\boldsymbol{\alpha}$	An auxillary vector or matrix from the representer theorem for kernels
$\boldsymbol{\epsilon}$	A vector-valued random variable in observational space
$\boldsymbol{\zeta}$	A vector-valued random variable in model space
$\boldsymbol{\theta}$	A vector of variational parameters in variational Bayes
$\boldsymbol{\lambda}$	A vector of ordered eigenvalues
$\boldsymbol{\phi}$	A vector of states in feature space
$\boldsymbol{\Phi}$	A column-wise matrix of state in feature space for all ensemble members
$\boldsymbol{\Sigma}$	An approximated covariance matrix for the posterior in variational Bayes
$\delta\boldsymbol{\phi}$	A vector of perturbations in feature space compared to the ensemble mean in feature space
$\delta\boldsymbol{\Phi}$	A column-wise matrix of perturbations in feature space for all ensemble members
$\mathbf{0}$	A vector of zeros with its shape given by the context
$\mathbf{1}$	A vector of ones with its shape given by the context

Other notations

Symbol	Meaning
$\boldsymbol{x}^{\mathrm{a}}$	An analysis / posterior vector
$\boldsymbol{x}^{\mathrm{b}}$	A background / prior vector
$\boldsymbol{x}^{\mathrm{o}}$	A vector of observations
$\boldsymbol{x}^{(\mathrm{i})}$	The i-th element of the matrix $\mathbf{X}$, often denoting the i-th ensemble member

$\mathbf{x}_t$	A vector valid at time t
$\mathbf{y}_{1:t}$	A vector with entries from time 1 to time t
$\mathbf{y}_{:t}$	A vector with entries from the beginning to time t
$H_t(\mathbf{x}_t)$	The observation operator valid at time t applied on the model state from the same time
$M_{t \mapsto t+1}(\mathbf{x}_t)$	The dynamical model mapping the model state from time t to time $t+1$
$\mathbf{Y}_t^T$	The adjoint of the dynamical model and observation operator, mapping from time t to time 0 in Chapter 5
$\varphi_t(\mathbf{y}_t)$	The feature extractor function valid at time t applied on the observational state from the same time
$f(\mathbf{y}_t)$	The inverse function applied on the observational state from time t
$\mathcal{L}(\mathbf{x})$	The variational data assimilation cost function in model space, evaluated with vector $\mathbf{x}$
$\widetilde{\mathcal{L}}(\boldsymbol{w})$	The variational data assimilation cost function in weight space, evaluated with vector $\boldsymbol{w}$
$J(\theta)$	The variational Bayes cost function with respect to parameters θ
$K(\mathbf{y}, \mathbf{y}')$	A matrix-valued kernel function, applied on $\mathbf{y}$ and $\mathbf{y}'$
$\widetilde{K}(\mathbf{y}, \mathbf{y}')$	A matrix-valued and centered kernel function, applied on $\mathbf{y}$ and $\mathbf{y}'$
$\mathbf{x}^T$	A transposed vector
$\mathbf{X}^T$	A transposed matrix
$\mathbf{x}^2$	A squared vector
$\mathbf{X}^{\frac{1}{2}}$	The square-root of a matrix
$\mathbf{X}^{-1}$	The inverse of a matrix
$\log \mathbf{X}$	The logarithm of a matrix
$\operatorname{tr} \mathbf{X}$	The trace of a matrix
$\overline{\mathbf{x}}$	The ensemble mean in the same space of the given vector
$\mathbb{E}_{p(\mathbf{x}_t)}(\mathbf{x})$	The expectation of a vector with respect to a probability density function
$\|\mathbf{x}\|$	The norm of a vector
$p(\mathbf{x}_t)$	The probability density function of a vector

$p(\mathbf{x}_t \mid \mathbf{y}_t)$	The probability density function of a vector conditioned on another vector
$q_\theta(\mathbf{x}_t)$	An approximated probability density function of a vector with variational parameters θ
$\mathcal{N}(\mu, \Sigma)$	The Gaussian distribution with μ as mean and Σ as covariance
$\mathrm{IG}(\alpha, \beta)$	The inverse gamma distribution with α and β as parameters
$D_{KL}(q \parallel p)$	The Kullback-Leibler divergence from probability density function q to probability density function p
$\mathbb{R}^n$	A n-dimensional real vector space
$\mathcal{H}$	A reproducing kernel Hilbert space
$\mathcal{O}(1)$	An order of magnitude

A.2 A recipe to implement the LETKF from Torch-Assimilate

In the following, I show how a generic data assimilation schema would work within the core, the interface, and the pipeline layer of my data assimilation environment "torch-assimilate" (see also Section 3.5.3), based on the implementation of the LETKF:

1. Read-in of output files that can be again used to restart the model into a dataset.
2. Transformation of the read-in dataset into a generic state format, needed for torch-assimilate, this will be then $\mathbf{X}_t^b$.
3. Read-in of observations that will be assimilated into an observational dataset, this will be then $\mathbf{y}_t^o$.
4. Read-in of the 2-metre-temperature output from COSMO into a pseudo observational dataset, which will then act as basis for the ensemble equivalent of the observations.
5. Transformation of the pseudo observational dataset into a valid state $\tilde{\mathbf{X}}_t^b$.
6. (Possible) pre-processing of all datasets that are used within the data assimilation.
7. Apply the observation operator to $\tilde{\mathbf{X}}_t^b$, resulting in $\mathbf{Y}_t^b = H(\tilde{\mathbf{X}}_t^b)$.
8. Estimate global quantities of the ensemble mean in observational space $\overline{\mathbf{y}}_t^b$, the mean innovations $\delta\mathbf{y}_t^o = \mathbf{y}_t^o - \overline{\mathbf{y}}_t^b$, and ensemble perturbations in

observational space $\delta \mathbf{Y}_t^b = \mathbf{Y}_t^b - \overline{\mathbf{y}}_t^b$.

