Data assimilation with autocorrelated model error

Author:

Haonan Ren

Supervisor:

Prof. Peter Jan VAN LEEUWEN, Dr. Javier AMEZCUA

A thesis submitted for the degree of Doctor of Philosophy

NOV 2022

Declaration of Authorship

I, Haonan Ren, declare that this thesis titled, "Data assimilation with autocorrelated model error" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at University of Reading.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: *Haonan Ren*

_____

Date: April 27, 2023

_____

Acknowledgements

The completion of this study could not have been achieved without the expertise of Prof. Peter Jan Van Leeuwen and Dr. Javier Amezcua. I would like to express my thanks and gratitude to the examiners for sitting on this panel and taking time to read the thesis. I also could not have undertaken this journey without my monitor committee which is consisted of by Prof. Sarah Dance and Prof. Remi Tailleux, who generously provided knowledge and expertise.

I am also grateful to my classmates and cohort members, especially my office mates, for their help, such as late-night feedback sessions, and moral support. Thanks should also go to the librarians, research assistants, and study participants from the university, who impacted and inspired me.

Lastly, I would be remiss in not mentioning my family, especially my parents. Their belief in me has kept my spirits and motivation high during this process. I would also like to thank the nurses and doctors from the Royal Berkshire Hospital. They have been working hard to keep me alive and healthy for the past 2 years.

PUBLICATIONS

Chapter 3 and 4 of this thesis are reproduced from the following publications respectively:

1. Ren, Haonan, Javier Amezcua, and Peter Jan Van Leeuwen (2021). "Effects of mis-specified time-correlated model error in the (ensemble) Kalman Smoother". *Quarterly Journal of the Royal Meteorological Society* 147.734, pp. 573–588.

2. Amezcua, J, Ren, H and van Leeuwen, PJ. 2023. Using the (Iterative) Ensemble Kalman Smoother to Estimate the Time Correlation in Model Error. *Tellus A: Dynamic Meteorology and Oceanography*, 75(1), 108–128.

All work undertaken in the first paper was carried out by Haonan Ren with co-authors providing guidance and review. All experiments in the second paper were performed by Haonan Ren with co-authors' guidance, and the paper was written mainly by Dr. Javier Amezcua.

Abstract

Data assimilation has often been performed under the perfect model assumption, but in reality, numerical models often contain model errors with spatial and temporal correlations. The objective of this thesis is to thoroughly investigate the impact of an inaccurate time correlation in the model error description on data assimilation results, both analytically and numerically using the ensemble Kalman Smoother (EnKS). Furthermore, we try to develop an efficient way to perform online estimation of certain model error autocorrelation parameters with the data assimilation scheme.

With a simple linear model and a single-parameter autocorrelation, we find that the performance of the data assimilation scheme can be impacted by the departures between the actual values of the parameter and the value proposed in the data assimilation process with sparse observations. However, the impact of the incorrect parameter can be diminished with dense observations. Furthermore, we show that the correct model error decorrelation timescale can be estimated after multiple simulation windows using the state augmentation method with the linear system.

More complex autocorrelation, in which decaying and oscillatory scales are considered, is later examined on the linear model and, furthermore, a nonlinear logistic map. It seems impossible for the EnKS to track both decaying and oscillatory parameters in the autocorrelation, and the iterative variant of the EnKS (IEnKS) is required. With the nonlinear logistic map, even the IEnKS fails to find the correct values for the two parameters and can get stuck in local minima. Fortunately, we can find the correct values of the parameters with careful tuning of the IEnKS and transformation of the solution space.

When the problem confronts a high-dimensional nonlinear system such as the quasi-geostrophic model, a large part of the state has to be observed in space, even for the

simplest case of the model error autocorrelation. In this case, it shows the limitations of our method for practical weather forecast systems since the observation cannot be as dense as needed for the parameter estimation to work, and result in an affordable scheme.

Contents

Bibliography   119

Chapter 1

Introduction

1.1 Motivation

With a dense population distributed in major cities, weather forecasts to prevent the damage of high-impact weather are increasingly more critical. New studies (such as Hanlon et al., 2021) have established that high-impact weather, such as extremely hot days, heavy rainfall, floods and wildfires, could become more frequent and intense based on the current global warming circumstance and its variability. According to the Fifth Assessment Report published in 2014 by the Intergovernmental Panel on Climate Change (IPCC) (Pachauri et al., 2014), the economic losses associated with extreme weather will continue to increase. Besides, extreme weather can also easily affect human well-being and health. A disaster-related to weather, climate or water hazard occurred every day on average over the past 50 years, killing 115 people daily (Ebi et al., 2021). Therefore, highly accurate weather forecasts are needed more than ever to prevent the loss of life and properties (see e.g. Bauer, Thorpe, and Brunet, 2015).

There are 3 main reasons that weather forecast is not and can never be exact: measured observation uncertainty, chaotic behaviour of nature and model limitations. The observational data of the numerical forecast prediction (NWP) system comes from different sources, terrestrial and space based. There is no specific way to know precisely

what is happening at present by analyzing the data, but the large-scale patterns can be defined with uncertainties. Another factor that diminishes the accuracy of weather forecast is the chaotic behaviour of the atmosphere. It has been suggested that a butterfly flapping its wings in Tokyo can create a storm in New York because small weather features can develop into large-scale events over a period of a few days (Lorenz, 1972). Forecast skills will be appreciable, although forecasts will only be somewhat reliable. By a week or so skill will be less and forecasts rather hit and miss, and skill will decrease to zero after 15 days (Toth and Buizza, 2019). This heavily depends on the variable. At last, the NWP system also has its limitations for various reasons, such as limited computational power and lack of knowledge of atmospheric and oceanic physics. Imperfect models create errors that can feed into the weather forecast results. Studies on predictability in forecasting models have shown that model error has indeed significant impact on the accuracy of weather forecast (Dalcher and Kalnay, 1987; Palmer, 2019; Alley, Emanuel, and Zhang, 2019; Krishnamurthy, 2019). However, due to various sources of the model error and its increase over time, measuring the model error is impossible for the present skills. More efforts to improve the forecast accuracy focus on state variable estimation through assimilation of the observation and ensemble forecasting (e.g. Anderson, 2001; Whitaker and Hamill, 2002). Ideally, forecast model deficiencies should be addressed by, for instance, improving the physical parameterizations (e.g. Couvreux et al., 2021) or applying higher resolution systems to resolve small-scale processes (e.g. Tonani et al., 2019; Mukhopadhyay et al., 2019). But even current state-of-the-art parameterization schemes and high-resolution numerical system cannot remove all the model errors in our forecast models (see in e.g. Zheng et al., 2016; Comin et al., 2018). Therefore, representing the model error in the weather forecasting system can be a more direct solution to improving the forecast accuracy.

Instead of neglecting the model error, one possible way to relax the strong-constraint

setting is to treat the model error as a random or an uncorrelated white noise with prescribed error covariances (such as in Heemink, Verlaan, and Segers, 2001). However, in reality, the model errors are often spatially and temporally correlated. Another approach is using data assimilation schemes with parameter estimation to estimate the model error or at least certain characteristics of the model error. Some research has shown that with the updated model error characteristics, data assimilation can provide significantly improved results (see discussions in Amezcua and Van Leeuwen, 2018; Ren, Amezcua, and Van Leeuwen, 2021). In this thesis, we investigate the impact of the autocorrelated model error on the data assimilation results, more specifically the impact of misspecified autocorrelation of the model error. Furthermore, we try to find a way to estimate certain parameters in the model error autocorrelation online in order to update the model error and improve the data assimilation results with the ensemble Kalman Smoother (EnKS) and its iterative variant (IEnKS).

1.2 Thesis aim

The objective of this thesis is to investigate the impact of the model error, more specifically, temporal auto-correlated model error, on data assimilation results and find a way to update the model error online within the data assimilation procedure. We propose different formulations of the model error autocorrelations, and examine the results with different models, both linear and nonlinear.

Then, we have the following research questions that need to be answered:

1. **What impact does the temporal autocorrelated model error have on the data assimilation results?**

   Specifically, we investigate:

- What is the impact of the model error characteristic on the ensemble spread\variance and RMSE?

- How sensitive are data assimilation results to observation density?

- How would the misspecified autocorrelation of the model error impact the performance of the data assimilation scheme?

2. **Is it possible to establish a way to update the parameters encoded in the model error autocorrelation online during the data assimilation procedure?**

   While the temporal autocorrelation of the model error is considered, we want to implement different formulations of the model error autocorrelations and examine them on different models, both linear and nonlinear. Specifically, we will look at:

   - How sensitive are the parameter estimation results to different model error formulations?

   - Do we need a more efficient data assimilation scheme for the parameter estimation in a nonlinear system instead of the EnKS?

   - How many observations are needed to obtain the optimal results for parameter estimation?

   - Furthermore, what impact of the initial guess will the parameters have on the parameter estimation results?

3. **How does our estimation method perform in more complex and higher-dimensional systems, such as the Quasi-geostrophic (QG) model?**

   The final goal of our project is to implement our method in a more realistic system, the QG model. We further investigate:

   - With the higher-dimensional and more complex system, would the same method and data assimilation scheme works for parameter estimation?

   - If the original settings fail, what kind of tuning is needed to perform successful parameter estimation with complex models? Or is there an alternative method

we can use to achieve our goal?

1.3 Principal new results

The principal new results of this thesis (numbered according to the order of chapters) are:

1. The misspecified model error autocorrelation can significantly affect the performance of the data assimilation scheme. We find that to have a fair estimation of the parameters encoded in the model error, frequent observations are required.

2. Estimating model error autocorrelation parameters is indeed a difficult task, even within a low-dimensional system. However, practical solutions can be found via a careful reformulation of the problem.

3. Parameter estimation using the state augmentation method on the high-dimensional nonlinear system is a much more complex problem compared with state variable update. We need much more dense observations in space to produce satisfactory parameter estimation that converges to the correct value of the parameter.

1.4 Outline of the thesis

The general structure of this thesis is as follows:

- **Chapter 2** generally introduces different data assimilation methods, and the general ideas of various data assimilation approaches. We also highlight the importance of including model error in the NWP systems and data assimilation process. Furthermore, the treatments of the model error in the data assimilation schemes in the past years are also discussed. Finally, we review the parameter estimation methods applied in the data assimilation schemes in the past years.

- **Chapter 3** starts with the formulation of the ensemble Kalman smoother (EnKS) we use in our experiments, and describes the model error autocorrelation structure. In detail, we examine the impact of the model error autocorrelation on the ensemble variance and mean-square error (MSE). Then, we move to a higher dimensional linear system and evaluate the performance of the EnKS with the autocorrelated model error by computing the ratio of root-mean-square error over the ensemble spread. The final part of the chapter is to perform parameter estimation of the decaying scale encoded in the autocorrelation. This chapter has been published as Ren, Amezcua, and Van Leeuwen (2021).

- **Chapter 4** further investigates the possibilities of performing online parameter estimation for the model error. After the results on a linear system with simple decaying autocorrelation of the model error, we extend our experiments to a highly nonlinear system, the logistic map. We also propose a different structure of the model error autocorrelation, including both decaying and oscillatory scales. First, we experiment with different formulations of model error autocorrelations on the simple linear model we used in **Chapter 3** using the EnKS and its iterative variant (IEnKS). Then, we perform the parameter estimation on the logistic map using the IEnKS with different tunning settings. This chapter has been recently published (Amezcua, Ren, and Van Leeuwen, 2023).

- **Chapter 5** introduces a more complex and realistic model, the quasi-geostrophic model, for our experiments. As mentioned in **Chapter 3**, the ultimate goal is to update the model error in the operational forecasting system. Since the results on both linear and nonlinear systems are quite promising, we extend our experiments to a quasi-geostrophic model. We start with the formulation of the 2-layer QG model and how we solve the model numerically. Then, we introduce an ensemble smoother in ensemble subspace to simulate more observations with less computational power. We offer the results for the state update on the QG model. Furthermore, we investigate the performance of the smoother. Then, we show the results

of parameter estimation using state augmentation with the ensemble smoother in ensemble subspace for the simplest case in the model error autocorrelation.

- **Chapter 6** summarises the results of this thesis regarding the research questions presented in **Chapter 1**. We also propose the potential aspects of future work on autocorrelated model error estimation in data assimilation schemes.

Chapter 2

Review of data assimilation methods and model error treatment

2.1 Data assimilation methods

Data assimilation combines prior information provided by numerical model simulations with observations from reality to produce the best possible description of the system of interest and its uncertainty (see in Evensen, Vossepoel, and Van Leeuwen, 2022). The general solution of the data assimilation problem is given by Bayes' theorem (Bayes, 1763):

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \tag{2.1}$$

where $\mathbf{x}$ represents the state of the system, and $\mathbf{y}$ denotes the observations. From Bayes' theorem, the probability density function (pdf) of the state variable given observations $p(\mathbf{x}|\mathbf{y})$ can be obtained knowing the pdf of the state variable $p(\mathbf{x})$, the conditional pdf of the observations given the current state of the system, and the marginal pdf of the observations, which is independent of state variables.

Various approaches in the field of data assimilation have been developed in the past years, such as the classic variational approach (Courtier and Talagrand, 1987), the statistical approach based on the Kalman filter formulation (Asch, Bocquet, and Nodet, 2016),

and new approaches such as the hybrid method (e.g. Hamill and Snyder, 2000; Lorenc et al., 2015; Goodliff, Amezcua, and Van Leeuwen, 2015) which tries to combine the benefits of the two classic approaches. Variational methods try to find the mode of the posterior pdf, assuming a Gaussian prior and Gaussian observation error. The maximum of the posterior pdf $p(\mathbf{x}|\mathbf{y})$ appears at the exact location in state space as the minimum of the cost function, defined as:

$$\mathcal{J}(\mathbf{x}) = -\log p(\mathbf{x}|\mathbf{y}). \tag{2.2}$$

Kalman-Filter-like methods, which assume a Gaussian prior, Gaussian observation error, and a linear observation operator, are popular and well-examined as members of the statistical data assimilation approaches. Then the solution can be found analytically using linear algebra methods. Ensemble variants have been developed to avoid having to store and propagate the large covariance matrix that is explicitly needed in a Kalman Filter (e.g. Evensen, 1994; Evensen and Van Leeuwen, 1996). Since both variational and Ensemble Kalman-Filters have their merits, hybrids have been formulated that try to combine the best of each methodology. Finally, fully nonlinear data-assimilation methods have been developed based on particle filters (such as in Gordon, Salmond, and Smith, 1993; Isard and Blake, 1998; Van Leeuwen et al., 2019) and particle flow filters (Daum and Huang, 2009; Daum, Huang, and Noushin, 2010; Hu and Van Leeuwen, 2021; Lucini, Van Leeuwen, and Pulido, 2021) that are rapidly approaching maturity (Van Leeuwen et al., 2019). In the following subsections, we discuss the two classic data assimilation methods in detail.

2.1.1 Variational data assimilation

Variational data assimilation, which is based on optimal control theory and was formally introduced by the meteorology community to treat the errors of the NWP system, aims

at minimizing a cost function that measures the departure between the model output and the measurement. As its representatives, 3-dimensional and 4-dimensional variational methods (3D-Var and 4D-Var) have been implemented in operational forecasting systems in the past years. Observations are collected at the analysis time in 3D-Var (see discussions in Courtier et al., 1998; Mazzarella et al., 2017). The 3D-Var tries to minimize the following cost function shown in Eq. (2.2) which can be reformulated as:

$$\mathcal{J}(\mathbf{x}) \propto -\log p(\mathbf{x}) - \log p(\mathbf{y}|\mathbf{x}), \tag{2.3}$$

in which $p(\mathbf{x})$ comes from the prior and $p(\mathbf{y}|\mathbf{x})$ from the likelihood. Then, with the Gaussian assumption for the prior and observation error, we have

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}), \tag{2.4}$$

where $\mathbf{x}, \mathbf{x}^b$ and $\mathbf{y}$ are the state of the system, prior/background and the measured state. $\mathbf{B}$ and $\mathbf{R}$ are the background error and observation error covariance matrices, and $\mathbf{H}$ represents the observation operator which maps the state vector into the observation space. The minimization of the cost function in Eq. (2.4) requires the gradient with respect to $\mathbf{x}$ to vanish:

$$\frac{\partial \mathcal{J}(x)}{\partial x} = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}) = 0. \tag{2.5}$$

The solution of Eq. 2.5 is given by (Lorenc, 1986):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b), \tag{2.6}$$

where $\mathbf{K}$ is the optimal gain matrix and can be computed as:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}. \tag{2.7}$$

When observation operator $\mathbf{H}$ is nonlinear, a Gauss-Newton method is typically used to find the state that minimizes the cost function (Gratton, Lawless, and Nichols, 2007), which consists of iteratively solving a set of linear problems such, and each linear problem is itself solved iteratively. A major disadvantage of 3D-Var is that the prior error covariance $\mathbf{B}$ is not evolving in time, hampering the correct spreading of information from the observations over the model domain.

As an extension of 3D-Var, in 4D-Var, the observations are assimilated at the exact time of measurement (Mazzarella et al., 2017). The analysis increment is propagated over the simulation period using tangent linear and adjoint models (Errico, 1997; Errico, Vukićević, and Raeder, 1993) which requires much higher computational power than 3D-Var. The cost function that 4D-Var tries to minimize is different from that in 3D-Var since the time index is taken into account:

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x_0} - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x_0} - \mathbf{x}_0^b) + \frac{1}{2}\sum_{l=0}^{L}(\mathbf{H}_l\mathbf{x}_0 - \mathbf{y}_l)^T \mathbf{R}_l^{-1}(\mathbf{H}_l\mathbf{x}_0 - \mathbf{y}_l), \quad (2.8)$$

in which

$$\mathbf{H}_l\mathbf{x}_0 = \tilde{\mathbf{H}}_l(\mathbf{M}_{0\to l}\mathbf{x}_0),$$

where $\mathbf{M}_{0\to l}$ denotes the nonlinear forward model and $\tilde{\mathbf{H}}$ is the observation operator for time l. Compared with the 3D-Var method, 4D-Var data assimilation requires higher programming power for the operational systems. Furthermore, the observation operator $\mathbf{H}_l$ is always nonlinear because it now contains the nonlinear forward model. However, 4DVar allows the time history of observations to be simulated fully to update the forecast. A weakness is the use of a fixed $\mathbf{B}$ matrix, which is one of the main reasons hybrid methods were put forward (Bonavita et al., 2016).

2.1.2 Kalman-Filter-based data assimilation

The Kalman Filter was formally introduced by Kalman (1960). The Kalman-Filter-based data assimilation methods either make Gaussian assumption on the pdf's or is based on the Best Linear Unbiased Estimator (BLUE) from estimation theory. It provides the optimal solution (minimum variance) to the data assimilation problem in the case of linear dynamics and Gaussian error assumption at the time observations occur. The Kalman filter algorithm contains two steps: the prediction of the system and the update of the prior. The prediction/prior is produced by a stochastic and linear system:

$$\mathbf{x}_{t+1}^b = \mathbf{M}_{t+1}\mathbf{x}_t^b, \tag{2.9}$$

where $\mathbf{M}_t$ is the model operator to propagate the model from time-step $t-1$ to time-step t and the covariance is propagated from t to $t+1$ via:

$$\mathbf{B}_{t+1} = \mathbf{M}_{t+1}\mathbf{B}_t\mathbf{M}_{t+1}^T + \mathbf{Q}, \tag{2.10}$$

in which $\mathbf{Q}$ is the model error covariance. An observation times Bayes' theorem is used with Gaussian prior and Gaussian observation errors, and $\mathbf{H}$ is assumed linear. With these assumptions, the posterior is also Gaussian, and the mean and the mode coincide. Using the same formulation with the cost fucntion we recover in the previous subsection, but in the Kalman Filter we explicitly solve this system, so we find:

$$\delta\mathbf{x}_t^a = (\mathbf{B}_t^{-1} + \mathbf{H}_t^T\mathbf{R}_t^{-1}\mathbf{H}_t)^{-1}\mathbf{H}_t^T\mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^b). \tag{2.11}$$

In the Kalman filter we rewrite this as

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^b), \tag{2.12}$$

where $\mathbf{K}_t$ is the Kalman Gain at given time-step t where the observations are available:

$$\mathbf{K}_t = \mathbf{B}_t \mathbf{H}_t^T (\mathbf{HB}_t \mathbf{H}_t + \mathbf{R}_t)^{-1}. \tag{2.13}$$

The prior covariance matrix is also updated by the Kalman Gain as:

$$\mathbf{A}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{B}_t. \tag{2.14}$$

Different from the Kalman Filter, the Kalman Smoother updates a whole simulation window instead of the observational time-step as the Kalman Filter, using all available observations from the past, present and possibly the future (see Asch, Bocquet, and Nodet, 2016). Since time as an extra factor is considered in the smoother, the formulation of the update equation of the smoother is slightly different from the update equation of the Kalman filter shown in Eq. (2.12):

$$\mathbf{x}_{0:\tau}^a = \mathbf{x}_{0:\tau}^b + \mathbf{K}_{0:\tau}(\mathbf{y}_{1:L} - \mathbf{H}_{1:L}\mathbf{x}_{0:\tau}^b), \tag{2.15}$$

where the state variable $\mathbf{x}_{0:\tau}^a$ and prior estimation $\mathbf{x}_{0:\tau}^a$ are collections of time series from the initial state to the end of the simulation window ($t = \tau$), L represents the number of observations in the simulation period, and $\mathbf{K}_{0:\tau}^a$ is now a gain over the whole simulation period:

$$\mathbf{K}_{0:\tau} = \mathbf{B}_{0:\tau} \mathbf{H}_{1:L}^T (\mathbf{H}_{1:L} \mathbf{B}_{0:\tau} \mathbf{H}_{1:L} + \mathbf{R}_{1:L})^{-1}. \tag{2.16}$$

The Kalman Smoother requires more computational power than the Kalman Filter since the state over the whole assimilation window is updated. According to the update equations of the Kalman Filter and Smoother, both methods require linear model and observation operators, which means they are not applicable for nonlinear systems such as the

numerical forecasting systems. The Extended Kalman Filter and Smoother (e.g. Jazwinski, 2007) can cope with nonlinear operators but require the computation of the Jacobian of the nonlinar model. Our project involves with both linear with linear and nonlinear systems. For the linear part, we apply the Kalman Smoother as shown in chapter 3. But for more complex circumstances, the Monte-Carlo formulations of the Kalman Smoother (EnKS) (Van Leeuwen and Evensen, 1996) is extensively used in both chapter 3 and chapter 4 and will be discussed in the following subsection.