9. Multiply the global innovations and ensemble perturbations with the inverse of the cholesky decomposition $\mathbf{R}^{-\frac{1}{2}}$ of the observational error covariance, resulting in $\widetilde{\delta \mathbf{y}}_t^o = \mathbf{R}^{-\frac{1}{2}} \delta \mathbf{y}_t^o$ and $\widetilde{\delta \mathbf{Y}}_t^b = \mathbf{R}^{-\frac{1}{2}} \delta \mathbf{Y}_t^b$.

The following steps are parallelized looped with Dask to gather localized weights for every grid point independently, here shown for the i-th grid point.

10. Based on the specified localization covariance functions, the current grid point and the observational positions, estimate the localization weights $\mathbf{l}_i$.

11. Multiply the globally normalized innovation and ensemble perturbations by the square-root of the localization weights to gather their localized equivalents $\widetilde{\delta \mathbf{y}}_{t,i}^o = \mathbf{l}_i^{\frac{1}{2}} \widetilde{\delta \mathbf{y}}_t^o$ and $\widetilde{\delta \mathbf{Y}}_{t,i}^b = \mathbf{l}_i^{\frac{1}{2}} \widetilde{\delta \mathbf{Y}}_t^b$.

12. Convert $\widetilde{\delta \mathbf{y}}_{t,i}^o$ and $\widetilde{\delta \mathbf{Y}}_{t,i}^b$ from Xarray DataArray to PyTorch Tensor.

The steps until this point were all data assimilation method-agnostic, the following few steps are specific for the implementation of the LETKF.

13. Calculate the dot product between the localized ensemble perturbations in observational space $\mathbf{K}_i = (\widetilde{\delta \mathbf{Y}}_{t,i}^b)^\intercal \widetilde{\delta \mathbf{Y}}_{t,i}^b$.

14. Perform an eigenvalue decomposition to $\mathbf{K}_i$, resulting into $\boldsymbol{\lambda}_i$ as vector of ordered eigenvalues and $\mathbf{V}_i$ as matrix of the corresponding eigenvectors.

15. Add $\frac{(k-1)}{\rho}$ to the eigenvalues $\boldsymbol{\lambda}_i$ as part of the prior covariance, with k the number of ensemble members and ρ as prior.

16. Estimate the posterior covariance in weight space as $\widetilde{\mathbf{P}}_{t,i}^a = \mathbf{V}_i \boldsymbol{\Lambda}_i^{-1} (\mathbf{V}_i)^\intercal$, where $\boldsymbol{\Lambda}_i^{-1}$ specifies a diagonal matrix with $\boldsymbol{\lambda}_i^{-1}$ as elements on the diagonal.

17. Calculate the dot product between the localized ensemble perturbations in observational space and the localized innovation $\mathbf{k}_{t,i}^o = (\widetilde{\delta \mathbf{Y}}_{t,i}^b)^\intercal \widetilde{\delta \mathbf{y}}_{t,i}^o$.

18. Estimate the mean weights as $\overline{\boldsymbol{w}}_{t,i}^a = \widetilde{\mathbf{P}}_{t,i}^a \mathbf{k}_i^o$.

19. Estimate the weight perturbations with the eigenvalues and eigenvectors as $\delta \boldsymbol{W}_{t,i}^a = \mathbf{V}_i [(k-1) * \boldsymbol{\lambda}_i]^{-\frac{1}{2}} (\mathbf{V}_i)^\intercal$.

The following steps are now again data assimilation method-agnostic such that they are applied in all implemented algorithms.

20. Add the mean weights column-wise to the weight perturbation to get the

combined weights $\boldsymbol{W}_{t,i} = \overline{w}^{a}_{t,i} + \delta \boldsymbol{W}^{a}_{t,i}$.

21. Conversion of $\boldsymbol{W}_{t,i}$ from PyTorch Tensor to Xarray DataArray.

22. Apply the ensemble weights to the ensemble perturbations in model space for the same grid point to get a localized posterior ensemble $\mathbf{X}^{a}_{t,i} = \overline{\mathbf{x}}^{b}_{t,i} + \delta \mathbf{X}^{b}_{t,i} \boldsymbol{W}_{t,i}$.

The steps afterwards are again done globally outside of the parallelized loop to get restart files for the model.

23. Collect the localized posterior ensemble into a global posterior ensemble $\mathbf{X}^{a}_{t}$.

24. (Possible) post-processing of the global posterior ensemble.

25. Transformation of the global posterior state into a valid restart dataset.

26. Writing of valid restart dataset to an analysis files from which the model could be restarted.

A.3 The iterative ensemble Kalman smoother

In the following, I describe my implementation of an iterative ensemble Kalman smoother (IEnKS) that follows closely the derivation of Bocquet and Sakov (2014). I iteratively optimize Eq. (5.5) with Eq. (5.6) in the IEnKS, here with a Gauss-Newton scheme. To compare the IEnKS to the ETKS, I use the transform version of the IEnKS, where the ensemble transformation (5.4) is applied after every iteration.

Similarly to the ETKS, I independently propagate every ensemble member into observational space based on the current solution of their weights $\mathbf{y}^{(i)}_{t} = H(M_{0 \mapsto t}(\overline{\mathbf{x}}^{b}_{0} + \delta \mathbf{X}^{b}_{0}(\mathbf{w} + \delta \mathbf{w}^{(i)})))$. The ensemble mean in observational space $\overline{\mathbf{y}}_{t} = \sum_{i=0}^{k} \mathbf{y}^{(i)}_{t}$ acts as approximated propagation of the current weight solution $H(M_{0 \mapsto t}(\overline{\mathbf{x}}^{b}_{0} + \delta \mathbf{X}^{b}_{0} \mathbf{w}))$. For the estimation of the approximated adjoint $\mathbf{Y}^{T}_{t}$, I use again the column-wise matrix of the ensemble perturbations in observational space $\delta \mathbf{Y}_{t}$, with $\delta \mathbf{y}^{(i)}_{t} = \mathbf{y}^{(i)}_{t} - \overline{\mathbf{y}}_{t}$ for the i-th column. In the IEnKS, I have now to account for changed state perturbations which were modified by $\mathbf{T}$,

$$\mathbf{Y}^{T}_{t} \approx (\mathbf{T}^{-1} \delta \mathbf{Y}_{t})^{T}. \tag{A.1}$$

Based on this approximated adjoint, I estimate the inverse of the updated ensemble covariance in ensemble space $[\tilde{\mathbf{P}}^{a}_{new}]^{-1}$, approximating the Hessian of $J(\mathbf{w})$ at the solution $\mathbf{w}$. I use an additional learning rate $0 < \tau \leqslant 1$, which can reduce the length of one update step to improve the convergence of the algorithm and to

stabilize the estimation of the Hessian (Khan et al., 2018),

$$[\tilde{\mathbf{P}}^{a}_{new}]^{-1} = (1-\tau)[\tilde{\mathbf{P}}^{a}]^{-1} + \tau \sum_{t=0}^{T} \mathbf{Y}_t^{T} \mathbf{R}_t^{-1} \mathbf{Y}_t, \tag{A.2}$$

$$\mathbf{w}_{new} = \mathbf{w} - \tau \tilde{\mathbf{P}}^{a}_{new} [(k-1)\mathbf{w} - \sum_{t=0}^{T} \mathbf{Y}_t^{T} \mathbf{R}_t^{-1} (\mathbf{y}_t^{o} - \overline{\mathbf{y}}_t)]. \tag{A.3}$$

The schema is iterated between the ensemble propagation and update of the weights until a fixed number of iterations N is reached. Afterwards, I propagate the estimated state again throughout the assimilation window as in the ETKS.

The computational costs for the IEnKS depend heavily on the number of update iterations N. I get as total computational costs for the IEnKS the following, $N \times (T \times P \times k + \Omega) + T \times P \times k$. This shows that the IEnKS is the most expensive algorithm in this thesis. Because of these computational costs, I will restrict in the following experiment the number of iterations for the IEnKS to two.

I use the same localization schema and radii as in the ETKS experiments in Chapter 5 based on the assumption that advection can be neglected in the case of land surface data assimilation. I multiplicatively inflate the ensemble before I estimate the ensemble weights with an inflation factor of $\gamma = 1.15$. In the propagation of the IEnKS, I use the background ensemble for all variables, except the soil moisture.

Result

In the following, I analyze the performance of our IEnKS implementation with two iterations. I compare the propagated solutions of the IEnKS to our LETKS (24 h) experiment (here, differently called LETKS Fwd 24h) for the soil moisture in root-depth as RMSE based on the nature run (Fig. A.1).

Although the IEnKS decreases the analysis error to the nature run compared to its background error, the LETKS experiment has a lower error than the IEnKS experiment, especially after 2015-08-02. This increased error is caused by the second loop of the IEnKS, which harms the performance of the IEnKS, especially after 2015-08-01 12:00 UTC. In the first analysis step at 2015-07-31 12:00 UTC, the IEnKS has a lowered error compared to the LETKS, indicating a possible gain by the IEnKS in sub-optimal cases, where the model and nature differ more. In my case, the problem of assimilating the 2-metre-temperature for the soil moisture seems to be linear enough such that the explicit linearization within the LETKS has almost no impact on the assimilation.

A.4 Centering in feature space

Here, I show how I can center the feature space solely based on kernels. This derivation closely follows Schölkopf et al., 1998. The solution for the kernelized