2.1.2.1 Ensemble Kalman Filters

The KF-based ensemble methods such as the Ensemble Kalman Filter (EnKF) (Evensen, 1994; Houtekamer and Mitchell, 1998) are Monte-Carlo approximations of the Kalman filter and proposed to avoid the difficulties of storing and propagating the error covariance of the state. In the case of the EnKF, the covariance matrix is replaced by the sample covariance matrix computed from the ensemble members (see e.g. Evensen, 2009; Asch, Bocquet, and Nodet, 2016; Evensen, Vossepoel, and Van Leeuwen, 2022). The ensemble methods also fix part of the problem of the inapplicability of Kalman Filter/Smoother in the nonlinear systems. The EnKF has the similar update equation for each ensemble member as the Kalman Filter shown in Eq. (2.12):

$$\mathbf{x}^{a,e} = \mathbf{x}^{b,e} + \mathbf{K}^e(\mathbf{y} - \mathbf{H}\mathbf{x}^{b,e} - \eta^e),\tag{2.17}$$

in which the index e represents the number of ensemble member, and the $\mathbf{y}$ is the observation with observation error $\boldsymbol{\eta}$ drawn from: $\boldsymbol{\eta} \in \mathcal{N}(0, \mathbf{R})$, and η^e is a perturbation of $\mathbf{H}\mathbf{x}^{b,e}$ drawn from the same distribution as η. The observation $\mathbf{y}$ is mapped from the true state of the system, $\mathbf{x}^r$, into the observation space as:

$$\mathbf{y} = \mathbf{H}\mathbf{x}^r + \boldsymbol{\eta}.\tag{2.18}$$

$\mathbf{K}^e$ can be computed as:

$$\mathbf{K}^e = \mathbf{P}^{b,e}\mathbf{H}^T(\mathbf{H}\mathbf{P}^{b,e}\mathbf{H}^T + \mathbf{R})^{-1}, \tag{2.19}$$

where $\mathbf{P}^{b,e}$ denotes to the prior ensemble error covariance that can be computed using the differences between ensemble mean $\bar{\mathbf{x}}^b$ and ensemble members $\mathbf{x}^{b,1:N_e}$ as:

$$\mathbf{P}^{b,e} = \frac{1}{N_e - 1}(\mathbf{x}^{b,1:N_e} - \bar{\mathbf{x}}^b)(\mathbf{x}^{b,1:N_e} - \bar{\mathbf{x}}^b)^T. \tag{2.20}$$

When the observation operator $\mathbf{H}$ is nonlinear we can use an iterative version of the ensemble data assimilation, based on a Gauss-Newton iteration as in 3D and 4D-Var. The EnKF has been extensively applied in diverse research fields including oceanography (Bertino, Evensen, and Wackernagel, 2003), numerical weather forecast (Szunyogh et al., 2005), and hydrology (Liu, Chen, and Zhang, 2008).

2.1.2.2 Ensemble Kalman smoother and its iterative variant

The main goal of our project is to perform online parameter estimation to update some of the characteristics of the model error. We choose the EnKS (e.g. Evensen and Van Leeuwen, 2000; Khare et al., 2008; Nerger, Schulte, and Bunse-Gerstner, 2014) and its iterative form (IEnKS) (e.g. Bocquet and Sakov, 2014; Evensen et al., 2019) as the main method in our experiments, so we can assimilate all observations available in an assimilation window to update the parameters and state of the system. In this case, the update equation for e_{th} ensemble member (N_e ensemble members in total) is different from Eq. 2.17:

$$\mathbf{x}_{0:\tau}^{a,e} = \mathbf{x}_{0:\tau}^{b,e} + \mathbf{K}_{0:\tau}(\mathbf{y}_{0:L} - \mathbf{H}_{0:\tau}\mathbf{x}_{0:\tau}^{b,e} - \eta_{0:\tau}^e), \tag{2.21}$$

where L is the number of observations available, and the time series of the Kalman gain of each ensemble member can be computed as:

$$\mathbf{K}_{0:\tau}^{e} = \mathbf{P}_{0:\tau}^{b,e} \mathbf{H}_{0:\tau}^{T} (\mathbf{H}_{0:\tau} \mathbf{P}_{0:\tau}^{b,e} \mathbf{H}_{0:\tau}^{T} + \mathbf{R}_{0:L})^{-1}. \tag{2.22}$$

To achieve successful parameter estimation with the EnKS, we implement a cycling scheme in which the simulation period is divided equally into multiple windows. Details of the formulation of the EnKS and cycling scheme are thoroughly demonstrated in chapter 3. Instead of the tangent linear of the evolution and observation models, the IEnKS requires sensitivities (gradient and Hessian) obtained from the ensemble. Out of the many possible choices, we consider the algorithm proposed by Evensen et al. (2019), in which the IEnKS is formulated in the ensemble subspace and requires less computational power when dense observations are available. First, we obtain the observations using Eq. (2.18). Then, the perturbed ensemble measurements for each of the ensemble member, $\mathbf{y}^{e}$, can be computed as:

$$\mathbf{y}^{e} = \mathbf{H}\mathbf{x}^{b,e} + \boldsymbol{\eta}^{e}. \tag{2.23}$$

We define a projection $\boldsymbol{\Pi} \in \mathcal{R}^{N_e \times N_e}$ as:

$$\boldsymbol{\Pi} = (\mathbf{I} - \frac{1}{N_e} \mathbf{1}\mathbf{1}^{T})/\sqrt{N_e - 1}, \tag{2.24}$$

with $\mathbf{1} \in \mathcal{R}^{N_e}$ a vector with ones and $\mathbf{I} \in \mathcal{R}^{N_e \times N_e}$ an identity matrix. The matrix $\boldsymbol{\Pi}$ is used to subtract the mean of the ensemble and scale the result with $1/\sqrt{N_e - 1}$. Then the centered measurement-perturbation matrix is defined as:

$$\mathbf{E} = \mathbf{y}^{en}\boldsymbol{\Pi}, \tag{2.25}$$

whose columns are sampled from $\mathcal{N}(\mathbf{0}, \mathbf{R})$. Now, we start the iteration with defining a quadratic matrix:

$$\mathbf{\Omega}^i = \mathbf{I} + \mathbf{W}^i \mathbf{\Pi}, \tag{2.26}$$

where i is the iteration index and

$$\mathbf{W}^i = \mathbf{W}^{i-1} - \gamma(\mathbf{W}^{i-1} - \mathbf{S}^{iT}(\mathbf{S}^i \mathbf{S}^{iT} + \mathbf{E}\mathbf{E}^T)^{-1}\tilde{\mathbf{D}}^i). \tag{2.27}$$

where W^0 is a matrix of zeros and the innovation term $\tilde{\mathbf{D}}^i$ can be computed as:

$$\tilde{\mathbf{D}}^i = \mathbf{S}^i \mathbf{W}^i + \mathbf{y}^{pert} - \mathbf{y}^{en}, \tag{2.28}$$

in which the term $\mathbf{S}$ is the product of the scaled ensemble measurements mean and the quadratic matrix:

$$\mathbf{S}^i = \mathbf{E}\mathbf{\Omega}^{i-1}. \tag{2.29}$$

Finally, we have the update for the prior of the system for i_{th} iteration:

$$\mathbf{x}^a = \mathbf{x}^b(\mathbf{I} + \mathbf{W}^i/\sqrt{N_e - 1}), \tag{2.30}$$

The coefficient γ is a factor that allows us to use this algorithm as a iterative variant of the EnKS. In the case of EnKS, we set $\gamma = 1$. To be able to fully concentrate on the parameter estimation problem without having to deal with issues related to a small ensemble size, our experiments are typically run with an ensemble size of 200 ($N_e = 200$).

2.2 Model error in data assimilation

2.2.1 Introduction to model error

Numerical models are imperfect since the discrete geophysical model is unable to properly represent all spatial and temporal scales, nor all physical processes in nature. Model

errors, which is essentially the mismatch between the model output and the true state of nature, generated during the forecast can proliferate and have impact further on a large scale (Palmer, 2001). In numerical weather prediction (NWP), model errors are normally attributed to model deficiencies and inaccurate initial conditions. Since the atmosphere is a chaotic system that is sensitive to the initial conditions (Lorenz, 1963), much attention has been paid to producing more accurate initial conditions. However, only considering the error in the initial condition can lead to a prior error covariance matrix smaller than the actual error covariance matrix (Li et al., 2009). The error not coming from the initial condition starts to play an essential role in the forecasting system and hence the data assimilation schemes (Ren, Amezcua, and Van Leeuwen, 2021).

Due to limited computational power and insufficient observations from reality, model error was originally considered in relatively simple cases, extending from one-dimensional systems to two-dimensional shallow-water systems (see e.g seminal papers by Ghil et al., 1981; Cohn and Parrish, 1991). Later, the increase in computational power allowed for the successful application of weak constraint DA into more complex models (Ghil, 1989). For simplicity, model error has often been treated as a random variable with Gaussian distribution and no time auto-correlation in time, i.e. white in time. The reality can be quite different. The impact of model error caused by unresolved processes on the forecast and DA results can last for several model time steps. Bennett (1992) extensively discussed the use of correlated model errors and solution of the problem using the representer method.

In a Bayesian formulation of data assimilation, generating the posterior update requires knowledge of or certain error characteristics and the prior. It is hard to represent or measure the model error in the data assimilation procedure due to the large size of a typical geophysical problem and enormous amount of information required (Carrassi and Vannitsem, 2010). Fortunately, with the improvement of numerical forecasting skills and data assimilation algorithms, there have been many methods developed to deal with model

error in the data assimilation process, which will be discussed in the following subsection.

2.2.2   Treatment of the model error in data assimilation

As mentioned in chapter 1, one way to deal with model error in the data assimilation (weak-constraint) process is to treat the error as a random variable or white noise and add it to the state prior at any given time (Evensen and Van Leeuwen, 1996). Alternatively, a random multiplicative factor can be inserted in the tendencies of the model governing equations (Arnold, Moroz, and Palmer, 2013). Of course, there are many other approaches to address the model error problem, such as inflating the prior error covariance with multiplicative inflation (Anderson and Anderson, 1999) and the covariance relaxation method (Zhang, Snyder, and Sun, 2004). Some researchers take the model error as bias and furthermore, develop a method to perform an online bias correction (see an example in Dee and Da Silva, 1998), which has been successfully tested on 3D-Var (e.g. Dee and Todling, 2000; Carton et al., 2000). Treating the model error as random or biased is one simple way to deal with it but far from the truth. Model error often has impact on different scales in space and time which depends on its sources (Maraun, 2016). This thesis is based on the idea of Amezcua and Van Leeuwen (2018), where the model error is considered correlated in both space and time. Our focus is on the temporal correlation with certain formulations (see details in chapter 3 and chapter 4). Amezcua and Van Leeuwen (2018) propose a decaying autocorrelation for the model error and formulate an EnKS to adapt this specific structured model error. In chapter 3 and chapter 4, we extend the experiments from evaluating the performance of the data assimilation scheme with the autocorrelated model error to establish a possible way to estimate the characteristics of model error autocorrelation online on different models with different formulations of autocorrelation of the model error.

2.3   Parameter estimation

Data assimilation is predominantly implemented for state estimation but can be used for parameter estimation. Ruiz, Pulido, and Miyoshi (2013) state that the suboptimal settings of model parameters can lead to a significant increase in model error. Data assimilation-based parameter estimation can provide us with the optimal values of parameters that can efficiently reduce the model error in a certain metric (Hansen and Penland, 2007). However, with imperfect models, a single optimal value of a parameter doesn't exist (Smith and Mees, 2000). We have to point out that parameter estimation can be extremely difficult and sometimes impossible to solve since the parameters are often unknown and cannot be compared with the actual values of the parameters since we have no direct observations of them. There may be situations in which parameter estimation is simple. It is not our case, but in general, there can be linear relationships. The experiments from observing system simulation experiments(OSSE) where the estimated parameters can be compared with the true parameters, show that the optimal solution can be found when estimating a single parameter (see a discussion in Kotsuki et al., 2018). However, previous research (Aksoy, Zhang, and Nielsen-Gammon, 2006; Tong and Xue, 2008) has shown that estimating multiple correlated parameters can degrade the accuracy of estimating the individual parameter.

In this project, we apply a widely implemented method to perform parameter estimation with the EnKS, the state augmentation method (see an example in Carrassi and Vannitsem, 2011). The state augmentation method is to formulate an augmented state vector by considering the parameters as an auxiliary state and then using the data assimilation method to estimate the original state variable and the parameters simultaneously. The state augmentation method has been previously proven successful in model error or bias estimation (see in e.g. Martin, Nichols, and Bell, 1999; Griffith and

Nichols, 2000; Bell, Martin, and Nichols, 2004). Trudinger et al. (2008) applies the technique using the extended Kalman filter and EnKF for parameter estimation with a simplified bio-geochemical model, and the technique is also evaluated with the Korteweg–de Vries–Burgers (KdVB) numerical model within the framework of maximum likelihood ensemble filter (MLEF) (Zupanski and Zupanski, 2006). However, we find that the state augmentation methods is not always valid for parameter estimation, especially for complex circumstances, and details will be demonstrated in chapter 3, 4 and 5.

2.4   Summary

In this chapter, we started by introducing the general concept of data assimilation. A brief overview of the basics of variational data assimilation has been provided. Then, a more detailed overview of the Kalman-Filter-like data assimilation methods has been given. As mentioned previously, more details about the formulation of the ensemble methods we used will be presented later in chapter 3 and chapter 4. After introducing the data assimilation methods, an introduction of the model error in data assimilation and its treatments has been introduced. At last, we briefly introduced the difficulties of parameter estimation in data assimilation schemes and the state augmentation method we used to perform parameter estimation in our experiments. In the following chapter, we will further discuss the formulation of the EnKS we use for the autocorrelated model error and its performance, including a brief test of parameter estimation.

Chapter 3

Effect of autocorrelated model error on data assimilation results with linear model

This chapter uses a model-error-space formulated EnKS introduced by Amezcua and Van Leeuwen (2018) to investigate the performance of the EnKS under the assumption of model errors being autocorrelated. In this chapter, we answer the first research question and part of the second question: 1. *What impact does the temporal autocorrelated model error have on the data assimilation results?* 2. *Is it possible to establish a way to update the parameters encoded in the model error autocorrelation online during the data assimilation procedure?* This chapter is strongly based on the paper published on *Quarterly Journal of the Royal Meteorological Society* (Ren, Amezcua, and Van Leeuwen, 2021).

3.1   Introduction

In the atmospheric and oceanic sciences various approximate data assimilation methods have been developed in the past few decades, typically originating in either variational approaches (Courtier and Talagrand, 1987) or (Ensemble) Kalman Filter-based approaches (Evensen, 1994). Recently there has been a surge in hybrid methods trying to combine the advantages of the variational and KF-based methods (e.g. Axell and Liu, 2016; Bannister,

2017; Lorenc and Jardak, 2018; Lee, Amezcua, and Bannister, 2022).

In the past few decades, data assimilation methods like 4-dimensional variational method (4DVar) have been often performed under the assumption that the numerical models are perfect, known as the strong-constraint setting (see a discussion in e.g. Amezcua and Van Leeuwen, 2018). Typically, it is assumed that the model errors can be neglected when compared with other error sources in the systems, such as the errors in the initial condition and observations (Tremolet, 2006). Since many dynamical systems of interest are chaotic, which means they are highly sensitive to the initial condition (Lorenz, 1963), a lot of research has focused on the errors in the initial condition in order to improve the accuracy of the weather forecast.

There are cases, however, when errors not coming from initial conditions become important in the accuracy of the forecasts and hence the data assimilation process. In fact, there is ample evidence that this is the case for most, if not all, geoscience disciplines (see e.g. Fox-Kemper et al., 2019; Bony et al., 2015; Kuma et al., 2018; Muelmenstadt and Feingold, 2018; Fisher and Koven, 2020; Fennel et al., 2019). These model errors are often hard to estimate, which has hampered their inclusion in the data assimilation process. However, there are many reasons why a proper estimate of model errors needs to be included, apart from the fact that they are there in our prediction models. Jazwinski (2007) points out that in order to obtain an optimal estimate of the system, we need a better understanding of the error covariance matrices from all error sources. Furthermore, including random model errors in smoothers for chaotic systems such as the atmosphere and the ocean makes these system less dependent on initial conditions, allowing for more efficient optimisation and longer smoother windows. Indeed, with better understanding of initial and observational errors, and a strong reduction in the former, there has been an increasing number of works taking model errors into account in data assimilation process (e.g. Carrassi and Vannitsem, 2010; Howes, Fowler, and Lawless, 2017; Amezcua and

Van Leeuwen, 2018; Farchi et al., 2021).

Model error is essentially the mismatch between the true evolution of the system and the forecast produced by the numerical model over one model time step. There are various sources for model errors in numerical models, such as numerical discretization of the underlying differential equations describing the system, incorrect parameterizations, missing physical processes, etc. Some works implement a random additive variable at any given time-step as model error (e.g Evensen and Van Leeuwen, 1996), or insert a random multiplicative factor in the tendencies of the model governing equations (Arnold, Moroz, and Palmer, 2013). For simplicity, model errors are often considered Gaussian random variables with zero mean and no correlation over time. Alternatively, the model error can be considered to be fixed over the simulation period, resulting in a model bias. However, in operational systems real model errors will be complex in both spatial and temporal behaviour, as can be inferred directly from the sources of these errors (e.g. Griffith and Nichols, 2000; Mao et al., 2015; Bonavita, 2021).

In this chapter we study the case in which the spatial structure of the model error is known, but its temporal structure is uncertain. In reality both space and time structure are unknown, but we focus on the latter. We consider that the nature run evolves with a true model error; i.e. a random model forcing with a certain decorrelation time scale ω^r. We label this time scale *memory*. The imperfect forecast model uses a guessed memory ω^g, which is different from the real one.

This chapter has two main purposes. The first is to investigate the effect of the incorrect time-correlated model error on data assimilation results under different observational frequencies, and different number of observations in an assimilation window. More specifically, we aim to quantify the change in performance of the Kalman Smoother when the time statistics of the model error are misspecified, and the sensitivity of this change to

different assimilation parameters. These results are extended to the ensemble case. The second objective is to use the data assimilation process to diagnose the memory of the model error. This is of great importance since it allows to discriminate between a bias and a completely time-independent model error, and identify cases in between.

Before continuing, a simple illustration can illuminate the issue. In Figure 3.1 we show results of a smoothing process for a simple one-dimensional system over a time window of 20 nature time steps. We use an ensemble Kalman Smoother with two different observation densities in time (the details are discussed in a later section). The memories in the nature model and the forecasts models do not coincide. We can see that with $\omega^r = 0.0$, when the actual model error is a white-in-time random variable, the evolution of the true state of the system behaves rather randomly with the present model settings. If we do not know the memory and assume the model error is a bias in the data assimilation process ($\omega^g \to \infty$), the estimation made by the data assimilation method is not even close to the truth, even with very dense observations in the simulation period, as shown in the left two subplots in Figure 3.1. On the other hand, if the model error in the true model evolution behaves like a bias, and we assume that the model error is white in time in the data assimilation process, the results are quite different with different observation frequencies. As shown in two subplots on the right in Figure 3.1, with very frequent observations, we can see a fairly good performance of the data assimilation process, but with a single observation, the estimation is still not accurate.

The general structure of this chapter is as follows. In section 2, we investigate the performance of the Kalman Smoother on a linear model with time-correlated model error analytically. When we are unable to find closed expressions, we numerically evaluate the (open) analytical expressions when necessary. We determine the behaviour of the posterior variance and mean-square error for different values of true and guessed memory.
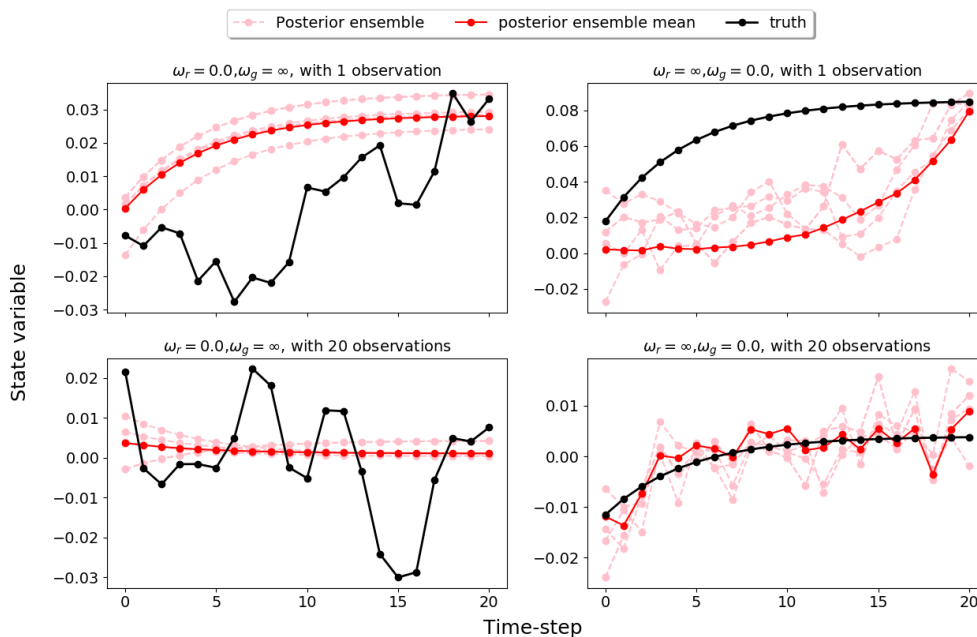
FIGURE 3.1. Plots of the trajectories of the true state of the system (black), three posterior ensemble members (pink, randomly chosen from 200 members), and the posterior ensemble mean (red). The left subplots show results for a true white noise model error and an assumed bias model error for two observation densities. Note that the posterior estimates are poor in both cases. The right subplots depict a bias true model error and an assumed white noise model error. The result with one observation is poor, while if many observations are present the assimilation result is consistent within the ensemble spread.

Next, a higher dimensional system is explored via numerical experiments using the Ensemble Kalman Smoother in section 3. In section 4 we use state augmentation to try to infer the memory time scale from the assimilation process, with satisfactory results. Section 5 contains a summary and a discussion of the results.

In this chapter we follow the notation introduced by Amezcua and Van Leeuwen (2018). Identifying different attributes in a variable can be difficult in some expressions. In general, superscripts are used as time indices. If there is a comma in the superindex it is because we have also added a label corresponding to the role in the data assimilation process. For instance, the variable $\mu_t^{b,x}$ would be the background mean of x at time t.