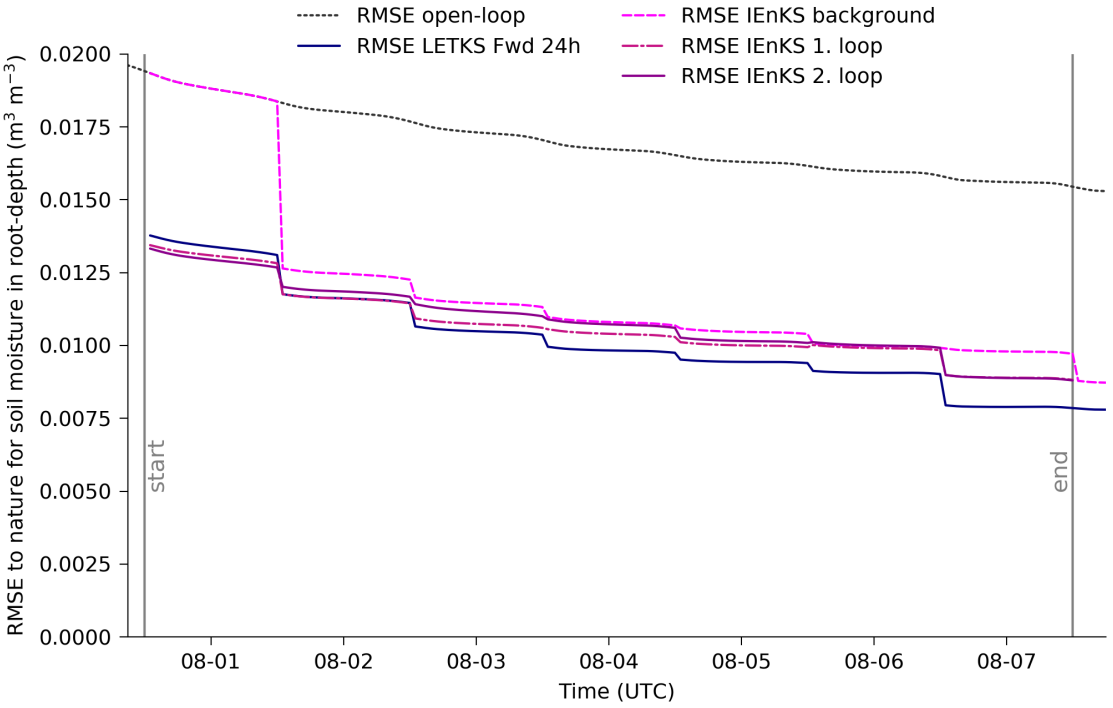


Figure A.1: Root-mean-squared-error of the Iterative Ensemble Kalman smoother compared to the nature within the simulation window as area average for the soil moisture in root-depth. The light-black dotted line and dark-blue line are the the open-loop run and the LETKS Fwd 24h experiment, respectively. The violet lines are different loops of the IEnKS with a 24 hour assimilation window. The first loop corresponds to the propagated state after the first update, whereas the second loop is the propagated state after the second and final update.

ETKF involves two centered kernel products $\widetilde{K}(\mathbf{Y}_t^b, \mathbf{Y}_t^b)$ and $\widetilde{K}(\mathbf{Y}_t^b, \mathbf{y}_t^o)$; both represent hereby a gram matrix and vector, respectively. In the following I derive the centered kernel for both products independently.

Let's start the derivation with a theorem of how the kernel matrix represents the feature space by data (the theorem closely follows Murphy, 2012). Following Mercer's theorem (Mercer and Forsyth, 1909), a positive-definite gram matrix $\mathbf{K}$ with its corresponding positive-definite kernel entries $k^{(i,j)} = K(\mathbf{y}_t^{(i)}, \mathbf{y}_t^{(j)})$ can be decomposed by eigendecomposition. The eigendecomposition results into a matrix with the eigenvectors $\mathbf{U}$ and a vector with the eigenvalues $\boldsymbol{\lambda}$, where $\boldsymbol{\Lambda}$ represent a diagonal matrix with the eigenvalues on its diagonal,

$$\mathbf{K} = \mathbf{U}^T \boldsymbol{\Lambda} \mathbf{U}.$$

As a consequence, I can represent the i,j-entry of this kernel matrix as product of the scaled eigenvectors $k^{(i,j)} = (\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{u}^{(i)})^T \boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{u}^{(i)}$. By denoting $\varphi_t(\mathbf{y}_t^{(i)}) = \boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{u}^{(i)}$, I can introduce the feature operator,

$$k^{(i,j)} = (\varphi_t(\mathbf{y}_t^{(i)}))^T \varphi_t(\mathbf{y}_t^{(j)}) = K(\mathbf{y}_t^{(i)}, \mathbf{y}_t^{(j)}).$$

This way, the entries of every positive-definite gram matrix can be represented by an inner product of feature vectors. These feature vectors are hereby implicitly

defined by the scaled eigenvectors of the gram matrix. Hence, every positive-definite kernel function has a corresponding feature vector that depends on the input data.

Now, I can take advantage of Mercer's theorem and derive the centered kernel functions. All kernel functions are centered by the prior ensemble mean in feature space. To simplify the derivation, I define $\mathbf{1}$ as all-ones vector, its length is given by the context, and $\boldsymbol{\Phi}_t^b$ as column-wise matrix, where the i-column is given as i-th prior ensemble member in feature space $\boldsymbol{\phi}_t^{b(i)} = \varphi_t(\mathbf{y}_t^{b(i)})$,

$$\begin{aligned}\overline{\boldsymbol{\phi}}_t^b &= n^{-1} \sum_{i=1}^{n} \varphi_t(\mathbf{y}_t^{b(i)}) \\ &= n^{-1} \mathbf{1} \boldsymbol{\Phi}_t^b \\ &= n^{-1} \mathbf{1} K(\mathbf{Y}_t^b, \cdot).\end{aligned}$$