There are some cases in which the function is unclear, for instance a superscript applied to a vector cannot mean exponentiation. In the case where superscripts correspond to exponents, this is clearly specified in the text. An example of more complicated uses is: K_t^{x,ω^g} which refers to the Kalman gain evaluated in x space at time t computed with the covariance matrix which uses the guessed memory ω^g. More complicated uses of the sub- and superscripts are clearly identified in the text, and we also recommend the reader to check Amezcua and Van Leeuwen (2018) for clarity.

3.2   Time-correlated model error in the Kalman Smoother

Let us consider a simple linear model with the governing equation over 1 model time step:

$$\mathbf{x}_{t+1} = \mathbf{M}_{t\to(t+1)}\mathbf{x}_t + \boldsymbol{\nu}_{t+1}, \tag{3.1}$$

where $\mathbf{M}_{t\to(t+1)} \in \mathcal{R}^{N_x \times N_x}$ represents the linear model operator and its size depends on the number of variables in the system N_x, $\mathbf{x}_t \in \mathcal{R}^{N_x}$ is the state variable at given time-step t, and $\boldsymbol{\nu}_{t+1} \in \mathcal{R}^{N_x}$ is an additive model error which contains correlation in time and space. The initial condition of the random variable, $\mathbf{x}_0 \in \mathcal{R}^{N_x}$, is drawn from a multivariate Gaussian distribution (MGD), $\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0^{b,x}, \mathbf{B})$, where $\boldsymbol{\mu}_0^{b,x} \in \mathcal{R}^{N_x}$ is the mean of the random variable and $\mathbf{B} \in \mathcal{R}^{N_x \times N_x}$ is its covariance matrix.

The time-correlated model error at time t also comes from a MGD, $\boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, with zero mean and the covariance matrix $\mathbf{Q} \in \mathcal{R}^{N_x \times N_x}$. We also consider spatial correlations for the model errors, hence $\mathbf{Q}$ is not diagonal. We follow Amezcua and Van Leeuwen (2018) and assume that the model errors are correlated in time as:

$$Cov(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j) = \phi(|i-j|, \omega)\mathbf{Q}, \tag{3.2}$$

where $\phi(|i-j|, \omega)$ represents the memory of the model errors, $|i-j|$ is the absolute difference between time steps i and j, and ω represents the characteristic memory timescale of the model error. Then, a one-step propagation of the model error can be formulated as:

$$\boldsymbol{\nu}_{i+1} = \phi(1, \omega)\boldsymbol{\nu}_i + \sqrt{1 - \phi(1, \omega)^2}\boldsymbol{\nu}_{i+1}^{random}, \tag{3.3}$$

where $\boldsymbol{\nu}_{i+1}^{random}$ is the random part of the model error at time-step $t+1$. The function ϕ decreases monotonically to 0 as $|i-j|$ increases, and the maximum value of the ϕ is 1.0 as the absolute difference between time steps i and j tends to 0. For simplicity, we choose an exponentially decaying memory for the model error:

$$\phi(|i-j|, \omega) = e^{-\frac{|i-j|}{\omega}}. \tag{3.4}$$

When the correlation timescale ω tends to 0.0, which indicates no temporal correlation in model errors, the ϕ function becomes a Kronecker delta function and the linear model becomes a 1st-order Markov model:

$$\phi(|i-j|, \omega) = \begin{cases} 1, & \text{if } i = j. \\ 0, & \text{otherwise.} \end{cases} \tag{3.5}$$

In the other limit when ω tends to infinity, the memory of the model errors becomes 1.0 and the model error is fixed in time.

3.2.1 Formulation of the Kalman Smoother

We start with the formulation of the Kalman Smoother as described in (Amezcua and Van Leeuwen, 2018). It uses an extended control variable, $\mathbf{z}_{0:\tau} \in \mathcal{R}^{(\tau+1)\times N_x}$ over $\tau + 1$ model time steps. This construction simplifies the representation of the covariance matrix and the exposition of the method. This extended variable can be written as the initial state of the system $\mathbf{x}_0 \in \mathcal{R}^{N_x}$, plus a collection of the model errors over time, $\boldsymbol{\nu}_{1:\tau} \in \mathcal{R}^{\tau\times N_x}$:

$$\mathbf{z}_{0:\tau} = \begin{bmatrix} \mathbf{x}_0 \\ \boldsymbol{\nu}_{1:\tau} \end{bmatrix}. \tag{3.6}$$

The extended variables can be transformed back to state space via:

$$\mathbf{x}_t = \mathbf{M}_{0:t}\mathbf{z}_{0:t}, \tag{3.7}$$

where $\mathbf{M}_{0:t} \in \mathcal{R}^{(t+1)N_x \times N_x}$ is the extended model operator and can be formulated as a block-matrix:

$$\mathbf{M}_{0:t} = \left[\mathbf{M}_{0\to t}, \mathbf{M}_{1\to t}, \mathbf{M}_{2\to t}, \mathbf{M}_{3\to t}, \cdots, \mathbf{M}_{(t-1)\to t}, \mathbf{I}\right]. \tag{3.8}$$

The extended form also follows a MGD $\mathbf{z}_{0:\tau} \sim \mathcal{N}(\boldsymbol{\mu}_{0:\tau}^{b,z}, \mathbf{D}_{0:\tau})$, with mean $\boldsymbol{\mu}_{0:\tau}^{b,z} \in \mathcal{R}^{(\tau+1)N_x}$:

$$\boldsymbol{\mu}_{0:\tau}^{b,z} = \begin{bmatrix} \boldsymbol{\mu}_0^{b,x} \\ \boldsymbol{\mu}_{1:\tau}^{b,\nu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_0^{b,x} \\ \mathbf{0}_{1:\tau} \end{bmatrix}. \tag{3.9}$$

In this case, the prior covariance matrix $\mathbf{D}_{0:\tau} \in \mathcal{R}^{(\tau+1)N_x \times (\tau+1)N_x}$ has a simple form, which can be written as a block-matrix:

$$\mathbf{D}_{0:\tau} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{1:\tau} \end{bmatrix}. \tag{3.10}$$

The covariance matrix of the extended control variable has two separate and independent parts: the part that comes from the initial condition, $\mathbf{B} \in \mathcal{R}^{N_x \times N_x}$, and the part that originates purely from the correlated model errors, $\mathbf{Q}_{1:\tau} \in \mathcal{R}^{\tau N_x \times \tau N_x}$. The covariance

matrix $\mathbf{Q}_{1:\tau}$ is a block-matrix and can be written as a Kronecker product of a *Toeplitz matrix* and the spatial covariance matrix of the model error, $\mathbf{Q}$:

$$\mathbf{Q}_{1:\tau} = \mathbf{\Phi}_{1:\tau} \otimes \mathbf{Q}, \tag{3.11}$$

where the Toeplitz matrix $\mathbf{\Phi}_{1:\tau} \in \mathcal{R}^{\tau \times \tau}$ contains all the memory coefficients. This Toeplitz matrix, $\mathbf{\Phi}_{1:\tau}$, has different forms in different scenarios:

- When $\omega \to 0$, the Toeplitz matrix becomes an identity matrix, $\mathbf{I} \in \mathcal{R}^{\tau \times \tau}$, and the Kronecker product $\mathbf{Q}_{1:\tau}$ becomes a block-diagonal matrix.

- When $\omega \to \infty$, the Toeplitz matrix $\mathbf{\Phi}_{1:\tau}$ is a matrix of ones and $\mathbf{Q}_{1:\tau}$ becomes a block-matrix, in which every block element is the spatial covariance matrix $\mathbf{Q}$.

To demonstrate the structure of the Kalman Smoother solution, we consider only one single observation at time-step τ. Details of the formulation with multiple observations can be found in Amezcua and Van Leeuwen (2018). Then, the Kalman gain acting upon the whole simulation period in extended-variable space, $\mathbf{K}_{0:\tau}^{z}$, can be computed as:

$$\mathbf{K}_{0:\tau}^{z} = \mathbf{D}_{0:\tau}(\mathbf{M}_{0:\tau})^{T}\mathbf{H}^{T}(\mathbf{H}\mathbf{M}_{0:\tau}\mathbf{D}_{0:\tau}(\mathbf{H}\mathbf{M}_{0:\tau})^{T} + \mathbf{R})^{-1}. \tag{3.12}$$

With the Kalman Gain we can update the extended control variable using the Kalman equation, assuming that the state initial $\mathbf{x}^{0}$, the observation error $\boldsymbol{\eta}$ and the model error $\boldsymbol{\nu}$ are statistically independent of each other. Hence, the analysis mean is:

$$\mathbf{z}_{0:\tau}^{a} = \mathbf{z}_{0:\tau}^{b} + \mathbf{K}_{0:\tau}^{z}\mathbf{d}, \tag{3.13}$$

where $\mathbf{d}$ is the innovation between observations and the model output at observational time t, which can be calculated as:

$$\mathbf{d} = \mathbf{y} - \mathbf{HM}_{0:\tau}\mathbf{z}_{0:\tau}^{b}. \tag{3.14}$$

The vector $\mathbf{y}$ represents the observations obtained from the true evolution of the system by the observation operator, $\mathbf{H} \in \mathcal{R}^{N_y \times N_x}$, including the observational error:

$$\mathbf{y}_t = \mathbf{Hx}_t^r + \boldsymbol{\eta}_t, \tag{3.15}$$

where $\boldsymbol{\eta}_t \in \mathcal{R}^{N_y}$ is the observational error which follows a zero-mean MGD $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ and its size depends on the number of variables observed from the system N_y, $\mathbf{x}_t^r$ represents the *real* state of the system at time-step t, and $\mathbf{R} \in \mathcal{R}^{N_y \times N_y}$ represents the covariance matrix of the observation errors. Note that the observational time can be anywhere inside the assimilation window $0 \leq t \leq \tau$. Finally, the covariance matrix is updated via:

$$\mathbf{A}_{0:\tau}^{z} = (\mathbf{I} - \mathbf{K}_{0:\tau}^{z}\mathbf{HM}_{0:\tau})\mathbf{D}_{0:\tau}. \tag{3.16}$$

Considering more than 1 observation per assimilation window does not yield simple expressions. Instead, it can be done in two ways. First, we can consider modified expressions as in the Appendix of Amezcua and Van Leeuwen (2018). Second, the observations can be assimilated serially one after the other. Since the observation error covariance matrix is assumed diagonal, this is equivalent to updating observations all-at-once.

3.2.2 Evaluating the performance of the Kalman smoother with time-correlated model error

Amezcua and Van Leeuwen (2018) established a framework to handle time time-correlated model errors in the Kalman Smoother and its ensemble implementation. Nonetheless, they did not evaluate the performance of the methods they discussed, and they did not study the consequences (in this performance) of using wrong memory of the model error in the forecast. This is one of the two new contributions of this work, and it is detailed in this chapter.

A data assimilation system should be able to produce accurate estimations of the posterior density function of the state variables. In practice assuming a unimodal posterior, it should at least be able to produce a mean trajectory which remains "close" to the (unknown) truth, and provide an uncertainty measure corresponding to the true uncertainty of the mean with respect to the truth.

One common approach is to compare the root-mean-square error (RMSE) which is the true error of the posterior mean, with the posterior standard deviation, or spread, which is the error estimated by the data assimilation method (Fortin et al., 2014). When the data assimilation results give us the "best" estimation of the system, the ratio of the RMSE and the spread should approximately be equal to 1.0. To simplify the situation, instead of comparing the RMSE with the spread, we use the ratio of the mean-square error (MSE) and the variance of the state variable.

Before proceeding to actual experiments we find the analytical expressions for both the MSE of the background and analysis. We also analyse in detail the variance expressions shown in Amezcua and Van Leeuwen (2018). To simplify calculations we assume that the state is one-dimensional and the model operator is a damping coefficient, α. The model is pure noise if α tends to 0.0, and a random walk model when $\alpha = 1.0$. We choose

a damping coefficient between 0.0 and 1.0 to ensure that the linear model is stationary. This leads to a model equation:

$$x_{t+1} = \alpha x_t + \nu_{t+1}. \tag{3.17}$$

For the next subsections we work in the state variable space, i.e. our control variable is $\mathbf{x}_{0:\tau}$. This follows two reasons: the meaning of the expressions is more tractable, and the implementation in the ensemble case is more straightforward. The general expressions are obtained as double sums which are not easy to visualise. In some cases these double sums can be evaluated, leading to expressions provided in the tables in the Appendix. In other cases we evaluate the expressions numerically and provide graphical illustrations.

3.2.2.1   Posterior variance in the Kalman smoother

The prior variance at any time and covariance between two different time steps in our scalar system have the following expressions:

$$
\begin{aligned}
Var(x_t^b) &= \alpha^{2t} b^2 + q^2 \sum_{i=1}^{t} \sum_{j=1}^{t} \alpha^{2t-i-j} \phi(|i-j|, \omega) \\
Cov(x_{t_1}^b, x_{t_2}^b) &= \alpha^{t_1+t_2} b^2 + q^2 \sum_{i=1}^{t_1} \sum_{j=1}^{t_2} \alpha^{t_1+t_2-i-j} \phi(|i-j|, \omega),
\end{aligned}
\tag{3.18}
$$

where b^2 is the variance of the initial x_0, the superscript b denotes the prior, and q^2 is the variance of the model error. In (3.18), in the expressions involving the scalars α, b and q the exponent actually means the constant raised to a power, as opposed to being a super-index.

According to (3.18), the prior covariance and variance have two sources: the initial condition which is the first term on the right-hand side (RHS), and the auto-correlated model errors as the double sum term on the RHS. Of course, Equation (3.18) is only suitable for $t > 0, t_1 > 0, t_2 > 0$. As a special case, since the initial condition x_0^b is independent from the model errors at any given time, its variance and covariance are given by:

$$Var(x_0^b) = b^2$$
$$Cov(x_0^b, x_t^b) = \alpha^t b^2. \tag{3.19}$$

Once more, the expression α^t means the constant α raised to the power t. To obtain a feeling for Equation (3.18), Tab. I, which is listed in the Appendix , contains results on limiting cases for ω and α where the results of the sums can be evaluated analytically. Figure 3.2 shows that the prior variance is a monotonically increasing function of both α and ω^g, and, not surprisingly, of time. The prior variance is almost constant when
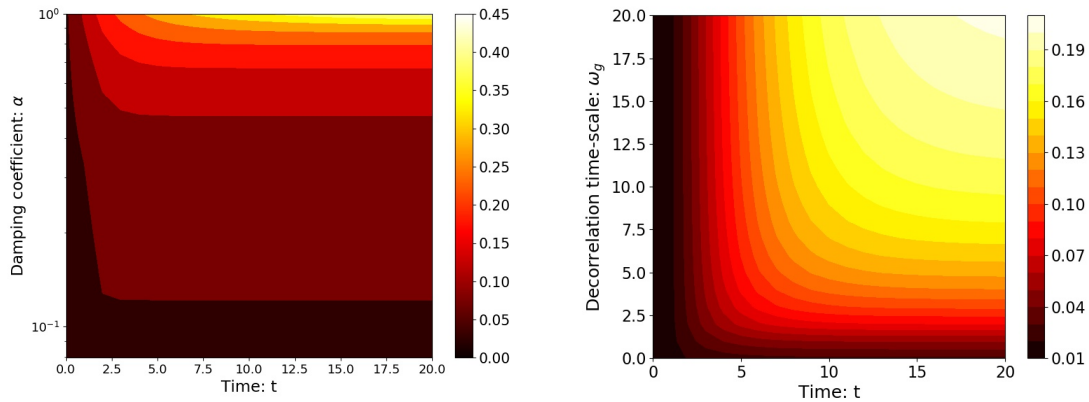

FIGURE 3.2. Prior variance as function of time in the window and of the damping coefficient with $\omega^g = 1.0$ (left), and as function of ω^g with $\alpha = 0.8$ (right). Note that the y-axis of the left panel has a logarithmic scale.

$\alpha < 0.5$. For larger values of α the prior variance increases much faster with ω^g. The

prior variance as function of ω^g shows the opposite behaviour: When ω^g is between 0.0 and 10.0 the prior variance increases significantly with increasing ω^g, but for larger ω^g values the increase of the variance slows down.

Since the posterior variance is the estimated error resulting from the data assimilation scheme, and in linear data assimilation the posterior variance is independent of the actual value of the observations, the posterior variance has no knowledge of the real decorrelation timescale of the model errors, ω^r. The posterior variance at a given time-step, assuming that we have a single observation at time-step τ, can be simplified as:

$$Var(x_t^a) = Var(x_t^b) - K_t^{x,\omega^g} Cov(x_\tau^b, x_t^b), \tag{3.20}$$

where K_t^{x,ω^g} is the Kalman Gain formulated in the x-space acting on the current time-step t and in this scalar case can be computed as:

$$K_t^{x,\omega^g} = \frac{Cov(x_t^b, x_\tau^b)}{Var(x_\tau^b) + r^2}, \tag{3.21}$$

where r^2 is the variance of the observation error, and clearly the exponent means the second power. We can see that the Kalman gain depends on the covariance between the state at the present time and at the observational time, the state variance at the observational time and the observation error. These expressions correspond to the state-space-formulation in Amezcua and Van Leeuwen (2018). We also compute some limiting cases on the posterior variance with a single observation for ω and α shown in Tab II, which is in the Appendix at the end of this thesis.

When more than one observation is present within an assimilation window, it is difficult to find simple analytical expressions and we refer to numerical evaluation. We start our numerical experiments with a fixed damping coefficient $\alpha = 0.8$, but with different memories ω. The results are shown in Figure 3.3.

The first thing that strikes the eye is the low posterior variance at observation times, which is as expected. Another clear trend is the decrease of posterior error with increasing ω^g. This is directly related to the spread of observation information in the system: a larger ω^g gives more memory in the system, and hence observations have a larger influence over time. In some plots the posterior variance is decreasing towards the initial time, while in others it is increasing. This, however, is mainly due to the different color scales in the plots, the posterior variance at initial time is mainly set by the prior variance, although observations do have an influence for larger decorrelation timescales. Finally, one can notice a decrease of the posterior variance for ω^g close to zero. This behavior has its roots in the behavior of the prior, which has minimal variance for small ω^g.
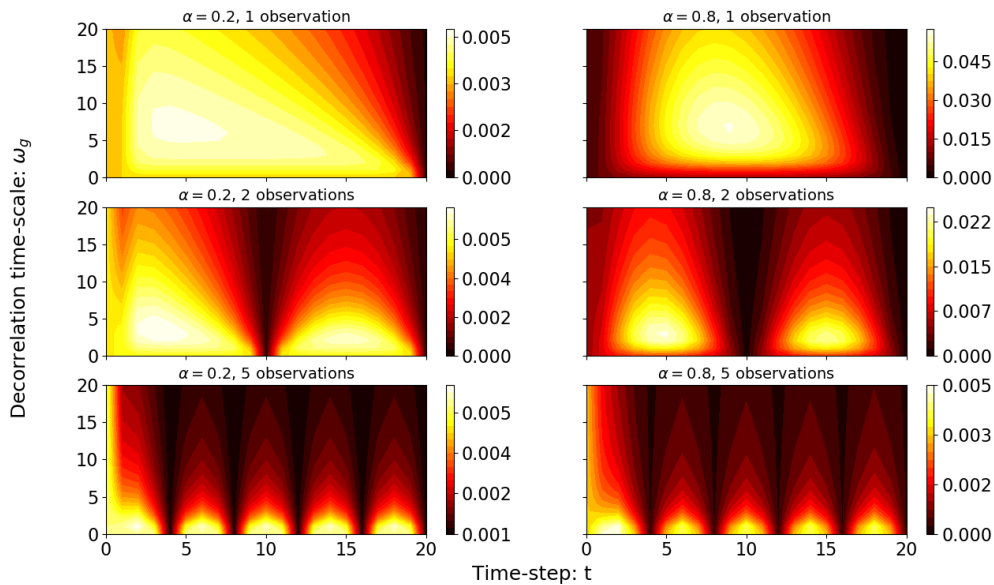

FIGURE 3.3. Posterior variance as function of time in the window and of ω^g, for two fixed damping coefficients, $\alpha = 0.2$ (left panels) and $\alpha = 0.8$ (right panels), using different numbers of observations in the simulation window (top, middle, bottom panels). Note the different color scales.

To see this latter point better we resort back to the analytical treatment of the case of a

single observation at the end of the simulation period, at $t = \tau$. We focus on the posterior variance at the initial time and at the observational time as those times show most interesting behaviour. As we have seen above, the initial variance and the covariance between the initial state and the state at any time is independent of the decorrlation timescale, so:

$$\frac{\partial Cov(x_0^b, x_t^b)}{\partial \omega^g} = \frac{\partial Var(x_0^b)}{\partial \omega^g} = 0. \tag{3.22}$$

Using this, we find for the Kalman gain from Equation (3.21) and Figure 3.2:

$$\frac{\partial K_0^{x,\omega^g}}{\partial \omega^g} = -\frac{\partial Var(x_\tau^b)}{\partial \omega^g} \frac{Cov(x_0^b, x_\tau^b)}{(Var(x_\tau^b) + r^2)^2} < 0, \tag{3.23}$$

and so the Kalman gain for initial time is a decreasing function of the decorrelation timescale. Using this we find for the posterior variance at initial time:

$$\frac{\partial Var(x_0^a)}{\partial \omega^g} = -\frac{\partial K_0^{x,\omega^g}}{\partial \omega^g} Cov(x_0^b, x_\tau^b) > 0, \tag{3.24}$$

which is an increasing function of the decorrelation timescale. At the observational time we can do a similar derivation:

$$\frac{\partial K_\tau^{x,\omega^g}}{\partial \omega^g} = \frac{\partial Var(x_\tau^b)}{\partial \omega^g} \frac{r^2}{(Var(x^b\tau) + r^2)^2} > 0, \tag{3.25}$$

leading to:

$$\begin{aligned}
\frac{\partial Var(x_\tau^a)}{\partial \omega^g} &= (1 - K_\tau^{x,\omega^g})\frac{\partial Var(x_\tau^b)}{\partial \omega^g} - \frac{\partial K_\tau^{x,\omega^g}}{\partial \omega^g} Var(x_\tau^b) \\
&= \frac{\partial Var(x_\tau^b)}{\partial \omega^g} \frac{r^4}{(Var(x_\tau^b) + r^2)^2} > 0.
\end{aligned} \tag{3.26}$$

We thus find that both at initial and at observation times the posterior variance increases with ω^g. In fact, this derivation shows that this is true for all values of ω^g, at initial

and final times, not only for small ω^g values as Figure 3.3 might suggest.