The last equality results out of the reproducing property of kernels. I start with the derivation of the centering operation for $\widetilde{K}(\mathbf{y}_t^{b(i)}, \mathbf{y}_t^{b(j)})$. I additionally define $\delta\boldsymbol{\phi}_t^{b(i)} = \boldsymbol{\phi}_t^{b(i)} - \overline{\boldsymbol{\phi}}_t^b$ as i-th perturbation in feature space and $\mathbf{K}^b$ as gram matrix with $k^{b(i,j)} = K(\mathbf{y}_t^{b(i)}, \mathbf{y}_t^{b(j)})$ as i,j-th entry,

$$\begin{aligned}\widetilde{K}(\mathbf{y}_t^{b(i)}, \mathbf{y}_t^{b(j)}) &= (\delta\boldsymbol{\phi}_t^{b(i)})^\mathsf{T} \delta\boldsymbol{\phi}_t^{b(j)} \\ &= (\boldsymbol{\phi}_t^{b(i)} - \overline{\boldsymbol{\phi}}_t^b)^\mathsf{T} (\boldsymbol{\phi}_t^{b(j)} - \overline{\boldsymbol{\phi}}_t^b) \\ &= (\boldsymbol{\phi}_t^{b(i)})^\mathsf{T} \boldsymbol{\phi}_t^{b(j)} - (\boldsymbol{\phi}_t^{b(i)})^\mathsf{T} \overline{\boldsymbol{\phi}}_t^b - (\overline{\boldsymbol{\phi}}_t^b)^\mathsf{T} \boldsymbol{\phi}_t^{b(j)} + (\overline{\boldsymbol{\phi}}_t^b)^\mathsf{T} \overline{\boldsymbol{\phi}}_t^b \\ &= K(\mathbf{y}_t^{b(i)}, \mathbf{y}_t^{b(j)}) - n^{-1} K(\mathbf{y}_t^{b(i)}, \mathbf{Y}_t^b)\mathbf{1} - n^{-1}\mathbf{1}K(\mathbf{Y}_t^b, \mathbf{y}_t^{b(j)}) + n^{-2}\mathbf{1}K(\mathbf{Y}_t^b, \mathbf{Y}_t^b) \\ &= k^{b(i,j)} - n^{-1}\mathbf{k}^{b(i)}\mathbf{1} - n^{-1}\mathbf{1}\mathbf{k}^{b(j)} + n^{-2}\mathbf{1}\mathbf{K}^b\mathbf{1}.\end{aligned}$$

Hence, I can describe the centering of $\widetilde{K}(\mathbf{y}_t^{b(i)}, \mathbf{y}_t^{b(j)})$ solely based on kernels. I continue with the derivation of the centering operation for $\widetilde{K}(\mathbf{y}_t^{b(i)}, \mathbf{y}_t^{o})$. I additionally define $\delta\boldsymbol{\phi}_t^o = \boldsymbol{\phi}_t^o - \overline{\boldsymbol{\phi}}_t^b$ as observational perturbation in feature space and $\mathbf{k}^o$ as gram vector with $k^{o(i)} = K(\mathbf{y}_t^{b(i)}, \mathbf{y}_t^o)$ as i-th entry,

$$\begin{aligned}\widetilde{K}(\mathbf{y}_t^{b(i)}, \mathbf{y}_t^{o}) &= (\delta\boldsymbol{\phi}_t^{b(i)})^\mathsf{T} \delta\boldsymbol{\phi}_t^{o} \\ &= (\boldsymbol{\phi}_t^{b(i)} - \overline{\boldsymbol{\phi}}_t^b)^\mathsf{T} (\boldsymbol{\phi}_t^{o} - \overline{\boldsymbol{\phi}}_t^b) \\ &= (\boldsymbol{\phi}_t^{b(i)})^\mathsf{T} \boldsymbol{\phi}_t^{o} - (\boldsymbol{\phi}_t^{b(i)})^\mathsf{T} \overline{\boldsymbol{\phi}}_t^b - (\overline{\boldsymbol{\phi}}_t^b)^\mathsf{T} \boldsymbol{\phi}_t^{o} + (\overline{\boldsymbol{\phi}}_t^b)^\mathsf{T} \overline{\boldsymbol{\phi}}_t^b \\ &= K(\mathbf{y}_t^{b(i)}, \boldsymbol{\phi}_t^{o}) - n^{-1} K(\mathbf{y}_t^{b(i)}, \mathbf{Y}_t^b)\mathbf{1} - n^{-1}\mathbf{1}K(\mathbf{Y}_t^b, \boldsymbol{\phi}_t^{o}) + n^{-2}\mathbf{1}K(\mathbf{Y}_t^b, \mathbf{Y}_t^b) \\ &= k^{o(i)} - n^{-1}\mathbf{k}^{b(i)}\mathbf{1} - n^{-1}\mathbf{1}\mathbf{k}^{o} + n^{-2}\mathbf{1}\mathbf{K}^b\mathbf{1}.\end{aligned}$$

The centering of $\widetilde{K}(\mathbf{y}_t^{b(i)}, \mathbf{y}_t^{o})$ is here a combination of $\mathbf{k}^o$ and $\mathbf{K}^b$ as gram vector and matrix and, thus, solely based on kernels.

I have proved here that the centering operations for feature-based data assimilation can be described solely on kernels. As a consequence, I do not need access to an explicit feature function for the kernelized ETKF. The kernelized ETKF therefore relies completely on the properties of the chosen kernel.

A.5 The kernelized ETKF as kernel ridge regression

In the following, I show that the kernelized ETKF has the same solution as kernel ridge regression for data assimilation (Sætrom and Omre, 2011; Yang, 2020). As data basis for the kernel ridge regression, I independently transform every prior ensemble member from model space into observational space $\mathbf{y}_t^{b(i)} = H_t(\mathbf{x}_t^{b(i)})$. These ensemble members are then the inputs for my function during the training. As output, I simply use the ensemble members in model space. The function $f(\mathbf{y})$ is therefore trained on the training dataset $\mathcal{D} = \{(\mathbf{y}_t^{b(i)}, \mathbf{x}_t^{b(i)}, i = 1 : k\}$ with k samples.