3.2.2.2 Mean-square error (MSE) of the posterior in the Kalman Smoother

Different from the posterior variance, for the MSE between the analysis mean and the true state of the system differences between the real decorrelation timescale and the one assumed in the data assimilation are important. We calculate the MSE of the prior as the difference between the prior mean μ_t^b and the truth. The truth is a realization of the true prior pdf at the initial time. The MSE at any time t is defined as:

$$MSE_t^b = \int (\mu_t^b - x_t^r)^2 p(x_t^r) \, dx_t^r. \tag{3.27}$$

When the statistics of the model error used in the data assimilation is different from that of the truth the prior pdf used in the data assimilation will deviate from the pdf that the truth is drawn from. Writing the pdf from which the truth is drawn as: $x_t^r \in \mathcal{N}(\nu_t, B_t)$, where ν_t is its mean at time-step t and B_t represents its variance the MSE at time t becomes:

$$MSE_t^b = E_{x_t^r}[(\mu_t^b - x_t^r)^2] = B_t + (\mu_t^b - \nu_t)^2, \tag{3.28}$$

in which the last term represents the bias in the prior. Using this in a Kalman Smoother we can compute the posterior MSE as:

$$\begin{aligned} MSE_t^a &= E_{x_t^b}[(\mu_t^a - x_t^b)^2] \\ &= E_{x_t^b}[(\mu_t^b - x_t^r) + K_t^{x,\omega^g}(x_\tau^r - \mu_\tau^b) + K_t^{x,\omega^g}(y_\tau - x_\tau^r)^2] \\ &= B_t + (\mu_t^r - \nu_t)^2 + \left(K_t^{x,\omega^g}\right)^2 \left(B_\tau + (\mu_\tau^b - \nu_\tau)^2 + r^2\right) \\ &\quad - 2K_t^{x,\omega^g}\left(Cov(x_t^r, x_\tau^r) + (\mu_t^r - \nu_t)(\mu_\tau^b - \nu_\tau)\right). \end{aligned} \tag{3.29}$$

In the ideal case when $\omega^g = \omega^r$, the MSE of the posterior can be simplified as:

$$
\begin{aligned}
MSE_t^a &= B_t + \left(K_t^{x,\omega^g} \right)^2 \left(B_\tau + r^2 \right) - 2K_t^{x,\omega^g} Cov(x_t^r, x_\tau^r) \\
&= B_t - K_t^{x,\omega^g} Cov(x_t^r, x_\tau^r).
\end{aligned}
\tag{3.30}
$$

As expected, the posterior MSE in the ideal case is the same as the posterior variance shown in Equation (3.20) because the statistics of the prior and the truth are the same in this ideal case. When more than one observation is present in the time window we can write the Kalman Smoother MSE as:

$$
\begin{aligned}
MSE_{0:\tau}^a &= E_{\mathbf{x}_{0:\tau}^r}[\boldsymbol{\mu}_{0:\tau}^a - \mathbf{x}_{0:\tau}^r][\boldsymbol{\mu}_{0:\tau}^a - \mathbf{x}_{0:\tau}^r]^T \\
&= E_{\mathbf{x}_{0:\tau}^r}[\boldsymbol{\mu}_{0:\tau}^b + \mathbf{K}_{0:\tau}^{x,\omega^g}(\mathbf{H}_{1:L}\mathbf{x}_{0:\tau}^r - \mathbf{H}_{1:L}\boldsymbol{\mu}_{0:\tau}^b + \eta_{0:\tau}) - \mathbf{x}_{0:\tau}^r][\ldots]^T \\
&= E_{\mathbf{x}_{0:\tau}^r}[(\mathbf{I} - \mathbf{K}_{0:\tau}^{x,\omega^g}\mathbf{H}_{1:L})(\boldsymbol{\mu}_{0:\tau}^b - \mathbf{x}_{0:\tau}^r) + \mathbf{K}_{0:\tau}^{x,\omega^g}\eta_{0:\tau}][\ldots]^T \\
&= (\mathbf{I} - \mathbf{K}_{0:\tau}^{x,\omega^g}\mathbf{H}_{1:L})\mathbf{B}_{0:\tau} \\
&\quad + (\mathbf{I} - \mathbf{K}_{0:\tau}^{x,\omega^g}\mathbf{H}_{1:L})(\boldsymbol{\mu}_{0:\tau}^b - \boldsymbol{\nu}_{0:\tau})(\boldsymbol{\mu}_{0:\tau}^b - \boldsymbol{\nu}_{0:\tau})^T(\mathbf{I} - \mathbf{K}_{0:\tau}^{x,\omega^g}\mathbf{H}_{1:L})^T,
\end{aligned}
\tag{3.31}
$$

where $\boldsymbol{\mu}_{0:\tau}^a$ is the time-series of the posterior mean from the Kalman Smoother, $\mathbf{B}_{0:\tau}$ represents the covariance matrix derived from the true pdf, $\mathbf{K}_{0:\tau}^{x,\omega^g}$ is the Kalman Gain matrix calculated with ω^g, and $\boldsymbol{\mu}_{0:\tau}^b$ is the prior mean time-series. The observation operator $\mathbf{H}_{1:L}$ maps L observations, from the state space into the observation space. Written in this form it is relatively easy to understand what the influence of a mis-specified model error is. However, this is slightly deceiving in that the result is written in terms of the true covariance and mean, which are unknown in the real world, and the Kalman Gain using the incorrect model error description. In the ideal scenario the MSE can be simplified to:

$$
MSE_{0:\tau}^a = (\mathbf{I} - \mathbf{K}_{0:\tau}^x\mathbf{H}_{1:L})\mathbf{B}_{0:\tau},
\tag{3.32}
$$

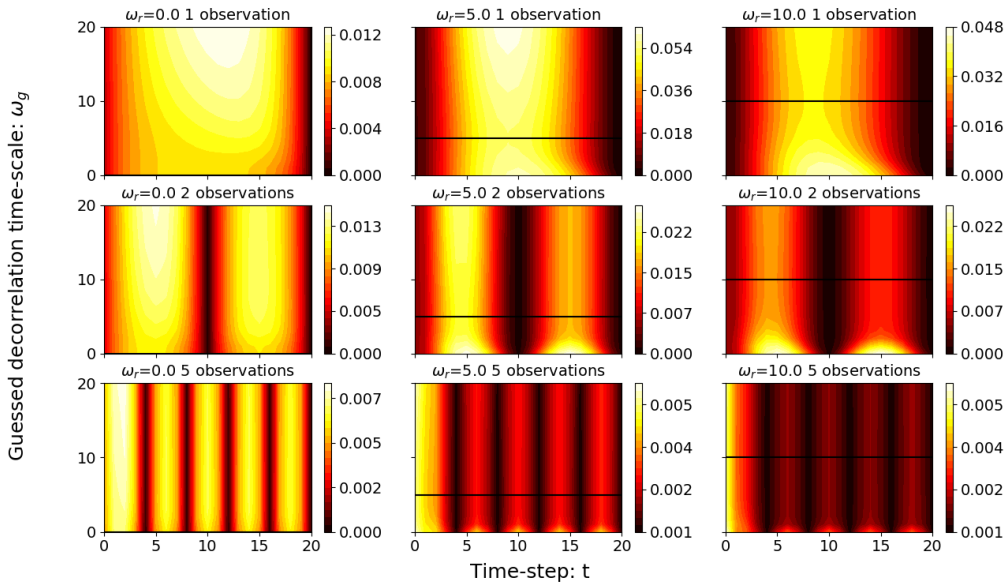where $\mathbf{K}_{0:\tau}^x$ is the optimal gain for the KS. Equation 3.32 shows the exact solution for the

FIGURE 3.4. Posterior MSE as function of time in the window and of ω^g. The different panels show results for different fixed ω^r (left, middle and right panels), and different number of observations in the simulation window (top, middle and bottom panels). The solid black line indicates where $\omega^g = \omega^r$.

posterior covariance matrix shown in Equation 3.16 in the variable space.

The behaviour of the posterior MSE when the memory in the prior differs with that of the true system, i.e. $\omega^g \neq \omega^r$, is found from the numerical evaluation of the analytical expressions shown in Equation (3.29), and the results are shown in Figure 3.4. Figure 3.4 shows that, in general, the magnitude of the posterior MSE decreases as the observation frequency increases. This matches the results shown in Figure 3.3 for the posterior variance. As we expected, the posterior MSE reaches its minimum at the observational time. From the top three subplots, we can see that with a single observation at the end of the simulation window, the MSE is minimized when $\omega^g = \omega^r$ for the time-steps that are away from the observational time and initial state. However, when the number of observations in the window increases the difference between ω^r and ω^g becomes less important: the solid lines do not dominate large changes in MSE.

The Appendix contains derivations and analytical results for the Ensemble Kalman Smoother, where we specifically study the influence of sampling errors.

3.2.3   Evaluation of the Kalman Smoother for a 1-dimensional system

To evaluate the performance of the Kalman Smoother we compute the ratio of the MSE over the variance of the posterior averaged over the simulation window, with different observational frequencies. Figure 3.5 shows the numerical evaluation of analytical expressions which contain ratio's of double sums and are hence difficult to visualise without plotting them.

As we can see, the Kalman Smoother works well when $\omega^g = \omega^r$ for all the cases, with the ratio of MSE over the variance equal to 1.0. With relatively high observational frequency, 5 observations or more in the simulation window, the MSE is larger than the estimated posterior variance when $\omega^g > \omega^r$, and vice versa. From the numerical results shown Figure 3.4, the mismatch between the two timescales ω^r and ω^g barely has any impact on the MSE. The ratio is dominated by the posterior variance.

To understand the behavior of the ratio in Fig 3.5 for small observation numbers we refer to Figure 3.6, which shows the time average posterior variance as function of ω^g for the case of one observation in the time window, as the black line. The concave shape is due to a combination of two effects. Firstly, the prior variance grows with ω^g as a larger ω^g gives rise to a larger decorrelation timescale, so errors persist in the time window. This effect leads to a growth in posterior variance with ω^g. Secondly, a larger ω^g reduces the posterior variance because the larger decorrelation timescale allows the observation observation information to spread more over the time window. These two competing effect

(A) with single observation
at the end of each window

(B) with 2 observations in a
simulation window

(C) with 5 observations in a
simulation window
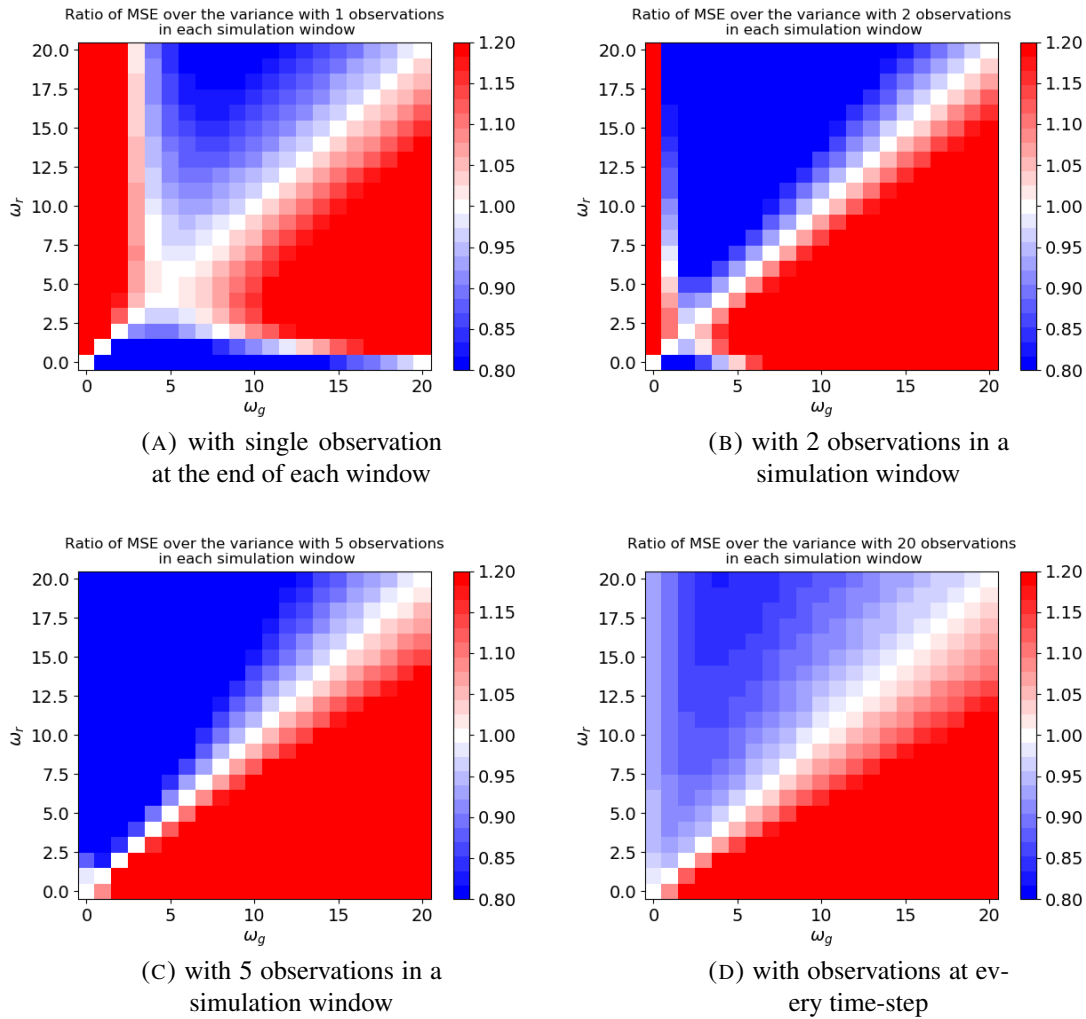
(D) with observations at every time-step

FIGURE 3.5.   Ratio of MSE over the posterior variance for the 1-dimensional system, calculated using numerical evaluation of the exact analytical expressions. The different panels show results for different numbers of observations.

lead to a maximum in posterior variance.

Fig 3.6 also shows the MSE for 3 different values of ω^r. The MSE curves are all convex, with a minimum when $\omega^g = \omega^r$, as expected since the minimum value of the MSE happens when the guess decorrelation time scale is equal to the real time scale. The ratio the MSE to the posterior variance is equal to one when their curves cross, and we see
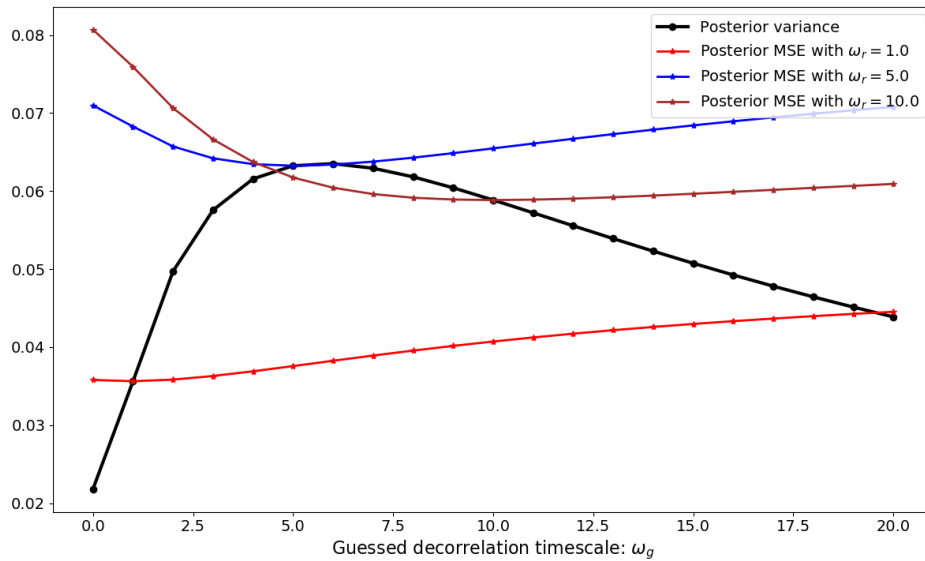
FIGURE 3.6. The time-averaged posterior variance (black) and posterior mean-square error with different ω^r (red,blue,brown), as a function of ω^g, using a fixed damping coefficient $\alpha = 0.8$.

immediately that the curves cross twice when the position of the minimum of the MSE is different from that of the maximum in the posterior variance. This is exactly what Fig 3.5 shows for 1 observation, and the structure of that solution is fully determined by the position of the peak in the posterior variance. For 2 observations we see qualitatively similar behavior, with the peak in the posterior variance shifting closer to zero. For 5 and 20 observations the peak in the posterior variance shifts all the way to zero because the influence of the observations becomes more important than the prior, so the posterior variance becomes a decreasing function of ω^g, as can also be observed in Figure 3.3. This means that the MSE and posterior variance curves only cross once, where $\omega^g = \omega^r$, as Figure 3.5 indeed shows.

Ideally we would be able to show this behavior exploring the analytical expressions of Eq (3.18), but the expressions become rather complicated as we would have to analytically evaluate ratio's of integrals over double sums, which we were unable to perform.

Finally, it should be noted that Figure 3.5 first calculates the MSE over posterior variance ratio and then averages over time, while Figure 3.6 and the argument above first average over time and then calculate the ratio. The results are qualitatively the same because Figure 3.3 and 3.4 show a similar behavior over time.

3.3 Time-correlated model error in a higher dimensional system

In this section we explore how the analytical results from the 1-dimensional system carry over to systems with a relatively higher dimensions. To this end we implement an Ensemble Kalman Smoother (EnKS) (Evensen and Van Leeuwen, 2000) using perturbed model forecasts (Van Leeuwen, 2020) with 200 ensemble members on a 10-dimensional system in which the deterministic model consists of a diagonal matrix with the damping coefficient on the diagonal,and spatially and temporally correlated model errors. This means that although the elements of the state are evolving independently over time they become more and more correlated due to the correlated model error. The large ensemble size with respect to the size of the state variable ensures that sample effects are small. Four cases with four different observation frequencies are discussed, similar to the experiments we do for the 1-dimensional system.

We generate the true trajectory of the system, $\mathbf{x}_{0:\tau}^r$ with ω^r, and all the prior ensemble members are generated using ω^g. The assimilation is run over 50 time windows, in which the results from one window provide the prior for the initial conditions for the next window (i.e. cycling). There are 20 time-steps ($\tau = 20$) in each time window.

We experiment with different combinations of ω^r and ω^g with the same range from 0.0 to 20.0, and the 4 observation settings explored above. To evaluate the performance

on the EnKS we calculate the ratio of the MSE and the ensemble variance. The MSE at a given time-step t is computed as:

$$MSE_t = \frac{(\mathbf{x}_t^r - \bar{\mathbf{x}}_t^a)^T(\mathbf{x}_t^r - \bar{\mathbf{x}}_t^a)}{N_x}, \tag{3.33}$$

where $\bar{\mathbf{x}}_t^a$ is the mean of the posterior ensemble and $\mathbf{x}_t^r$ is the true state of the system.

After obtaining the MSE and the variance for each time step, we calculate the their ratio and the average of this ratio over the whole simulation period, and the results are shown in Figure 3.7. We find that the ratio of MSE over the posterior variance matches with the results for 1-dimensional system shown in Figure 3.5. The EnKS performs well with a correct guessed decorrelation timescale ($\omega^g = \omega^r$). With different observation frequencies, the ratio shows a similar behavior as the ratio of MSE over the variance of the posterior in the 1-dimensional system, but the differences are about 10% larger in the higher dimensional case. For the higher dimensional system, it seems that the posterior ensemble spread is still the main factor to the ratio, which has a non-monotonic behaviour with ω^g and becomes monotonically decreasing as ω^g increases.

3.4   Estimation of the memory in the model error

In the previous sections we showed that using an incorrect memory timescale in the model error can have a significant impact on the data assimilation results. Unfortunately, in many practical situations we do not know this memory timescale. However, it is possible to treat that correlation time scale as an unknown quantity and perform parameter estimation along the state estimation.

Parameter estimation via state augmentation has been used before, e.g. with an extended Kalman Filter (Carrassi and Vannitsem, 2011), and even to determine parameters

(A) with single observation at the end of each window

(B) with 2 observations in a simulation window

(C) with 5 observations in a simulation window
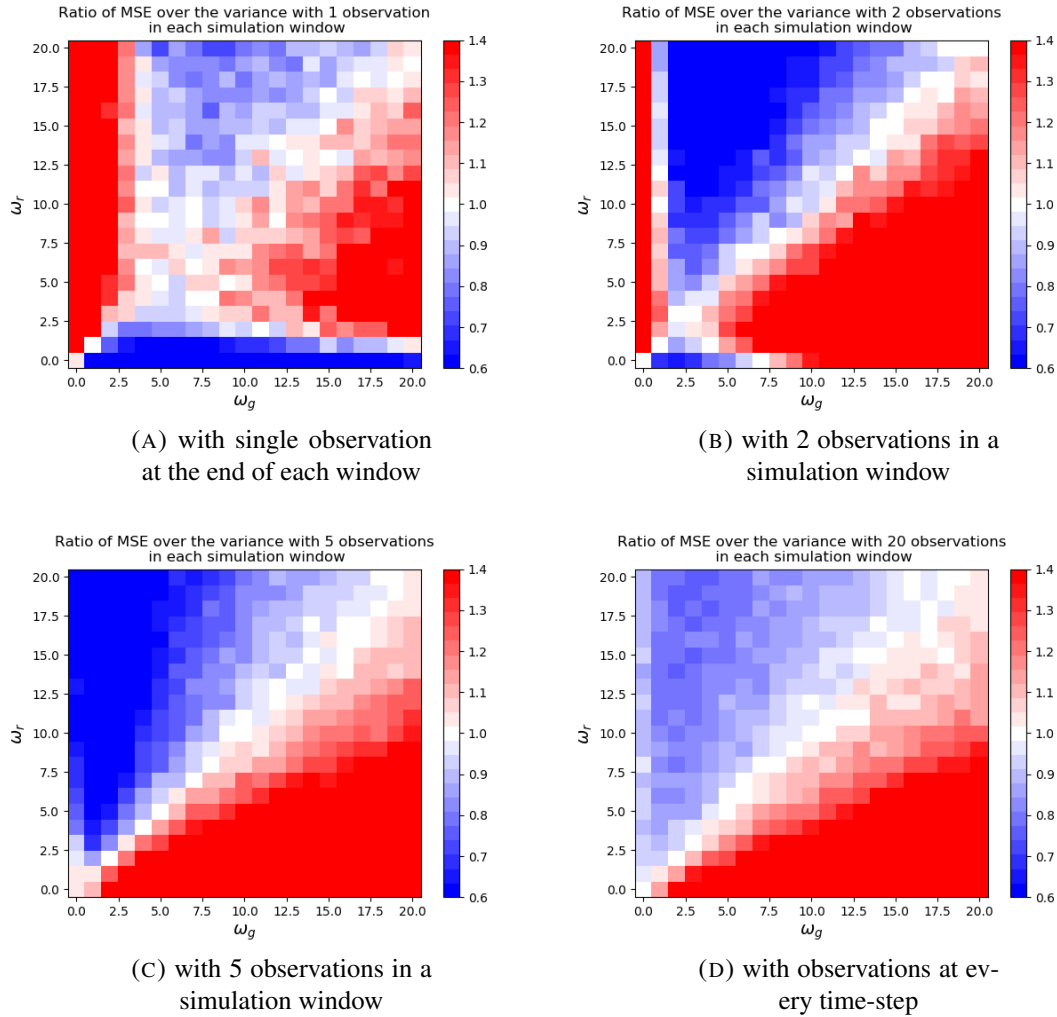
(D) with observations at every time-step

FIGURE 3.7. Ratio of MSE over the variance of the posterior for a 10-dimensional system with different observation frequencies. These plots come from numerical experiments with a 200-member EnKS. Note that the results are qualitatively and also quantitatively very similar to those in figure 3.5.

in Lagrangian Data Assimilation (Kuznetsov, Ide, and Jones, 2003). Evensen (2009) uses the EnKF to update the state while performing the parameter estimation.