To fit the inference function, I consider a regularized least-square loss and a RKHS $\mathcal{H}$ as hypothesis space. I am interested in the data assimilation increment $\Delta\mathbf{x}_t^a = \bar{\mathbf{x}}_t^a - \bar{\mathbf{x}}_t^b$. I therefore reduce the empirical risk with $f(\mathbf{y})$ to the ensemble perturbations $\delta\mathbf{x}_t^{b(i)} = \mathbf{x}_t^{b(i)} - \bar{\mathbf{x}}_t^b$ as difference of the ensemble members to the ensemble mean $\bar{\mathbf{x}}_t^b = \sum_{i=1}^k \mathbf{x}_t^{b(i)}$. With λ as regularization parameter, this results into

$$\underset{f\in\mathcal{H}}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^{k} \|\delta\mathbf{x}_t^{b(i)} - f(\mathbf{y}_t^{b(i)})\|^2 + \lambda\|f\|_{\mathcal{H}}^2. \tag{6.11}$$

The regularization controls the smoothness of the function $f(\mathbf{y}_t^{b(i)})$ and is connected to the strength of the prior belief (Kanagawa et al., 2018). The representer theorem (Kimeldorf and Wahba, 1970; Schölkopf et al., 2001) defines that the optimal function must have the following form,

$$f(\cdot) = \sum_{i=1}^{k} \alpha_i K(\mathbf{y}_t^{b(i)}, \cdot). \tag{A.4}$$

As a result from this representer theorem, I can reformulate (6.11) into

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}(\boldsymbol{\alpha})^T K(\mathbf{Y}_t^b, \mathbf{Y}_t^b)^2\boldsymbol{\alpha} - (\boldsymbol{\alpha})^T K(\mathbf{Y}_t^b, \mathbf{Y}_t^b)\delta\mathbf{X}_t^b + \frac{\lambda}{2}(\boldsymbol{\alpha})^T K(\mathbf{Y}_t^b, \mathbf{Y}_t^b)\boldsymbol{\alpha} \tag{A.5}$$

Set the gradient of (A.5) with respect to $\boldsymbol{\alpha}$ to 0, I get the following unique solution for $\boldsymbol{\alpha}$,

$$\boldsymbol{\alpha} = \delta\mathbf{X}_t^b[K(\mathbf{Y}_t^b, \mathbf{Y}_t^b) + \lambda\mathbf{I}]^{-1}. \tag{A.6}$$

From this solution, I can recover the kernelized ETKF solution by setting $\lambda = k - 1$ and using centered kernels. For the ensemble mean, the kernelized ETKF therefore solves a kernelized least-square problem, mapping from observational space into the space spanned by the ensemble perturbations.

A.6 The ETKF from the Variational free energy

Here, I prove that the ETKF is the optimal solution to the variational free energy in the linear-Gaussian case. In Section 6.2, I introduce the variational free energy as proxy for the reverse KL-divergence between the approximated posterior and the unknown but true posterior,

$$\begin{aligned} D_{KL}(q_\theta(x_t) \parallel p(x_t \mid y^o_{1:t})) \propto J(\theta) = &- \mathbb{E}_{q_\theta(x_t)} \log p(y^o_t \mid x_t) \\ &+ D_{KL}(q_\theta(x_t) \parallel p(x_t \mid y^o_{1:t-1})). \end{aligned} \tag{6.15}$$

In the ETKF, the analysis is estimated in weight space. In this weight space, the prior distribution is presumably given as Gaussian distribution $p(x_t \mid y^o_{1:t-1}) = \mathcal{N}(\mathbf{0}, (k-1)^{-1}\mathbf{I})$ with $\mathbf{0}$ as mean and $(k-1)^{-1}\mathbf{I}$ as covariance. Additionally, the posterior is also presumably a Gaussian distribution $q_\theta(x_t) = \mathcal{N}(w^a_t, \widetilde{P}^a_t)$ with w^a_t as mean and $\widetilde{P}^a_t$ as covariance. As for the LETKF, I express the covariance as symmetric square-root $\widetilde{P}^a_t = (k-1)^{-1}\delta W^a_t (\delta W^a_t)^T$ with δW^a_t as posterior perturbation in weight space. The mean and the perturbations are therefore the variational parameters that are optimized in the ETKF; in the following, I use for the approximated posterior $q_\theta(x_t)$ interchangeably $\widehat{q}$. With the reparametrization trick (6.16), I can sample from this posterior $w_t \sim \mathcal{N}(w^a_t, \widetilde{P}^a_t)$. Furthermore, I assume that these samples from the weight space can be translated into observational space by a linearized observation operator,

$$y_t = \overline{y}^b_t + \delta Y^b_t w_t. \tag{3.17}$$

As last assumption, the observational likelihood is Gaussian distributed $p(y^o_t \mid x_t) = \mathcal{N}(y^o_t - y_t, (\sigma^o)^2\mathbf{I})$ with $y^o_t - \overline{y}^b_t - \delta Y^b_t w_t$ as mean and $(\sigma^o)^2\mathbf{I}$ as covariance. In this assumption, I simplify for convenience that the covariance of the observational likelihood is a diagonal matrix with $(\sigma^o)^2$ on its diagonal. Nevertheless, the following proof is also valid for a full observational covariance $\mathbf{R}$ with cross-correlations. With these assumptions in mind, I prove that the ETKF equations (3.19), (3.20), and (3.23) are the optimal solution for the variational Bayes problem (6.15). This proof is inspired by the proof that a linearized variational autoencoder equals a probabilistic principal components analysis in Dai et al., 2017, Lemma 1.

In a first step, I reform the negative log-likelihood as expectation of the approximated posterior. Following the Gaussian distribution of the approximated posterior and the observational likelihood, the expectation of the negative log-likelihood results into

$$\begin{aligned} -\mathbb{E}_{\widehat{q}} \log p(y^o_t \mid x_t) \propto &\frac{1}{2} \mathbb{E}_{\widehat{q}}[(y^o_t - \overline{y}^b_t - \delta Y^b_t w_t)^T (\sigma^o)^{-2} \mathbf{I} (y^o_t - \overline{y}^b_t - \delta Y^b_t w_t)] + C \\ = &\frac{1}{2} (\sigma^o)^{-2} \mathbb{E}_{\widehat{q}}[(y^o_t - \overline{y}^b_t)^T (y^o_t - \overline{y}^b_t) - 2(y^o_t - \overline{y}^b_t)^T \delta Y^b_t w_t \\ &+ (\delta Y^b_t w_t)^T (\delta Y^b_t w_t)] + C. \end{aligned}$$

The constant C includes here all terms that are constant with respect to the optimized posterior. I can additionally state the Gaussian distribution of $\delta Y^b_t w^a_t$

with the Gaussian distribution of the approximated posterior $q_\theta(\mathbf{x}_t)$ and the linear observation operator $\delta \boldsymbol{Y}_t^b$,

$$\delta \boldsymbol{Y}_t^b \boldsymbol{w}_t \sim \mathcal{N}(\delta \boldsymbol{Y}_t^b \boldsymbol{w}_t^a, (k-1)^{-1} \delta \boldsymbol{Y}_t^b \delta \boldsymbol{W}_t^a (\delta \boldsymbol{Y}_t^b \delta \boldsymbol{W}_t^a)^\mathsf{T}).$$

This allows me to an analytical formulation of the negative observational log-likelihood,

$$\begin{aligned} -\mathbb{E}_{\hat{q}} \log p(\mathbf{y}_t^o \mid \mathbf{x}_t) \propto \frac{1}{2}(\sigma^o)^{-2}[&-2(\mathbf{y}_t^o - \overline{\mathbf{y}}_t^b)^\mathsf{T} \delta \boldsymbol{Y}_t^b \boldsymbol{w}_t^a \\ &+ (k-1)^{-1} \operatorname{tr}((\delta \boldsymbol{Y}_t^b \delta \boldsymbol{W}_t^a)^\mathsf{T} \delta \boldsymbol{Y}_t^b \delta \boldsymbol{W}_t^a) \\ &+ (\delta \boldsymbol{Y}_t^b \boldsymbol{w}_t^a)^\mathsf{T} \delta \boldsymbol{Y}_t^b \boldsymbol{w}_t^a] + C. \end{aligned} \tag{A.7}$$

In a second step, I reformulate the KL-divergence between the approximated posterior and the prior distribution in weight space,

$$\begin{aligned} D_{KL}(q_\theta(\mathbf{x}_t) \parallel p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}^o)) &= \frac{1}{2}[(k-1)(\boldsymbol{w}_t^a)^\mathsf{T} \boldsymbol{w}_t^a + (k-1)\operatorname{tr}(\widetilde{\boldsymbol{P}}_t^a) \\ &\quad - \log(\det(\widetilde{\boldsymbol{P}}_t^a))] + C \\ &= \frac{1}{2}[(k-1)(\boldsymbol{w}_t^a)^\mathsf{T} \boldsymbol{w}_t^a + \operatorname{tr}(\delta \boldsymbol{W}_t^a (\delta \boldsymbol{W}_t^a)^\mathsf{T}) \\ &\quad - \log(\det(\delta \boldsymbol{W}_t^a (\delta \boldsymbol{W}_t^a)^\mathsf{T}))] + C. \end{aligned} \tag{A.8}$$

To find its minimum, I have to take the derivative of $J(\theta)$ (6.15) with respect to the mean weight vector $\boldsymbol{w}_t^a$ and the weight perturbations $\delta \boldsymbol{W}_t^a$ and set this derivative to 0,

$$\frac{\partial J(\theta)}{\partial \boldsymbol{w}_t^a} = (\sigma^o)^{-2}[-(\delta \boldsymbol{Y}_t^b)^\mathsf{T}(\mathbf{y}_t^o - \overline{\mathbf{y}}_t^b) + (\delta \boldsymbol{Y}_t^b \boldsymbol{w}_t^a)^\mathsf{T} \delta \boldsymbol{Y}_t^b] + [(k-1)(\boldsymbol{w}_t^a)^\mathsf{T}] = 0, \tag{A.9}$$

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \delta \boldsymbol{W}_t^a} &= (\sigma^o)^{-2}[(k-1)^{-1}(\delta \boldsymbol{Y}_t^b \delta \boldsymbol{W}_t^a)^\mathsf{T} \delta \boldsymbol{Y}_t^b] + [\delta \boldsymbol{W}_t^a - ((\delta \boldsymbol{W}_t^a)^{-1})^\mathsf{T}] \\ &= (k-1)^{-1}[(k-1)\mathbf{I} + (\sigma^o)^{-2}(\delta \boldsymbol{Y}_t^b)^\mathsf{T} \delta \boldsymbol{Y}_t^b] \delta \boldsymbol{W}_t^a (\delta \boldsymbol{W}_t^a)^\mathsf{T} = 0. \end{aligned} \tag{A.10}$$