Even for the simple linear regressive model that we used in the previous section, since the correlation timescale is deeply encoded inside the governing equation of the system, parameter estimation becomes a non-linear problem. As an example of such a correlation

time scale estimation problem we will use state augmentation in an EnKS, in which the time scale is simply added to the state vector.

Instead of the memory timescale, ω^g, we use the log scale of the memory timescale to avoid negative memory estimates. The initial log-timescale values are drawn from a normal distribution: $\ln \omega^g_i \in \mathcal{N}(\ln \omega^g, 1.0)$. Hence we assume that the prior distribution of the memory time scale is lognormal distributed.

The results are shown in Figure 3.8. The top two plots show experiments with only one
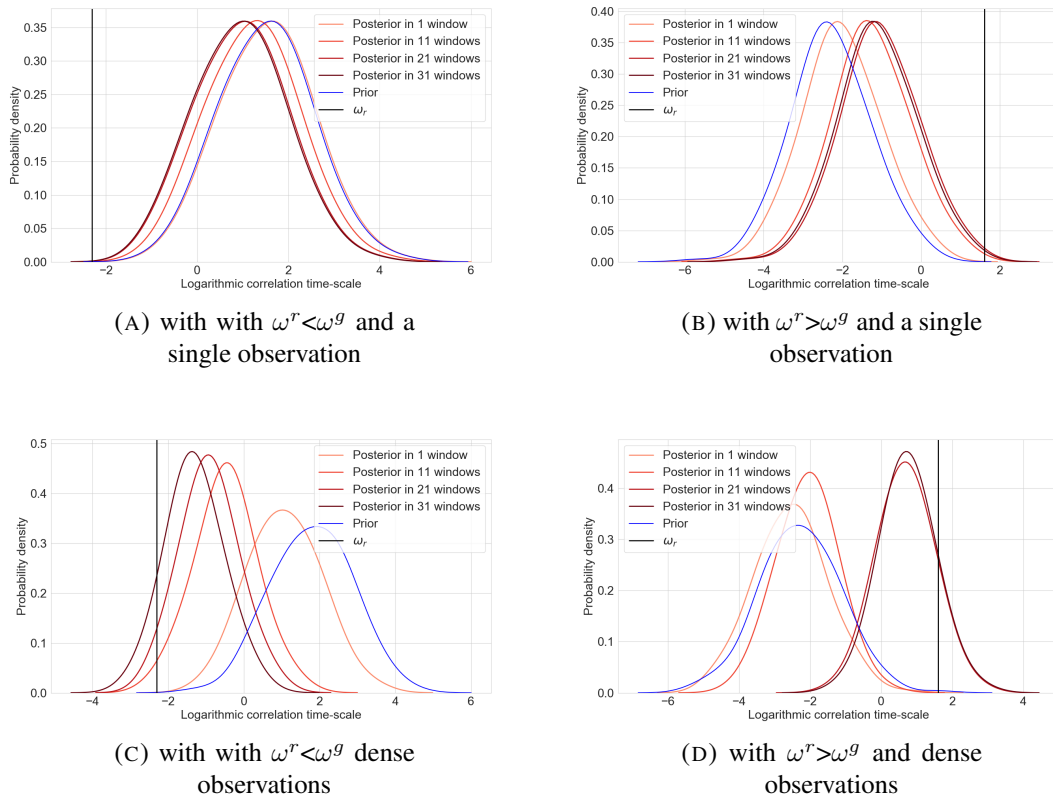

(A) with with $\omega^r < \omega^g$ and a single observation

(B) with $\omega^r > \omega^g$ and a single observation

(C) with with $\omega^r < \omega^g$ dense observations

(D) with $\omega^r > \omega^g$ and dense observations

FIGURE 3.8. PDFs of the prior (blue) and posterior (reddish colors) estimated ω^g, using an increasing number of assimilation windows. The different panels show results for different observation densities and prior mean larger (A,C) or smaller (B,D) than ω^r. The vertical black line denotes the true value ω^r.

observation at the end of the window, in which either our first estimate of the timescale is larger or smaller than the real timescale. With an increasing number of windows we obtain better estimates, but the variance of the estimate does not change. Also, the convergence is slow. We experimented with different values for first guess and true timescales, and in some cases the solution did not converge to the correct value. This is not surprising given the highly nonlinear character of the parameter estimation problem, especially with only one observation per window.

When we observe every time step the convergence is much faster, and the variance in the estimate decreases, as shown in the lower two subplots. In this case we always found fast convergence with different first guess and true timescale combinations, demonstrating that more observations bring us closer to the truth, and hence make the parameter estimation problem more linear.

3.5 Conclusion

In this section we investigated the influence of a miss-specified model error decorrelation time scale in linear models, using an (Ensemble) Kalman Smoother, and investigated estimation of that timescale in an EnKS.

Using a Kalman Smoother, analytical results were derived for the posterior variance and Mean-squared Error (MSE) for a zero-dimensional model. We find that the posterior variance, which only depends on the guessed correlation timescale ω^g, has different behaviour with different observation frequencies. With a single observation, the posterior variance has a maximum at a certain ω^g value, and that maximum and the ω^g value decrease when propaating in time. When we increase the number of observations, the posterior variance becomes a monotonic decreasing function of ω^g.

Since the posterior variance represents the error of the posterior estimated by the data assimilation process, with more information from the observations the estimated error becomes significantly smaller. The MSE, which is the actual error of the posterior mean, decreases as well when more observations are included. But unlike the posterior variance, the mean-square error of the posterior mean does not only depend ω^g, but also on the real memory time scale ω^r. The results for the posterior MSE with a single observation show that it increases with both ω^r and the mismatch between ω^g and ω^r. It means if we don't have a fair estimate of the correlation timescale, the actual posterior error will be larger.

For a higher-dimensional model we used an EnKS. The results agree with the results from the analytical and numerical evaluations of the Kalman Smoother. For many observations we found that the MSE is larger than the estimated error for $\omega^g > \omega^r$, and vice versa. For a low number of observations a new regime appears where for very small ω^r the MSE is smaller than the estimated error, and vice versa for very small ω^g. This behavior is mainly dictated by the behavior of the estimated error.

Since the influence of an incorrect decorrelation timescale in the model error can be significant we investigated the estimation of this timescale within an EnKS. We found that when the observation density is high state augmentation is sufficient to obtain converging results. However, with only one observation in a time window the problem becomes too nonlinear and the estimation process is slow, or does not even converge. These results are consistent with parameter estimation via state augmentation in the literature. The new element is that online estimation is possible beyond a relatively simple bias estimate of the model error.

As a next step we will explore the influence of incorrectly specified model errors in nonlinear systems, with the goal to come up with a robust estimation method for time-correlated model errors.

Chapter 4

Using the (iterative) ensemble Kalman smoother to estimate the time correlation in model error

We further investigate the possibilities of estimating the parameters encoded in the model error autocorrelation with different assumptions of the formulations of the model error autocorrelations, with both linear and nonlinear models. In this chapter, we investigate further for the second research question: *Is it possible to establish a way to update the parameters encoded in the model error autocorrelation online during the data assimilation procedure?* This chapter is reproduced from the paper published on *Tellus A: Dynamic Meteorology and Oceanography* (Amezcua, Ren, and Van Leeuwen, 2023).

4.1 Introduction

Forecast models have often been considered perfect representations of the processes in the real world, the so-called strong constraint in the DA literature. This leaves the uncertainty in initial conditions as the sole culprit for any forecast errors. In reality, however, model error can become as important as initialisation error in degrading forecast accuracy (Orrell et al., 2001). In numerical weather prediction (NWP) systems, model error arises from time and space discretisations, approximate parameterisations of physical processes

that are not represented explicitly, unresolved sub-grid processes, etc. While these errors are often ignored in the DA process, it is common knowledge that more accurate solutions can be obtained if they are included.

The impact of model error caused by unresolved processes on the forecast and DA results can last for several model time steps. Bennett (1992), typically way ahead of his time, extensively discussed the use of correlated model errors and solution of the problem using the representer method, Amezcua and Van Leeuwen (2018) formulated the time-correlated problem for ensemble smoothers, and Evensen (2021) extended this to iterative ensemble smoothers. An obstacle in this endeavor, however, is that it is hard to describe a prior on the model errors, especially if one is to include non-trivial probabilistic elements in both space and time. As a result, there has been interest in estimating model errors in DA schemes in the last two decades (Brasseur et al., 2005; Crommelin and Vanden-Eijnden, 2008; Zhu, Van Leeuwen, and Zhang, 2018; Lucini, Van Leeuwen, and Pulido, 2021; Bonavita and Laloyaux, 2020; Brajard et al., 2021; Pathiraja and Leeuwen, 2017; Evensen, 2021).

It is important to recall that DA has two components: the forecast, where an initial condition is evolved using the dynamical model, and the analysis, where forecast and observations from the truth are combined (e.g. Asch, Bocquet, and Nodet, 2016). Including model error in DA, therefore, requires actions in both steps. In the forecast, it is necessary to simulate the action of model error in the evolution of model trajectories. This can be by inflating the forecast error covariance in a single-trajectory setting, or by explicitly including realisations of the model error in an ensemble setting like the ensemble Kalman filter (Evensen, 1994). In the analysis step, one has to be able to update the model error parameters (or actual realisation values), i.e. go from background to analysis values. This task is often more complicated, and thus is often skipped, or done in an approximate

manner (e.g. Howes, Fowler, and Lawless, 2017). Using the wrong model error parameters, however, decreases the ability of the analysis step to obtain correct updated values for the state variable. This was shown for time-correlated model error in Amezcua and Van Leeuwen (2018) and Ren, Amezcua, and Van Leeuwen (2021).

Let us provide a simple illustration of the way time auto-correlated model error arises. Consider the time evolution of the logistic map with coefficient $\gamma_{true} = 3.75$, which is shown in Figure 4.1. This zero-dimensional non-linear discrete map was made popular
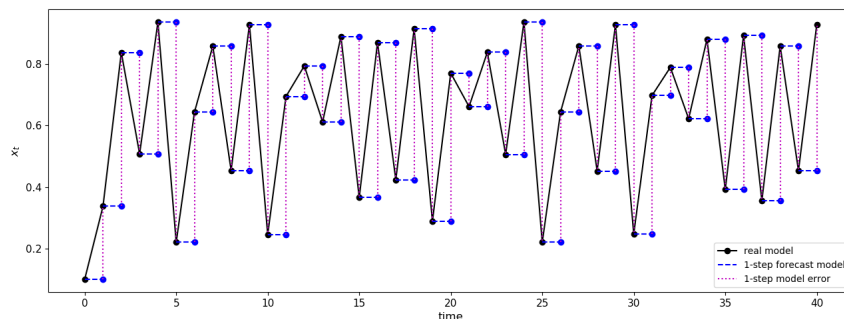

FIGURE 4.1. Simple illustration for the origin of auto-correlated model error. The system evolves under a real system represented by the logistic map (black line). 1-lag forecasts are produced with an imperfect model (blue line), persistence. The 1-lag model errors are computed by taking the differences of the two values (dashed magenta lines).

by May (1976), and is described with more detail in section 3. The nature (or true) evolution is shown by the black solid lines and the black dots. Now, consider an imperfect forecast model. The simplest one we can think of is persistence, i.e. no evolution over one model time step. If we take the exact values of the nature run at every model time step, and evolve them with this imperfect forecast model over one step, we obtain the blue lines. Taking the difference between the true value of the variable and the imperfect 1-step forecast renders the dashed magenta lines. These are 1-step model error values diagnosed offline, and therefore useful only retrospectively.

To characterise this model error, we run the described process for a long time period (10^4 model steps), save the model error values for each time step, and compute statistics on these values (similar to Evensen and Fario, 1997). Figure 4.2 shows three statistics: mean (left panel), standard deviation (centre panel), and lag-1 auto-correlation (right panel).
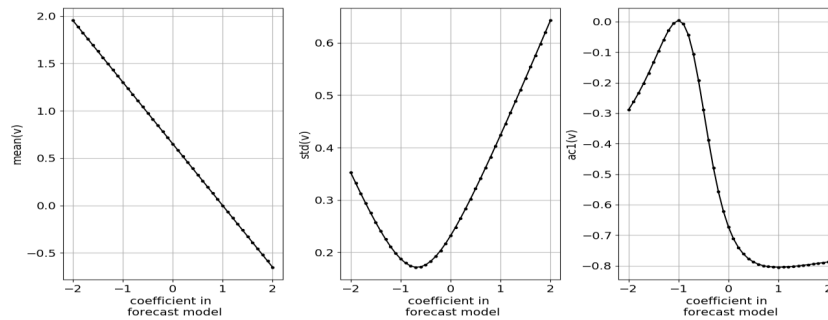

FIGURE 4.2. Model error statistics (mean in left panel, standard deviation in centre panel, lag-1 auto-correlation in the right panel). These statistics are computed off-line, after a long model run, in the way illustrated in figure 4.1. These are computed for different coefficients in the forecast model (linear) (horizontal axis in panels).

To make the experiment more general, we keep the true model fixed, but in the forecast model we vary the coefficient multiplying the value in the previous time step (persistence is when the coefficient is one). In each panel, the horizontal axis corresponds to values of this coefficient, while the vertical axis corresponds to the value of the statistic. The three statistics vary as function of the coefficient in the forecast model. Moreover, the lag-1 auto-correlation shows a non-trivial behaviour, and is zero in only one occasion. Hence, only for a very particular coefficient choice the model error is independent in time, and in general it is not. This model error has been diagnosed offline. The question we aim to answer is whether it is possible to obtain these estimates online.

This chapter has the specific objective to perform online estimation of the time-related characteristics of this model error. Our goal is to improve the accuracy of the forecast by sequentially updating the error using an ensemble Kalman-based method (Kalman,

1960; Kalman and Bucy, 1960). In particular, we use the Ensemble Kalman Smoother (Evensen and Van Leeuwen, 2000; Evensen, 2018; Amezcua and Van Leeuwen, 2018). As a smoother, instead of only updating fields at observation time, it updates the whole trajectory over a simulation window using all available observations in that window. The EnKS uses ensemble integrations to approximately represent the density for the prior model evolution. This ensemble is then used to solve the DA problem under the Gaussian assumption for model states, parameters and observation errors, and an observation operator that does not deviate too much from linear. Ren, Amezcua, and Van Leeuwen (2021) performed parameter estimation for the model error with spatial and temporal autocorrelation using an EnKS, and, while successful in some cases, the parameter estimation failed in others. It was argued that the failures were due to the linear correlations that are assumed in an EnKS. We will investigate this claim by using a nonlinear iterative EnKS (IEnKS) (Sakov, Oliver, and Bertino, 2012; Evensen, 2018). The IEnKS can be regarded as an ensemble variational method that does not require the tangent linear of the evolution and observation models, nor the adjoint of these models (Bocquet and Sakov, 2014).

This chapter is organised as follows. In section 2 we show how the problem of jointly estimating state variables and time-related parameters in the model error is a difficult one. We show that even when the state variables follow linear dynamics, the time-related model error parameters appear in a very non-linear fashion within the associated cost function. This complicates the estimation problem, and makes the minimum variance and maximum a posteriori solutions differ considerably. In practice, it requires the use a non-linear DA method such as the IEnKS for the joint estimation problem. In section 3 we describe our experimental setup. We choose two types of memory for the model error: one with pure exponential decay, and one with a mixed oscillatory-exponential decay. In this section we also describe the two models used in our experiments: the simple model of section 2, and also the logistic map, which provides insight into what happens with non-linear dynamics. Results of these experiments are presented and discussed in section

4. Finally, section 5 provides a summary and discussion of the work.

4.2   Estimating auto-correlated model error in a simple linear model

Estimating model error is a difficult problem, especially when these errors are correlated in time. We explore a systematic approach in which we aim to estimate parameters of the time-correlation part of the model error formulation. Specifically, we assume a model error structure that is separable in space and time. We parameterise the time-dependent behaviour with decaying and oscillatory parameters, and then use use DA to estimate them.

This section illustrates how the parameters related to time-dependent model error are involved in a nonlinear manner in the imperfect model evolution. Therefore, the online estimation of these parameters becomes a challenge for Bayesian estimation. For simplicity, let us consider zero-dimensional systems, which serves two purposes. First, it allows for analytic steps to be feasible and provide important insight. Second, we do not need to worry about spatial structures in the model error and can focus on the time structures.

In the following we first formulate the problem of jointly estimating state variables and parameters, followed by the full Bayesian problem set up. We show that, because the time-related parameters in the model error are related nonlinearly to the state variables, the parameter estimation problem is always nonlinear, *even when the prior on the para-maters is Gaussian and the model is linear in the state*. Then we formulate two solutions, one that finds the mode of the posterior, and the other first linearizes the problem and then finds the mean. The solution without linearization will be more accurate, but also much more computationally expensive as it relies on an iterative procedure that employs

the adjoint model, and does not provide an error estimate. The solution to the linearized problem is computationally much more efficient, but differs significantly from the solution to the nonlinear problem. We then formulate an iterative ensemble method, which is a more accurate solution to the full nonlinear problem. No explicit solution exists for this ensemble method, but the method is computationally efficient. This is the method we use in our numerical experiments we perform to achieve online estimation.

4.2.1 Problem formulation

Let $x \in \Re$ be the state variable of our system with initial conditions x_0 at $t = 0$. Consider τ independent model error jumps ϵ_t, for $t = \{1, \cdots, \tau\}$ with zero mean. We denote the control variable as the column vector $\mathbf{z} \in \Re^{\tau+1}$, with background distribution $N\left(\boldsymbol{\mu}^{b,z}, \mathbf{D}^z\right)$:

$$\mathbf{z} = [x_0, \epsilon_1, \cdots, \epsilon_\tau]^{\mathbf{T}}, \tag{4.1a}$$

$$\boldsymbol{\mu}^{b,z} = \left[\mu_0^b, 0, \cdots, 0\right]^{\mathbf{T}}, \tag{4.1b}$$

$$\mathbf{D}^z = \text{diag}\left[b^2, q^2, \cdots, q^2\right], \tag{4.1c}$$

where $\boldsymbol{\mu}^{b,z} \in \Re^{\tau+1}$ and $\mathbf{D}^z \in \Re^{(\tau+1)\times(\tau+1)}$ are the background mean vector and background error covariance matrix, respectively. The scalars b^2 and q^2 represent the background and model error variances. Note that $\mathbf{z}$ has a diagonal covariance matrix since we consider statistical independence amongst its elements. We use these elements to construct a simple linear system with time auto-correlated model errors in its evolution. The real linear evolution over one time step is prescribed by:

$$x_{t+1} = \alpha x_t + v_{t+1}, \tag{4.2}$$

i.e. a simple auto-regressive component plus a model error realisation, where the autocorrelated model error can be generated as:

$$v_{t+1} = \phi\left(1, \theta\right) v_t + \sqrt{1 - \phi\left(1, \theta\right)^2} \epsilon_{t+1}, \qquad (4.3)$$

in which $\phi\left(1, \theta\right)$ represents the lag-1 autocorrelation of the model error. This model error has distribution $v(t) \sim N\left(0, q^2\right)$, and the following structure in time:

$$\text{Corr}(v_t, v_{t'}) = \phi\left(|t - t'|, \theta\right). \qquad (4.4)$$

Eq. (4.4) indicates that the model errors are auto-correlated in time, and this only depends on the lag $|t - t|'$ and a vector of N_θ parameters $\theta \in \Re^{N_\theta}$. For τ time steps, this yields an auto-correlation matrix $\boldsymbol{\Phi} \in \Re^{\tau \times \tau}$. This symmetric Toeplitz matrix has the following elements:

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & \phi(1) & \cdots & \phi(\tau - 2) & \phi(\tau - 1) \\ \phi(1) & 1 & \cdots & \phi(\tau - 3) & \phi(\tau - 2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi(\tau - 2) & \phi(\tau - 3) & \cdots & 1 & \phi(1) \\ \phi(\tau - 1) & \phi(\tau - 2) & \cdots & \phi(1) & 1 \end{bmatrix}. \qquad (4.5)$$

Being a positive-definite symmetric matrix, a Cholesky decomposition $\boldsymbol{\Phi}$ is possible, i.e.:

$$\boldsymbol{\Phi} = \mathbf{L}\mathbf{L}^{\mathbf{T}}, \qquad (4.6)$$

where $\mathbf{L} \in \Re^{\tau \times \tau}$ is lower triangular.

We now take Eq. (4.2) and write the time evolution of the system from $t = 0$ to $t = \tau$ in terms of the control vector $\mathbf{z}$ and the Cholesky factor $\mathbf{L}$. Explicitly:

$$x_\tau = \mathbf{M}_{0:\tau} \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{L} \end{bmatrix} \mathbf{z}, \qquad (4.7)$$

where $\mathbf{M}_{0:\tau} \in \Re^{1 \times (\tau+1)}$ is a row matrix with the model evolution from 0 to any time t. For this simple model, the elements are decreasing powers of α:

$$\mathbf{M}_{0:\tau} = \left[\alpha^{\tau}, \alpha^{\tau-1}, \cdots, 1 \right]. \tag{4.8}$$

Let us define the composed evolution matrix $\hat{\mathbf{M}} \in \Re^{1 \times (\tau+1)}$:

$$\hat{\mathbf{M}} = \mathbf{M}_{0:\tau} \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{L} \end{bmatrix}. \tag{4.9}$$

This includes the effect of the deterministic dynamics and the auto-correlation of the model error, and it will become useful in the next subsections. Separating the initial condition and the model errors, we write Eq. (4.7) as:

$$x_{\tau} = \alpha^{\tau} x_0 + \mathbf{M}_{1:\tau} \mathbf{L} \boldsymbol{\epsilon}, \tag{4.10}$$

where $\mathbf{M}_{1:\tau} \in \Re^{1 \times \tau}$ and $\boldsymbol{\epsilon} \in \Re^{\tau}$ are obtained by removing the first element of $\mathbf{M}_{0:\tau}$ and $\mathbf{z}$ respectively. In fact, we can recover the value x_t at any time (not just the final time) using (4.10) with the corresponding truncated elements in the second term.