The optimal solution for the mean weights can be derived from (A.9), whereas the solution for the weight perturbations results from (A.10),

$$\boldsymbol{w}_t^a = \sigma^{-2}[(k-1)\mathbf{I} + \sigma^{-2}(\delta \boldsymbol{Y}_t^b)^\mathsf{T} \delta \boldsymbol{Y}_t^b]^{-1}(\delta \boldsymbol{Y}_t^b)^\mathsf{T}(\mathbf{y}_t^o - \overline{\mathbf{y}}_t^b), \tag{A.11}$$

$$\delta \boldsymbol{W}_t^a = \sqrt{(k-1)}[(k-1)\mathbf{I} + \sigma^{-2}(\delta \boldsymbol{Y}_t^b)^\mathsf{T} \delta \boldsymbol{Y}_t^b]^{-\frac{1}{2}}. \tag{A.12}$$

These solutions for $\boldsymbol{w}_t^a$ and $\delta \boldsymbol{W}_t^a$ correspond to (3.19) and (3.23) in the case of a diagonal observational covariance matrix. Under the previously stated assumptions of Gaussian distributions and a linearized observation operator, the variational free energy (6.15) has (A.11) and (A.12) as unique solutions. These unique solutions are therefore the global optimum. This completes the proof. □

This shows that the ETKF is the global optimum of the variational free energy in the Gaussian-linearized case. The ETKF is the linearized version of the MLEF/IEnKF (Zupanski, 2005; Sakov et al., 2012) if only observations at the same time as the update step are considered. This therefore also means that the MLEF reduces the variational free energy in the case of a non-linear observation operator and a Gaussian assumption in weight and observational space. In this generalized case, the variational free energy can have multiple minima. As a consequence, the solution of the MLEF is not unique. Nevertheless, this relates the MLEF to the Variational Online-Newton method from Khan et al., 2018, Appendix D.

This page intentionally left blank

Danksagung

An erster Stelle möchte ich meiner Freundin Clara danken. Diese hat mir, auch in schwierigen Zeiten, den Rücken gestärkt, stand zu mir, hat aber auch mal ihre Meinung kund getan, wenn einiges nicht so lief, wie es sollte. Natürlich möchte ich mich an dieser Stelle auch bei meinen beiden Betreuern Felix Ament und Gernot Geppert bedanken. Diese haben durch, teilweise doch etwas längliche, aber immer konstruktive Diskussionen mit dazu beigetragen, was hier auf Papier verewigt worden ist.

Ich möchte auch meinem Büronachbarn Finn danken, dass er es mit mir ausgehalten hat und immer für ein Käffchen mit wissenschaftlicher Diskussion zu haben war. Das hat mir vorallem in Zeiten der Corona-Pandemie sehr geholfen. Hierbei sind wir auch schon beim richtigen Thema, denn Corona hat auch eine gute Sache hervor gebracht, dass sich eine kleinere Runde gebildet hat, bestehend aus: Heike, Basti, Henning und Jule. Diese Runde war jeden Morgen ein Fixpunkt im Kalender, der dazu begetragen hat die doch schwierige Zeit erfolgreich zu überstehen. Zusätzlich möchte ich auch den weiteren Arbeitsgruppen-Mitgliedern danken. Diese haben durch das freundliche Aufnehmen vor sechs Jahren mit dafür gesorgt, dass die Arbeit entstanden ist. Vorallem möchte ich hier Akio hervorheben, mit dem ich 5 Jahre lang „NinJo 2.0“ als interaktive Vorlesung geleitet habe und der immer wieder hilfreiche Antworten parat hatte.
Aus einem wissenschaftlichen Gesichtspunkt möchte ich auch den Mitgliedern der DFG-Forschergruppe 2131 danken. Durch ihre Mithilfe und Diskussion sind einige der Ergebnisse in dieser Arbeit verbessert worden. Hierbei ist auch noch Clemens Simmer hervorzuheben. Dieser hat die Arbeit erst dadurch ermöglicht, dass ich 3 Jahre lang bei der Universität Bonn innerhalb der Forschergruppe angestellt war. An dieser Stelle möchte ich auch Johanna Baehr danken, dass diese als Moderation meines Doktorpanels jedes halbe Jahr kritisch nachgefragt hat und dabei auch hartnäckig geblieben ist. Ein weiter Dank geht an die Mitglieder des IMPRS-Office, diese waren immer für organisatorische Fragen da und haben einem als Doktorand das Leben vereinfacht.

Ich möchte auch meiner Doppelkopf-Gruppe danken, unter anderem bestehend aus: Laura, Fabi, Corinna und Toni. Es hat mit denen immer wieder Spaß gemacht, die Sorgen des Doktorandenalltags, bei ausgiebigen Runden, zu vergessen. Hierbei sind auch noch ehemalige Kommolitonen und weitere Freunde zu nennen, die trotz teilweise etwas größerer Entfernung immer wieder für Späße gut waren.

Zu guter Letzt möchte ich natürlich auch meiner Familie, im speziellen meiner Mutter, meinem Vater und meinen Großeltern, danken, ohne diese wäre ich wahrscheinlich nicht dort, wo ich Heute bin. Deshalb würde es ohne diese die vorliegende Arbeit auch so nicht geben.

Hinweis / Reference

Die gesamten Veröffentlichungen in der Publikationsreihe des MPI-M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

sind über die Internetseiten des Max-Planck-Instituts für Meteorologie erhältlich:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html

All the publications in the series of the MPI -M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

are available on the website of the Max Planck Institute for Meteorology:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html