We consider an observation of the truth at the end of the forecast window, i.e. at time $t = \tau$. For simplicity, let the observation operator be the identity, so the observation equation is:

$$y = x_{\tau} + \eta, \tag{4.11}$$

with the observation error: $\eta \sim N(0, r^2)$. Obtaining the analysis values for $\mathbf{z}$ was already discussed in Amezcua and van Leeuwen (2018), and Ren et al (2020). In this chapter we discuss the solution of the joint state-variable estimation problem.

4.2.2 Bayesian solution for the joint state-parameter estimation

Considering both the control variable $\mathbf{z}$ and parameters $\boldsymbol{\theta}$ to be random variables, the Bayesian solution of this problem is to obtain the posterior joint pdf of $\mathbf{z}$ and $\boldsymbol{\theta}$ given the observation y. Namely,

$$\mathrm{p}(\mathbf{z}, \boldsymbol{\theta}|y) = \frac{\mathrm{p}(y|\mathbf{z}, \boldsymbol{\theta})\mathrm{p}(\mathbf{z}, \boldsymbol{\theta})}{\mathrm{p}(y)}. \tag{4.12}$$

The numerator is the joint pdf of $\mathbf{z}$, $\boldsymbol{\theta}$ and y. This is obtained as the product of the likelihood of y times the prior joint pdf of $\mathbf{z}$ and $\boldsymbol{\theta}$. If we consider these two to be statistically independent, then we have:

$$\mathrm{p}(\mathbf{z}, \boldsymbol{\theta}) = \mathrm{p}(\mathbf{z})\mathrm{p}(\boldsymbol{\theta}). \tag{4.13}$$

The marginal pdf of the observations is:

$$\mathrm{p}(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{p}(\mathbf{z}, \boldsymbol{\theta}, y)\mathrm{d}\mathbf{z}\mathrm{d}\boldsymbol{\theta}. \tag{4.14}$$

The prior distribution for the control variable and the likelihood are easy to characterise. Recall that we have:

$$\mathbf{z} \sim N\left(\boldsymbol{\mu}^{b,z}, \mathbf{D}^z\right), \tag{4.15a}$$

$$y|\mathbf{z}, \boldsymbol{\theta} \sim N\left(\hat{\mathbf{M}}\left(\boldsymbol{\theta}\right)\mathbf{z}, r^2\right). \tag{4.15b}$$

From now on, we explicitly note that $\hat{\mathbf{M}}\left(\boldsymbol{\theta}\right)$ depends on $\boldsymbol{\theta}$. Note that the joint estimation problem is complicated since we have the product $\hat{\mathbf{M}}\left(\boldsymbol{\theta}\right)\mathbf{z}$ in the likelihood, which limits the possibility of obtaining an analytical expression for the posterior $p\left(\mathbf{x}, \boldsymbol{\theta}|y\right)$. For this reason, we now discuss two solutions based in statistics of this pdf: the maximum-a-posteriori solution, and a popular approximation, the extended Kalman smoother solution.

4.2.3 The Maximum-a-posteriori solution

We can try maximising the joint pdf to obtain a maximum-a-posteriori (MAP) solution. This is equivalent of finding the minimum of the cost function:

$$J(\mathbf{z}, \boldsymbol{\theta}|y)) = -\ln\left(\mathrm{p}(\mathbf{z}, \boldsymbol{\theta}|y)\right). \tag{4.16}$$

Using the distributions in Eq. (4.15) and an arbitrary prior for $\boldsymbol{\theta}$, the minus logarithm of Eq. (4.13) is:

$$J(\mathbf{z}, \boldsymbol{\theta}|y) = \text{constant} + \frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}^{b,z})^{\mathbf{T}}\mathbf{D}^{z^{-1}}(\mathbf{z} - \boldsymbol{\mu}^{b,z}) + \frac{1}{2r^2}(y - \hat{\mathbf{M}}(\boldsymbol{\theta})\mathbf{z})^2 - \ln(\mathrm{p}(\boldsymbol{\theta})). \tag{4.17}$$

The minimisers $\{\mathbf{z}^*, \boldsymbol{\theta}^*\}$ of the cost-function can be obtained by taking the gradient of $J(\mathbf{z}, \boldsymbol{\theta}|y)$ with respect to both control variables and parameters and equating to zero:

$$\begin{bmatrix} \nabla^{\mathbf{z}} J \\ \nabla^{\boldsymbol{\theta}} J \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{4.18}$$

with the gradients $\nabla^{\mathbf{z}} J \in \Re^{\tau+1}$ and $\nabla^{\boldsymbol{\theta}} J \in \Re^{N_\theta}$. If we also assume that the parameters follow a MVG –i.e. $\boldsymbol{\theta} \sim \left(\boldsymbol{\mu}^{\boldsymbol{\theta}}, \mathbf{D}^{\boldsymbol{\theta}}\right)$– we are able to compute the gradients explicitly. This yields the following system of $\tau + 1 + N_\theta$ equations:

$$\mathbf{D}^{z^{-1}}\left(\mathbf{z} - \boldsymbol{\mu}^{b,z}\right) - \frac{1}{r^2}\hat{\mathbf{M}}(\boldsymbol{\theta})^{\mathbf{T}}\left(y - \hat{\mathbf{M}}(\boldsymbol{\theta})\mathbf{z}\right) = \mathbf{0}, \tag{4.19a}$$

$$\mathbf{D}^{\boldsymbol{\theta}^{-1}}\left(\boldsymbol{\theta} - \boldsymbol{\mu}^{b,\theta}\right) - \left(\frac{\partial\hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^{\mathbf{T}}\mathbf{z}\left(\frac{y - \hat{\mathbf{M}}(\boldsymbol{\theta})\mathbf{z}}{r^2}\right) = \mathbf{0}, \tag{4.19b}$$

with the Jacobian matrix $\frac{\partial\hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \in \Re^{(\tau+1)\times N_\theta}$ defined as:

$$\frac{\partial\hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \begin{bmatrix} \nabla^{\boldsymbol{\theta}\mathbf{T}}\hat{\mathbf{M}}_1(\boldsymbol{\theta}) \\ \nabla^{\boldsymbol{\theta}\mathbf{T}}\hat{\mathbf{M}}_2(\boldsymbol{\theta}) \\ \vdots \\ \nabla^{\boldsymbol{\theta}\mathbf{T}}\hat{\mathbf{M}}_{\tau+1}(\boldsymbol{\theta}) \end{bmatrix}. \tag{4.20}$$

One can solve $\mathbf{z}$ from Eq. (4.19a) to get:

$$\mathbf{z} = \boldsymbol{\mu}^{b,z} + \mathbf{K}\left(y - \hat{\mathbf{M}}(\boldsymbol{\theta})\boldsymbol{\mu}^{b,z}\right), \tag{4.21}$$

with $\mathbf{K}$ and γ^2 defined as

$$\mathbf{K} = \frac{1}{\gamma^2}\mathbf{D}^z\hat{\mathbf{M}}(\boldsymbol{\theta})^{\mathbf{T}}, \tag{4.22a}$$

$$\gamma^2 = \hat{\mathbf{M}}(\boldsymbol{\theta})\mathbf{D}^z\hat{\mathbf{M}}(\boldsymbol{\theta})^{\mathbf{T}} + r^2, \tag{4.22b}$$

and similarly from Eq. (4.19b):

$$\boldsymbol{\theta} = \boldsymbol{\mu}^{b,\theta} + \frac{\mathbf{D}^{\theta}}{r^2}\left(\frac{\partial\hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^{\mathbf{T}}\mathbf{z}\left(y - \hat{\mathbf{M}}(\boldsymbol{\theta})\mathbf{z}\right). \tag{4.23}$$

Eq. (4.21) is the Kalman analysis equation for the posterior mean (Kalman, 1960; Kalman and Bucy, 1961). The complication, however, is in our case $\boldsymbol{\theta}$ is an unknown. Hence, Eq. (4.21) needs to be solved in tandem with Eq. (4.23), which cannot be done analytically in general.

To actually calculate the derivative of the model with respect to the parameters, we note the following. For the j_{th} parameter θ^j, Eq. (4.20) can be readily computed using the Cholesky factor $\mathbf{L}$ defined in Eq. (4.6) in the following manner:

$$\left(\frac{\partial\hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial\theta^j}\right)^{\mathbf{T}} = \mathbf{M}_{0:\tau}\begin{bmatrix} 1 & 0 \\ 0 & \frac{\partial\mathbf{L}}{\partial\theta^j} \end{bmatrix}. \tag{4.24}$$

The derivative of the Cholesky matrix can be found using Theorem A.1 of Särkkä (2013):

$$\frac{\partial\mathbf{L}}{\partial\theta^j} = \mathbf{L}\mathbf{T}\left(\mathbf{L}^{\mathbf{T}}\frac{\partial\boldsymbol{\Phi}}{\partial\theta^j}\left(\mathbf{L}^{-1}\right)^{\mathbf{T}}\right), \tag{4.25}$$

where the matrix $\mathbf{T} \in \Re^{\tau \times \tau}$ is defined as:

$$
\mathbf{T}^{ij}(\mathbf{A}) = \begin{cases} \mathbf{A}^{ij} \text{ if } i > j, \\[2mm] \frac{1}{2}\mathbf{A}^{ij} \text{ if } i = j, \\[2mm] 0 \text{ if } i < j. \end{cases} \tag{4.26}
$$

4.2.4 The extended Kalman Smoother solution

The extended Kalman Smoother solution can be derived directly from the MAP solution by centering the derivative of the model to the state and the parameters on the background values. Tracing back these derivatives we can rewrite Eq. (4.21) and Eq. (4.23) as:

$$
\mathbf{z} = \boldsymbol{\mu}^{b,z} + \frac{1}{\gamma^2}\mathbf{D}^z\hat{\mathbf{M}}(\boldsymbol{\mu}^{b,\theta})^{\mathbf{T}}\left(\hat{\mathbf{M}}(\boldsymbol{\mu}^{b,\theta})\mathbf{D}^z\hat{\mathbf{M}}(\boldsymbol{\mu}^{b,\theta})^{\mathbf{T}} + r^2\right)^{-1}\left(y - \hat{\mathbf{M}}(\boldsymbol{\theta})\boldsymbol{\mu}^{b,z}\right), \tag{4.27}
$$

$$
\boldsymbol{\theta} = \boldsymbol{\mu}^{b,\theta} + \frac{\mathbf{D}^\theta}{r^2}\left(\left.\frac{\partial\hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\mu}^{b,\theta}}\right)^{\mathbf{T}}\boldsymbol{\mu}^{b,z}\left(y - \hat{\mathbf{M}}(\boldsymbol{\theta})\mathbf{z}\right). \tag{4.28}
$$

Furthermore, in Eq. (4.28) we use a first-order Taylor expansion for $\hat{\mathbf{M}}(\boldsymbol{\theta})$ as:

$$
\hat{\mathbf{M}}(\boldsymbol{\theta}) = \hat{\mathbf{M}}\left(\boldsymbol{\mu}^{b,\theta}\right) + \left(\left.\frac{\partial\hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\mu}^{b,\theta}}\right)^{\mathbf{T}}\left(\boldsymbol{\theta} - \boldsymbol{\mu}^{b,\theta}\right). \tag{4.29}
$$

We can solve then solve Eq. (4.27) and Eq. (4.28) by introducing a new augmented variable $\tilde{\mathbf{z}} = \left(\mathbf{z}^{\mathbf{T}}, \boldsymbol{\theta}^{\mathbf{T}}\right)^{\mathbf{T}}$ and after some algebra we obtain:

$$
\boldsymbol{\mu}^{a,\tilde{z}} = \boldsymbol{\mu}^{b,\tilde{z}} + \tilde{\mathbf{K}}\left(y - \hat{\mathbf{M}}(\boldsymbol{\mu}^{\theta,b})\boldsymbol{\mu}^{b,z}\right), \tag{4.30a}
$$

$$
\mathbf{A}^{\tilde{z}} = \mathbf{D}^{\tilde{z}} - \tilde{\mathbf{K}}\left[\hat{\mathbf{M}}\left(\boldsymbol{\mu}^{b,\theta}\right), \left(\boldsymbol{\mu}^{b,z}\right)^{\mathbf{T}}\left.\frac{\partial\hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\mu}^{b,\theta}}\right]\mathbf{D}^{\tilde{z}}, \tag{4.30b}
$$

where $\mathbf{A}^z$ is denoted as the posterior variance in model error space. In this case, the augmented gain $\tilde{\mathbf{K}} \in \Re^{(\tau+1+N_\theta)\times 1}$ and augmented residual variance $\tilde{\gamma}^2 \in \Re_{\geq 0}$ are different from Eq. (4.22)

$$\tilde{\mathbf{K}} = \frac{1}{\tilde{\gamma}^2} \begin{bmatrix} \mathbf{D}^z \hat{\mathbf{M}}(\boldsymbol{\mu}^{b,\theta})^{\mathbf{T}} \\ \mathbf{D}^\theta \left(\left. \frac{\partial \hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\mu^{b,\theta}} \right)^{\mathbf{T}} \boldsymbol{\mu}^{b,z} \end{bmatrix}, \tag{4.31a}$$

$$\tilde{\gamma}^2 = \mathbf{M}(\boldsymbol{\mu}^{b,\theta}) \mathbf{D}^z \mathbf{M}(\boldsymbol{\mu}^{b,\theta})^{\mathbf{T}} + \left(\boldsymbol{\mu}^{b,z}\right)^{\mathbf{T}} \left. \frac{\partial \hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\mu^{b,\theta}} \mathbf{D}^\theta \left(\left. \frac{\partial \hat{\mathbf{M}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\mu^{b,\theta}} \right)^{\mathbf{T}} \boldsymbol{\mu}^{b,z} + r^2. \tag{4.31b}$$

In the last step we have again assumed that state variables and parameters are statistically independent in the prior. Notice that *even with a Gaussian prior $p(\boldsymbol{\theta})$ and a model that is linear in the state $\mathbf{z}$ the MAP and the extended KS solution can differ considerably.* The MAP solution results in an implicit equation that needs to be solved iteratively. However, even if we manage to do this an uncertainty estimate is still lacking. It is possible, if the system is low-dimensional, to solve this problem with e.g. a particle filter. Unfortunately, for higher dimensions particle filters are degenerate, and a potential solution for that, localization, is difficult to use with global parameters. Instead, we resort to a popular approximate solution based on an iterative Ensemble Kalman smoother that can easily be scaled up to higher dimensions.

4.2.5   The (Iterative) Ensemble Kalman Smoother

The Iterative Ensemble Kalman Smoother (IEnKS) does solve the parameter estimation problem approximately, but it does introduce nonlinearity in the solution procedure. Hence, it will provide a solution that is closer to the MAP than the Extended Kalman Smoother solution. In an IEnKS parameter estimation is typically performed via state

augmentation, as in the extended Kalman Smoother. We define a variable $\mathbf{u} \in \Re^{1+\tau+N_\theta}$ as:

$$\mathbf{u} = [x_0, x_1, \cdots, x_\tau, \theta^1, \cdots, \theta^{N_\theta}]^{\mathbf{T}}, \tag{4.32}$$

where the augmented variable includes the state variable at all time steps and all the model error parameters, sometimes called the state formulation. Notice that this is slightly different than Eq. (4.1), where the augmented variable contains the initial conditions, the model error jumps, and the model error parameters, leading to the so-called forcing formulation. In a linear system, it is trivial to go from one formulation to the other, as explained in Tremolet (2006) and Amezcua and Van Leeuwen (2018). For ensemble implementations it is easier to work with state variables directly, so this is how we implement the experiments in the chapter.

Ren, Amezcua, and Van Leeuwen (2021) studied the consequences of wrongly prescribing the model error decorrelation time scale, which was the only parameter. This is, they performed DA with a model error auto-correlation which was different from the one 'guessed' in the ensemble forecast. In the last part of the work, they used state augmentation in an EnKS –as formulated in Amezcua and Van Leeuwen (2018)– to estimate one time parameter in the model error. The authors did not, however, perform a systematic analysis with different types of model error structure. In this work we perform a deeper analysis and more extensive experiments.

In this work we perform state augmentation experiments using two DA methods, the stochastic EnKS and its iterative variant (IEnKS), as formulated in Evensen (2018). This formulation aims to find the minimum of a cost function using ensemble members, either in one step (EnKS) or multiple steps (IEnKS). In the IEnKS a cost function is defined for each ensemble member e by rewriting the cost function Eq. (4.17) in terms of the

extended variable $\mathbf{u}$ as:

$$J^e(\mathbf{u}) = \text{constant} + \frac{1}{2}(\mathbf{u} - \boldsymbol{\mu}^{\mathbf{u},e})^{\mathbf{T}}\mathbf{D}^{\mathbf{u}-1}(\mathbf{u} - \boldsymbol{\mu}^{\mathbf{u},e}) + \frac{1}{2r^2}\left(y^e - \hat{\mathbf{M}}(\mathbf{u})\,\mathbf{u}\right)^2, \quad (4.33)$$

in which $\boldsymbol{\mu}^{u,e}$ is the prior ensemble member e, and

$$y^e = y + \eta^e, \tag{4.34}$$

is a perturbed observation with the perturbation drawn from the observation error pdf, so $\eta^e \sim N(0, r^2)$. In Eq. (4.33) it is clear that the product $\hat{\mathbf{M}}(\mathbf{u})\mathbf{u}$ is responsible for the difficulty of this problem. Evensen (2018) proposed a simple Gauss-Newton iteration for each ensemble member as:

$$\mathbf{u}^{i+1,e} = \mathbf{u}^{i,e} - \delta\mathbf{C}^{-1}\left.\nabla J^e(\mathbf{u}^{i,e})\right|_{\mathbf{u}^{i,e}}, \tag{4.35}$$

where i is the iteration index, $\mathbf{C} \in \Re^{(1+\tau+N_\theta)\times(1+\tau+N_\theta)}$ is a symmetric approximation of the Hessian of the cost function in which the second derivative of $\hat{\mathbf{M}}(\mathbf{u})\,\mathbf{u}$ is neglected, so

$$\mathbf{C} = \mathbf{D}^{u-1} + \frac{1}{r^2}\left(\mathbf{u}^{\mathbf{T}}\frac{\partial\hat{\mathbf{M}}(\mathbf{u})}{\partial\mathbf{u}} + \hat{\mathbf{M}}(\mathbf{u})\right)^{\mathbf{T}}\left(\mathbf{u}^{\mathbf{T}}\frac{\partial\hat{\mathbf{M}}(\mathbf{u})}{\partial\mathbf{u}} + \hat{\mathbf{M}}(\mathbf{u})\right), \tag{4.36}$$

and $\nabla J^e(\mathbf{u}) \in \Re^{1+\tau+N_\theta}$ is the gradient of the cost function for ensemble member e. The iteration step length δ is tunable. The iteration is started at $\mathbf{u}_0^e = \boldsymbol{\mu}^{u,e}$.

The simple descent algorithm can be used with the ensemble gradient and the ensemble approximated Hessian information, as shown in equation (36) of Evensen (2018). For efficiency, we use the particular implementation described in Evensen et al. (2019). Tuning factor δ is an art. In our implementation we reduce the step size the closer we get to the minimum via:

$$\delta^{i+1} = c\delta^i, \tag{4.37}$$

with $0 < c < 1$. The stochastic EnKS can be obtained from this formulation by using only one iteration and setting $\delta = 1$.

4.3 Experimental setup

For our experiments, we use two types of model error (with one and two parameters), and two types of evolution models (linear and nonlinear).

4.3.1 Model error formulations

The first type of model error was used both in Amezcua and Van Leeuwen (2018) and Ren, Amezcua, and Van Leeuwen (2021). The correlation of model error between two time steps is:

$$\text{Corr}(v_t, v_{t'}) = \exp\left[-\frac{|t - t'|}{\omega}\right].$$ (4.38)

In this case, the auto-correlation decays exponentially, and it only has one parameter – the decay time scale ω ($\omega > 0$). When ω tends to 0 the model error becomes a white-noise process. When ω tends to infinity, the model error becomes fixed (a bias). Summarizing:

$$\text{Corr}(v_t, v_{t'}) \rightarrow \begin{cases} 1, & \text{if } t = t' \text{ or } \omega \rightarrow \infty \\ 0, & \text{if } |t - t'| \rightarrow \infty \text{ or } \omega \rightarrow 0. \end{cases}$$ (4.39)

The second model error structure we explore contains both decaying and oscillatory elements:

$$\text{Corr}(v_t, v_{t'}) = \phi^{|t-t'|} \cos(2\pi f |t - t'|).$$ (4.40)

The first term is a geometric memory term, with base $-1 < \phi < 1$, which can be considered a decay parameter. The second term is oscillatory with frequency f. To avoid instability of the system, the decay parameter ϕ is bounded as $0 \leq \phi \leq 1$, and the frequency f is bounded by $0 < f < 0.5$. When the frequency is 0.5, the covariance becomes purely decaying with time and it becomes purely oscillatory when the decaying parameter tends to 1.0. Summarising:

$$\text{Corr}(v_t, v_{t'}) = \begin{cases} \phi^{|t-t'|}, & \text{if } f = 0.5 \\ \cos(2\pi f |t - t'|), & \text{if } \phi = 1. \end{cases} \tag{4.41}$$

4.3.2   Evolution models

To illustrate the issues that arise when estimating model error parameters that are related to temporal correlations, we restrict our experimental set up to two zero dimensional models. The first is the simple linear scalar model used for the analysis in section 2. The true model is:

$$x_{t+1} = \alpha x_t + v_{t+1}. \tag{4.42}$$

Note that this true model is a stochastic model and we assume $v \sim N(0, q^2)$. The true model uses a real decay timescale ω^r, while the forecast model assumes a guessed time auto-correlation parameter ω^g. This value needs to be updated in the DA cycle.

The second set of experiments is much more ambitious since we use a non-linear map as the real system. In particular, we use the well-known logistic map (May, 1976):

$$x_{t+1}^r = \gamma x_t^r (1 - x_t^r), \tag{4.43}$$

where γ is the parameter that gives rise to stable solutions when $0 \leq \gamma \leq 4$; by stable

we mean solutions that do not leave the interval $[0, 1]$. In our experiments we choose $\gamma = 4$ which puts the model in a chaotic regime. Note that in these experiments, the true model is deterministic. This seemingly simple non-linear map is non-invertible, with initialisation errors growing and saturating quickly. It has therefore been used before to test the performance of different configurations of the EnKF (Mitchell and Houtekamer, 2009). The true autocorrelation of the model error (computed offline as described in the introduction) has both positive and negative values, while decreasing in absolute magnitude as a function of the lag. Therefore, a 2-parameter memory is more appropriate for this case, although it increases the difficulty of the problem by having to estimate an extra parameter.

As a forecast model, we propose a stochastic linear model

$$x_{t+1} = \gamma^g x^r + v^{t+1}, \tag{4.44}$$

where γ^g is a prescribed damping coefficient for the forecast model, with its value within $0 < \gamma_g < 1$ to keep the model stable. The model error has the same properties as in the experiments where the real model is linear. In section 1, we described an off-line process to compute the statistics of this model error from a long-time collection of differences between perfect and imperfect forecasts. What we look for, instead, is to estimate the time-related parameters online, along with the state estimation.

4.3.3 Implementation details

Our experiments follow the fraternal (non-identical) twin set up. A real initial condition is prescribed and the true model is run for a long time. In the case of the true stochastic linear model shown in Eq. (4.2), a particular realisation of the model error is used for this true run. For this model we use $\alpha = 0.8$ for all experiments. Synthetic observations

are obtained at different times by applying Eq. (4.11) with $r^2 = (0.01)^2$. We generate synthetic observations every model time step, although we use different subsets in different experiments. The ensemble forecast is generated by selecting N_e initial conditions coming from $\mathcal{N}(0, b^2)$, with $b^2 = (0.1)^2$. The model error realisations are generating using Eq. (4.10). This is, uncorrelated model errors are generated from $\mathcal{N}(0, q^2)$, with $q^2 = (0.1)^2$. The time auto-correlation is inserted by computing the desired Cholesky factor $\mathbf{L}(\boldsymbol{\theta})$ which depends in the parameter vector $\boldsymbol{\theta}$, performing the operation indicated in Eq. (4.10). To separate the true system with the forecast model, the trick is to use a guessed parameter in the forecast model and data assimilation process, which is different from the one used in the true run. For model error parameters we have the following:

- For the one-parameter model error, the prior distribution is $\omega^b \sim N(\mu^{b,\omega}, \sigma^{\omega 2})$ with $\mu^{b,\omega} = 0.3$ and $\sigma^{\omega 2} = (0.5)^2$.

- For the two-parameter model error, we consider the two parameters to be uncorrelated. Their marginal background distributions are: $\phi^b \sim N(\mu^{b,\phi}, \sigma^{\phi 2})$ with $\mu^{b,\phi} = 0.3$ and $\sigma^{\phi 2} = (0.2)^2$, and $f^b \sim N(\mu^{b,f}, \sigma^{f 2})$ with $\mu^{b,f} = 0.3$ and $\sigma^{f 2} = (0.2)^2$.

Both the EnKS and IEnKS use stochastic implementations, (Van Leeuwen, 2020) i.e. one needs to add observational noise to observation values predicted by each ensemble member. In this way the real observation and the predicted observations are generated in the same way. In all our experiments, we use an assimilation window of $\tau = 20$ time steps plus the initial condition. When we assimilate 1 observation per window, this is located at the end time. When we assimilate 20 observations per window, these correspond to all time steps except that of the initial conditions. In order to avoid the introduction of sampling errors –which would be difficult to disentangle from the other effects we study–, we use a relatively large ensemble with $N_e = 200$. This size is kept constant for all experiments.

For the IEnKS, we vary the number of iterations and the step length δ. We consider the IEnKS to have converged to a solution when the difference $|\mathbf{u}^{i+1} - \mathbf{u}^i| < 0.01$, which is approximately 1% of the range of the parameters.

4.4 Experimental results

4.4.1 Illustration of state estimation

In the first set of experiments we aim to illustrate the performance of the EnKS for state estimation. To this end, we use the 1-parameter memory setting and choose background values very close to the real parameter, both the linear and non-linear models. The results of these experiments are displayed in Figure 4.3. In all panels, the truth is shown with a black thick line, randomly selected background ensemble members in blue, and randomly selected analysis members in red. For the analysis, the thick red line corresponds to the mean. We do not show all members to avoid visual cluttering. For this experiment, we only study a single assimilation window.

The top row of this figure shows the results with the linear model. In this case, the EnKS is capable of recovering an analysis mean quite close to the truth. The availability of more observations makes an important difference, as noted in the results shown in the two columns (left shows 1 observation per window, right shows 20 observations per window). In this simple linear case, even with very sparse observations. the EnKS can still provide a fair estimation of the state.

The bottom row shows the results with the logistic map. In the case with a single observation at the end of the simulation window, the EnKS was unsuccessful, which is expected considering the length of the assimilation window, and that joint estimation is being performed. To obtain a better results, more information from the observations is
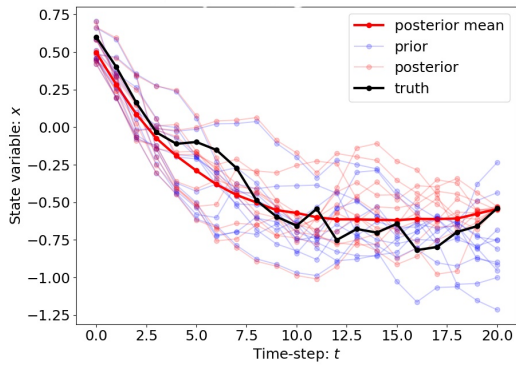
(A) Linear model, 1 observation per window
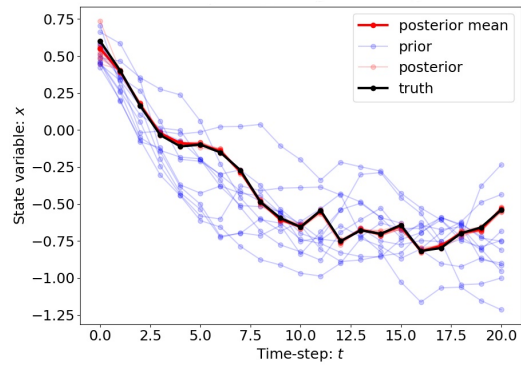
(B) Linear model, 20 observations per window
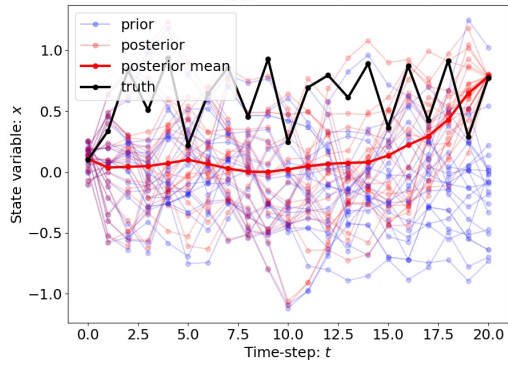
(C) Logistic map, 1 observation per window
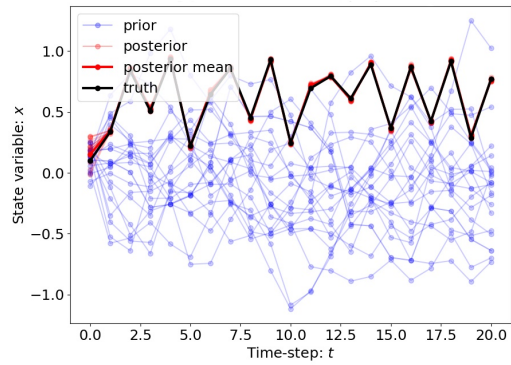
(D) Logistic map, 20 observations per window

FIGURE 4.3. State update for both linear model and logistic map with different number of observations over a single simulation window using EnKS.

needed. With dense observations, observing every time-step for instance, the EnKS provides a correct estimation of all the state values at all time steps.

4.4.2 Parameter estimation in the linear model

4.4.2.1 One-parameter estimation

Our first experiments with the linear model consist of trying to estimate the parameter ω via the EnKS, using the one-parameter autocorrelation shown in Eq. (4.38). The results are shown in Figure 4.4. This figure has four panels. Panels (A), (B) and (C) have the same format, the difference amongst the panels is the number of observations per assimilation window. In each panel, the black vertical line shows the true value of the parameter. Different pdfs are displayed, computed as histograms (smoothed using Kernel Density Estimation (KDE)) using the ensemble members. The blue line shows the background pdf for the estimated parameter, whereas the orange-red lines correspond to the analysis pdfs after different number of assimilation windows. We see that the DA system works well for one-parameter estimation, especially with multiple observations. In fact, even with a single observation at the end of each simulation window, the analysis pdf moves towards the true ω, and the variance of the posterior pdfs is smaller than that of the prior, and it gets smaller as the number of consecutive assimilation windows increases. The effect of using more observations in each simulation window is to accelerate the convergence of the posterior pdf towards a stationary distribution. Panel (D) illustrates more clearly the evolution of the analysis estimator for sequential assimilation windows (in the horizontal axis). The blue horizontal line shows the value of the background mean $\bar{\boldsymbol{\omega}}^b \approx 1$, whereas the horizontal black line shows the true value of the parameter $\omega^r = 0.5$. In this case, we only track the analysis mean (solid colour lines) and the analysis standard deviation (shaded colour areas). With one observation at the end of the assimilation window, the convergence of the analysis mean is slow, taking about 50 assimilation windows before the analysis mean and standard deviation stabilise. There is, however, a considerable bias with respect to the true parameter. Increasing the number of observations in the assimilation window accelerates the convergence of the mean estimator. The cases with 5 and 20 observations per assimilation window approach their final value between 10 and 20 assimilation windows. The final estimate with 5 observations is still slightly larger than
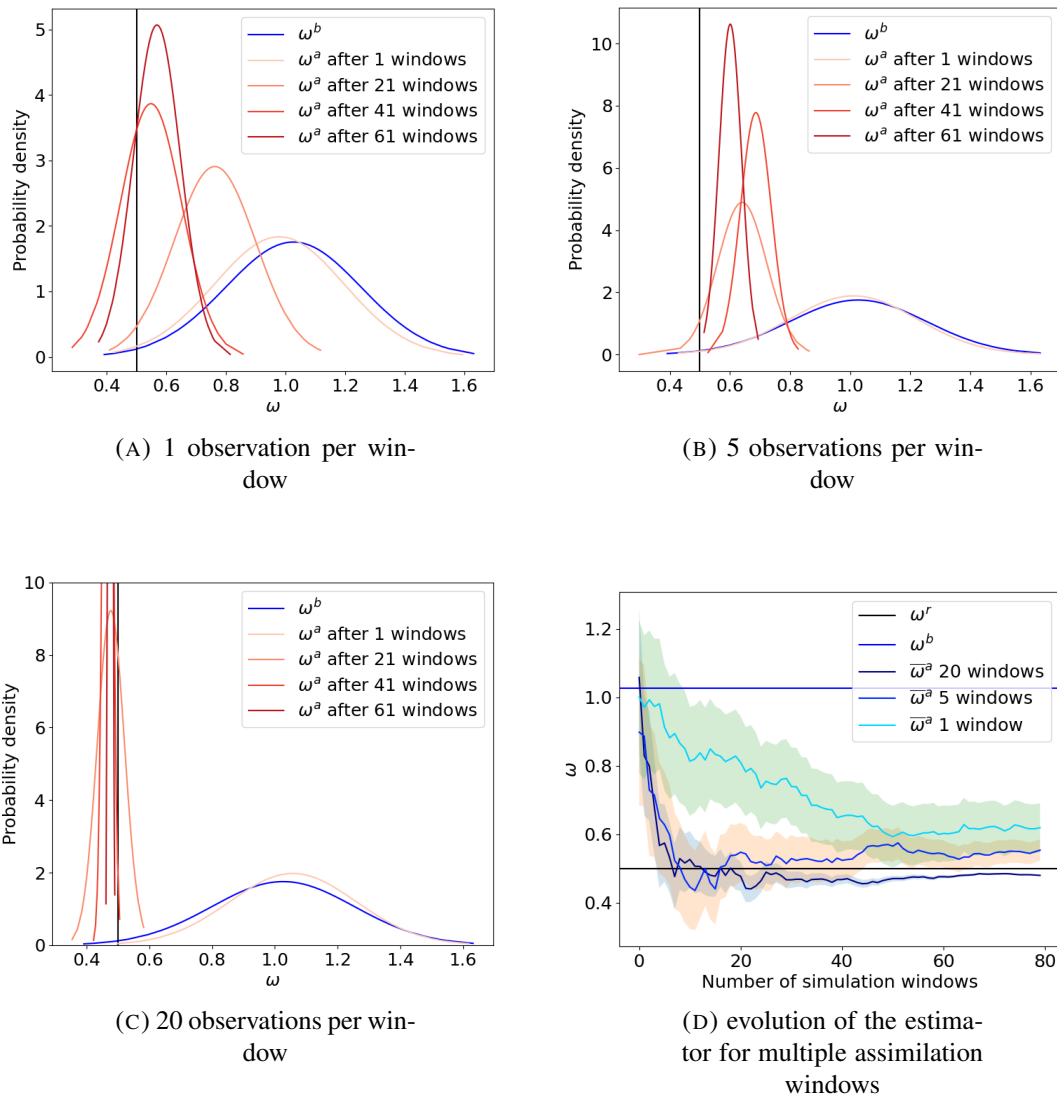
(A) 1 observation per window

(B) 5 observations per window

(C) 20 observations per window

(D) evolution of the estimator for multiple assimilation windows

FIGURE 4.4. (A)~ (C)Exponential scale estimation with different numbers of observations and simulation windows using the EnKS and (D) the convergence of the mean of posterior pdf with the number of simulation windows (the shading is the standard deviation of each case).

the true parameter. The estimate with 20 observations satisfactorily converges toward the real parameter from below, although at the end of the 80 assimilation windows a small gap still remains, much larger then the estimated uncertainty in the ensemble.

We perform similar experiments with the IEnKS. Recall that in this case the number of iterations can influence the performance of the smoother. We analyse the effect of the number of iterations in Figure 4.5. This figure has two similar panels corresponding to

(A) 1 observation per window

(B) 20 observations per window
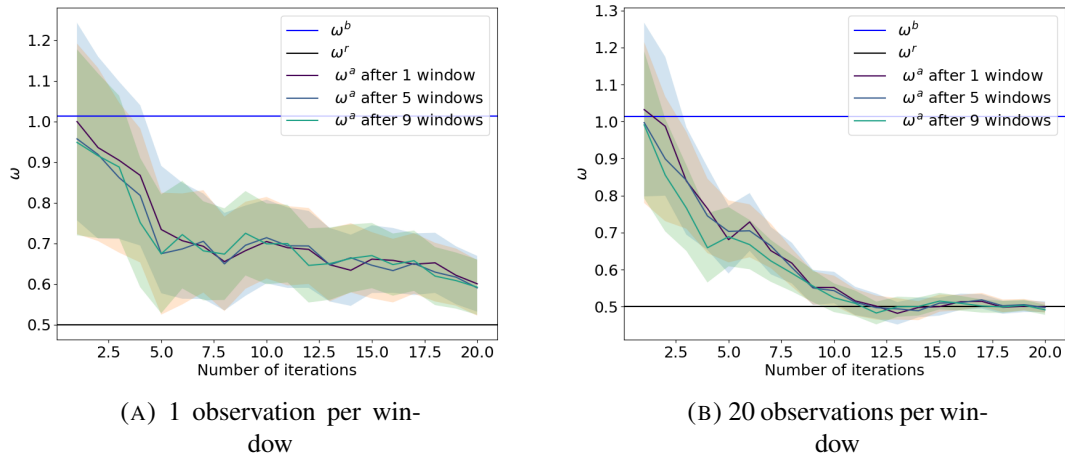
FIGURE 4.5. Analysis mean and standard deviation resulting from using IEnKS with different number of observations (panels), iterations (horizontal axis), after different number of simulation windows (lines).

1 observation (left) and 20 observations (right) per window. The blue horizontal line denotes the background mean, and the black line the real parameter value. In each panel we plot several lines corresponding to analysis mean after different number of assimilation windows, as well as a shaded area corresponding to the analysis standard deviation. In the horizontal axis we have the number of iterations in each window. These iterations use a fixed step length δ. Perhaps the most important message of this figure is that the IEnKS results are independent of the number of assimilation windows, and there is not a noticeable difference between the different lines. For one observation at the end of the assimilation window, the estimator has not fully converged after 20 iterations. It seems that at least 5 iterations per window are necessary for the IEnKS to show reasonable performance. In the case of 20 observations per window, it takes between 15 to 20 iterations for the IEnKS to converge to the true value of the parameter. Again, the number of assimilation windows does not matter.

Based on these results we fix the number of iterations in the IEnKS to $N_{iter} = 10$ for the rest of our experiments. In practice, one aims to perform as few iterations as possible to minimise computational expenses. We use the IEnKS with a different numbers of observations and simulation windows. These results shown in Figure 4.6. The over-
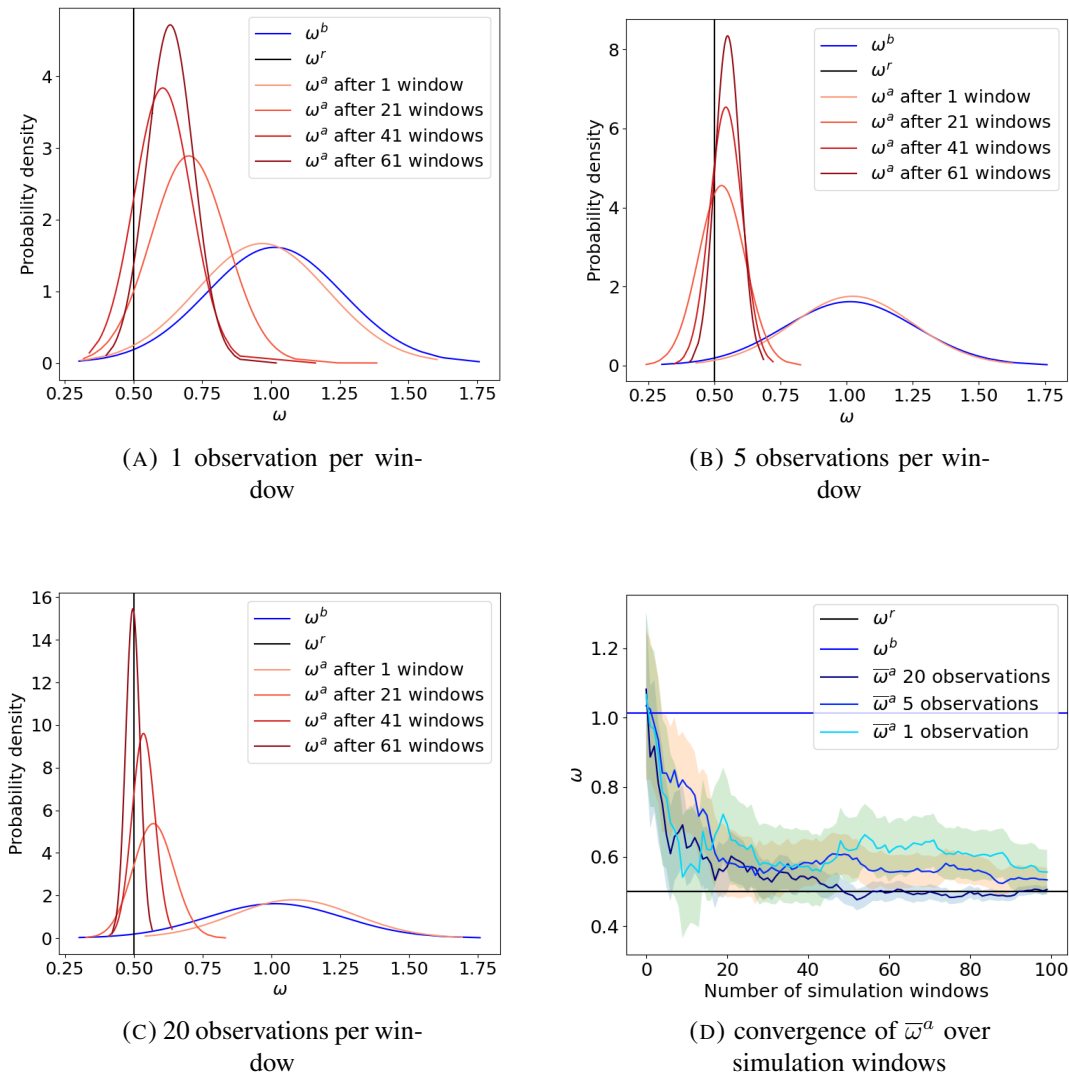

(A) 1 observation per window

(B) 5 observations per window

(C) 20 observations per window

(D) convergence of $\overline{\omega}^a$ over simulation windows

FIGURE 4.6. (A)∼ (C)Exponential scale estimation with different numbers of observations observations and simulation windows using IEnKS and (D) the convergence of the mean of posterior pdf with the number of simulation windows (the shading is the standard deviation of each case).

all format of this figure is the same as that of Figure 4.4. Panels (A) to (C) reveal that the convergence to the analysis pdf is faster with more observations per window, and the resulting consecutive analysis pdf's have smaller variance. Panel (D) reveals that for the three observation frequencies, the major changes to the analysis mean occur in the first 20 assimilation windows. For the cases with 1 and 5 observations per window, the bias in the final analysis mean is reduced considerably with respect to the estimators obtained by the EnKS. With 20 observations per window, the estimation is remarkably accurate, with the analysis mean oscillating around the real parameter value after 50 windows and remaining there.

In this subsection we have shown that EnKS converges to the real parameter value but a small bias remains, while its uncertainty estimate tends to contain that real value. The results show significant improvement increasing in the number of observations and the EnKS benefits from more assimilation windows. Next step is to extend our experiments to a more complex model error setting with two unknown parameters.

4.4.2.2 Two-parameter estimation

For this part of the experiments, we implement the model error with formulated in Eq. (4.40), which contains both decaying and oscillatory parameters. We start by experimenting with the EnKS. This method, however, fails to find the correct values for both parameters even in the case with observations every time step. This case is shown in Figure 4.7. This figure shows the background (blue line) and analysis (orange-red lines) pdf's for the oscillatory (left) and decaying (right) parameters. The vertical black lines correspond the true parameter values. The figure illustrates how the mean of the analysis pdf converges towards wrong values of the parameters, with noticeable variance reduction after successive assimilation windows. This behaviour suggests that the minimisation process in the IEnKS is converging to an incorrect local minimum. To explore the failure of the EnKS in

(A) 20 observations per window
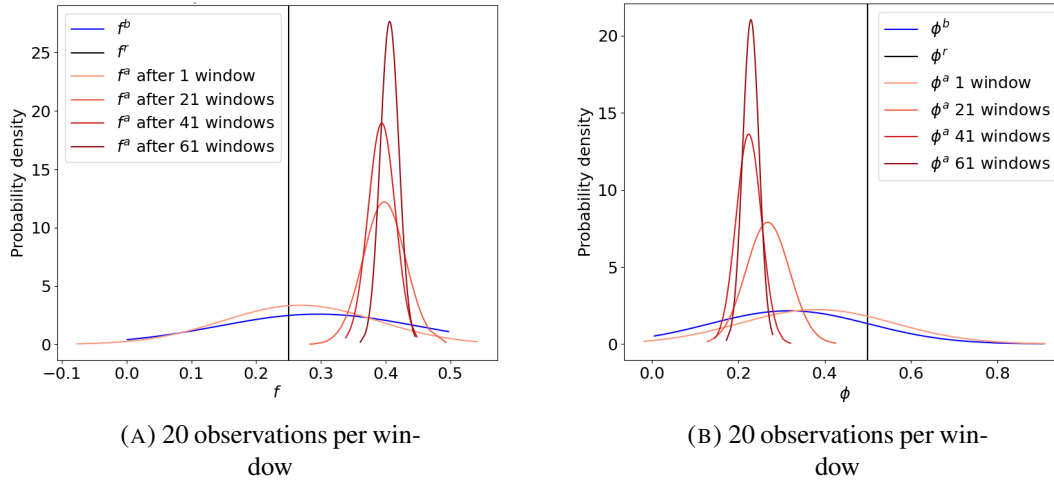
(B) 20 observations per window

FIGURE 4.7. Two-parameter estimation, f (left) and ϕ (right), using the EnKS with 20 observations and different number of simulation windows.

this case, it is useful to display the cost function of the problem. This requires computing Eq. (4.33) using Eq. (4.40) and Eq. (4.42). For simplicity, we do this for the case of a single observation at the end of the assimilation window. Before explaining the result, we need to recall some aspects of this cost function:

- The input of this function is the vector $\mathbf{u}$ and the output is a scalar. The function maps $1 + \tau + N_\theta$ values into a single one.

- The background elements needed by the cost function are the background mean $\boldsymbol{\mu}^{b,u}$ and and covariance $\mathbf{D}^u$.

- The observation elements needed by the cost function are the actual observation y and its variance r^2,

In Figure 4.8 we start with a simple setting.

We use the exact mean $\boldsymbol{\mu}^{b,u}$ and covariance $\mathbf{D}^u$, as well as the observational variance r^2 from section 4.3.3. Since we are not interested in the behaviour of the cost function with respect to the state variables, we set the state variables to a fixed value. and we only
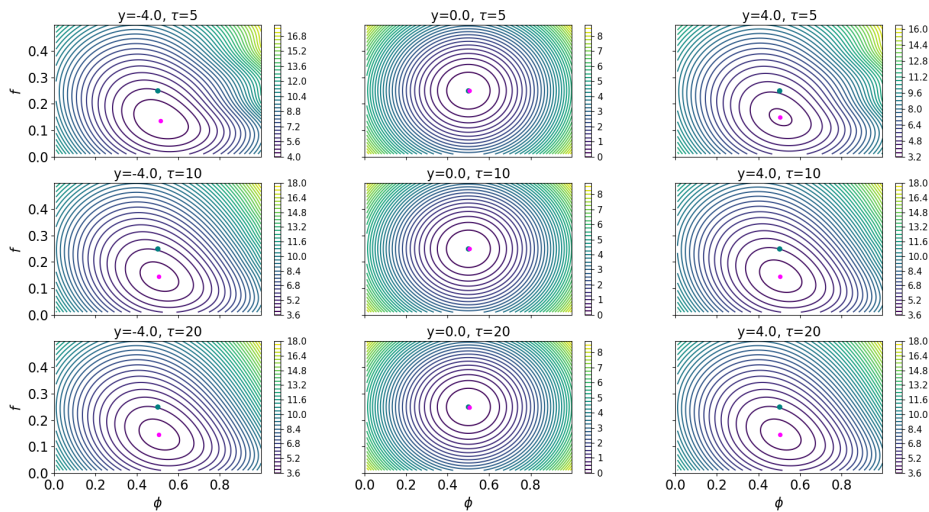
FIGURE 4.8. Exact cost function including the two-parameter model error, the state variable, observations with different number of time-steps and values of observations using the EnKS. The green point represents the analysis value predicted after 1 EnKS step (with no extra iterations), and the pink point represents the exact global minimum.

let the values $\boldsymbol{\theta} = [f, \phi]^\top$ vary freely. We plot a cross-section of the cost-function in the two-dimensional parameter space. We choose $x_0 = 0$ and $\epsilon_1 = \cdots = \epsilon_\tau = 0.1$ –note that we have to choose a value different from zero, otherwise the cost function would be blind to the variation in model error parameters.

Figure 4.8 has nine panels. Each one of the rows correspond to a different length of the assimilation window (recall that the observation is taken at the end), and each one of the columns correspond to a different given observation. We already know that the resulting cost function is not quadratic, and that the deformation from a quadratic cost function grows as the number of time step grows, and as the difference between the given observation and the predicted observation is larger. This is exactly what we observe in the figure. Both the first and third columns, when the observations are $y = \mp 4.0$ respectively, show the largest deformation of the cost function and this grows for the rows from top to bottom. The middle panel ($y = 0$) sees little deformation from a paraboloid. For each

panel, the green dot represents the background mean values of the parameters, and the pink dot corresponds to the global minimum of the cost function. In all cases, the cost function is still concave and there is a unique minimum to be found.

The simple case shown before is not what the EnKS encounters in practice. First of all, we do not know the real background covariance matrix, so this comes from a sample estimator. Second, the cost function is applied to each ensemble member separately, and therefore the background mean is just the background value for each member. The final analysis estimator is the mean of the N_e estimated minima. We explore this in Figure 4.9. First of all, we fix given observation to $y = -4$. Each column represents a different
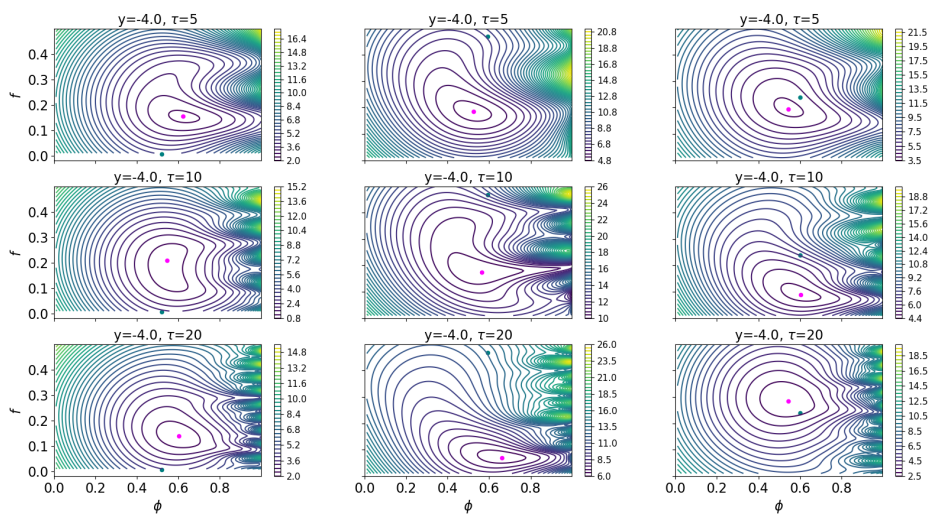
FIGURE 4.9. Sample cost function of different ensemble members (from left to right, $N_e = 2, 4, 8$) including the two-parameter model error, the state variable, observations with different number of time-steps and values of observations using the EnKS. The green point represents the analysis value predicted after 1 ENKS step (with no iterations), and the pink point represents the exact global minimum.

ensemble member (3 members out of the 200 we used to compute background statistics), and each row represents a different length of the assimilation window. To produce these cost functions, the background values are random realisations of the distributions for the

initial conditions, model errors and parameter values shown in section 4.3.3. Again, we are interested in the behaviour in the two-dimensional parameter space, so we fix all the state variables to the same constants as before. What we see is very different from Figure 4.8. Figure 4.9 shows cost functions which are not convex, in fact they have very intricate topographic features such as narrow elongated ridges and valleys, and even some local mimima and maxima. The global cost minimum is difficult to converge to in a single application of the EnKS. In fact, this does not happen. We see that as the assimilation window length increases, the complicated features of the cost function increase. This agrees with the unsuccessful results we had seen in Figure 4.7. Therefore, it is necessary to apply a number of smaller steps via iterations in an IEnKS.

We now apply the IEnKS to avoid getting stuck in a valley or a local minimum in the minimisation process. The results for the two-parameter estimation using IEnKS are shown in Figure 4.10 and 4.11. We apply the IEnKS with 10 iterations and a fixed step length $\delta = 0.3$.

Figure 4.10 shows the ensemble background and analysis pdfs (after different number of assimilation windows) for the oscillatory (top row) and decaying (bottom row) parameters, for 1 observation (top row) and 20 observations (bottom row). This figure reveals that the IEnKS works fairly well with more observations in this case. Similar to the results with the single parameter estimation, the two-parameter estimation results improve with the increase in the number of observations and assimilation windows in the simulation period. The IEnKS estimation converges towards the correct values of both parameters. In the case of only 1 observation in the assimilation window, the pdf stops short of the real value (black line), while in the case of 20 observations the estimation is better. Figure 4.11 tracks the evolution of the analysis mean and standard deviation for the two parameters (one for each panel) as the number of assimilation windows increases, and for three total number of observations: 1, 5 and 20. The true and background mean values are

(A) 1 observation per window

(B) 20 observations per window

(C) 1 observation per window
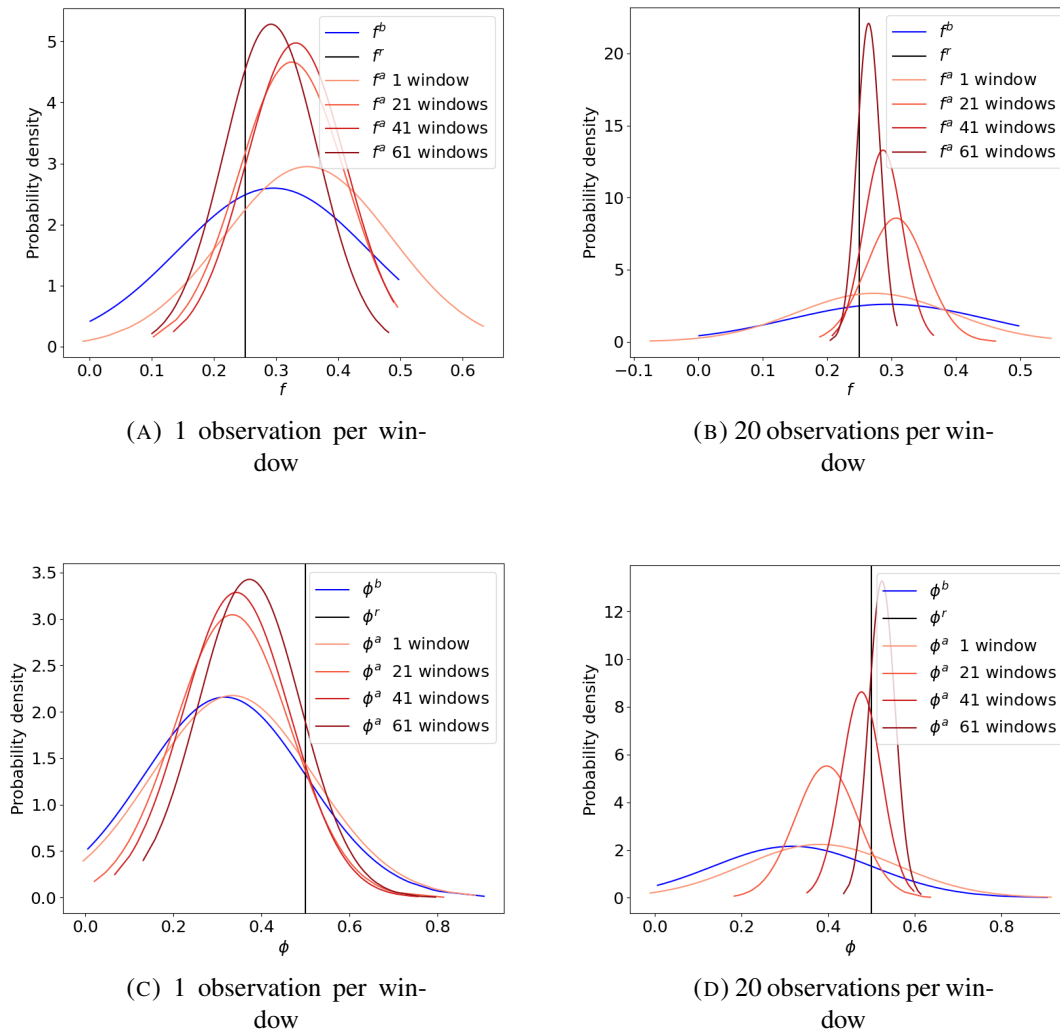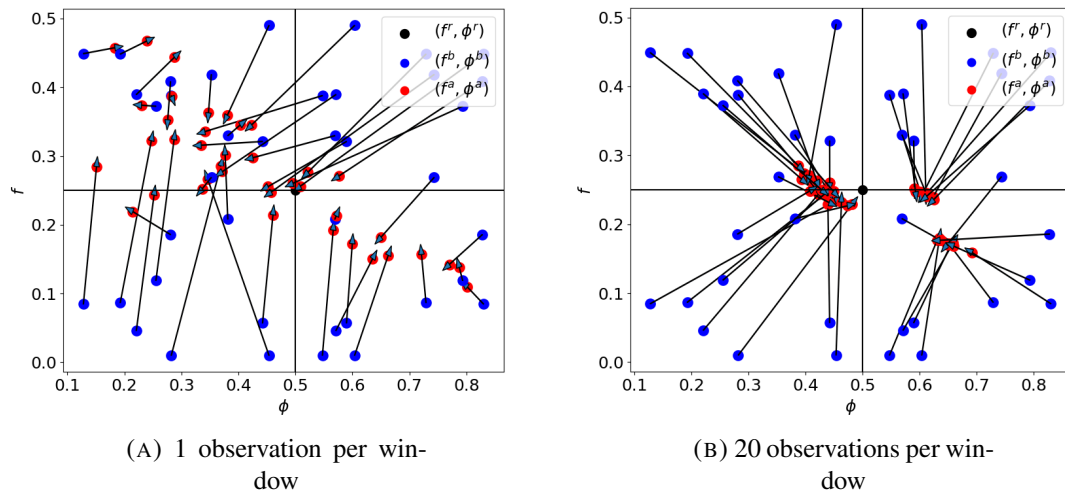
(D) 20 observations per window

FIGURE 4.10. Two-parameter estimation using the IEnKS with different number of observations and simulation windows, and 10 iterations.

shown with horizontal lines, black and blue correspondingly. Note the improvement in the estimation as the number of observations increase. Even for observations every time step, there is a bias in the final estimation of the frequency parameter, while the decay parameter is captured in the ensemble uncertainty. Compared with the results from the one-parameter estimation, the two-parameter estimation problem seems to be much more complicated, and requires more observations to make the estimation work.

(A) $\overline{f}^a$ with the number of windows

(B) $\overline{\phi}^a$ with the number of windows
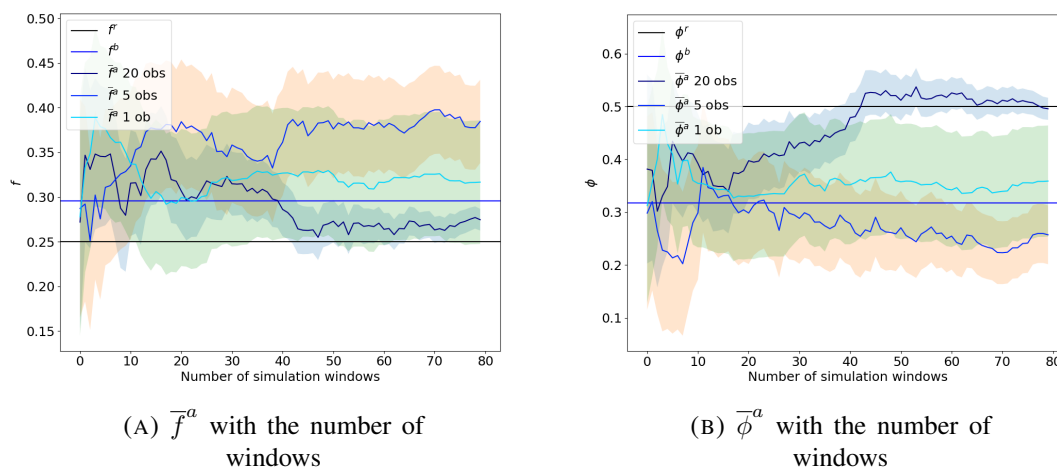
FIGURE 4.11. Posterior mean of the two parameters over the number of simulation windows with different number of observations.

In closing the experiments for this subsection, we want to see how the IEnKS works with different initial guess for the two parameters. Given the complicated shape of the cost function, it is conceivable that we may get stuck in local minimums, even when using this sophisticated iterative method. We divide the two-dimensional parameter space into four quadrants, and we choose starting points in each one of the quadrants. These results are shown in Figure 4.12.

This figure shows results in the case of one observation (left) and 20 observation (right). In both cases, the true values of the parameters are at the centre of the quadrants (denoted with a black dot). We see that the position of the initial condition can have serious impact on the estimation results. When we only observe at the end of the window most initial conditions do not lead to a value close to the true value. In fact, we see that the DA system cannot distinguish between parameter values lying roughly on a straight line defined by the red dots. On the other hand, when we have observations at every time-step, the posterior seems to get fairly close to the true values of the parameters, but many local minima exist around the true values. The results show the importance of the initial guess of the parameters on the parameter estimation results.

(A) 1 observation per window

(B) 20 observations per window

FIGURE 4.12.   Two-parameter estimation for different priors using the IEnKS with different number of observations and simulation windows after 10 iterations. The blue dots show different background values, used as initial conditions for the minimisation. The red dots show the obtained analysis values. The black dot in the centre shows the true values for the parameters.

After experimenting with the linear model, our next step is extending the experiments of two-parameter model error autocorrelation to the nonlinear model, the logistic map.

4.4.3   Parameter estimation in a non-linear model

The last experiments we perform are also the most challenging. In this case our true model is the logistic map with the damping coefficient $\gamma = 3.99$, and the proxy model is linear model with auto-correlated model error. The true model error autocorrelation in this case has both an oscillatory and decaying behaviour (this can be diagnosed offline as described in section 4.4), so we directly try two-parameter estimation. Since the EnKS failed on the two-parameter estimation even with the linear model, we decide to only apply the IEnKS with 10 iterations, over times windows of 10 time steps long.

In our first experiments we try to estimate f and ϕ using the IEnKS with a fixed iteration step length $\delta = 0.3$ as used in the linear experiments. The results of these experiments are shown in Figure 4.13 for both parameters (rows) and for 1 and 20 observations in the window (columns). With 1 observation at the end of the window the smoother does not

(A) 1 observation per window

(B) 10 observations per window

(C) 1 observation per window

(D) 10 observations per window
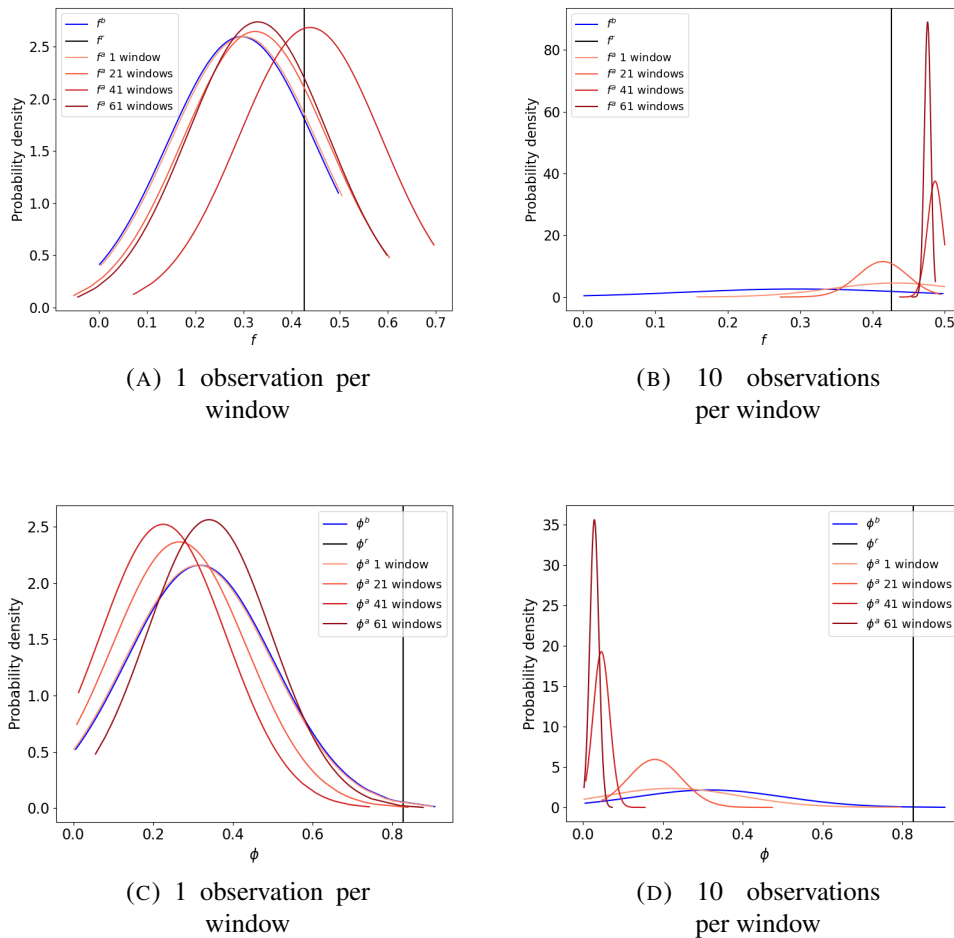
FIGURE 4.13. Two-parameter estimation using the IEnKS with the logistic map, using a fixed iteration step length ($\delta = 0.3$), 10 iterations per window, different number of observations and simulation windows.

seem to converge, and with 20 observations convergence seems to occur, but to incorrect values of the parameters. Increasing the number of iterations per window or the number of windows did not improve results.

Since the IEnKS with a fixed step length δ fails to estimate the parameters directly, we explored three further experimental settings. In the first we transform the estimation problem to a more linear setting by estimating lag-1 and lag-2 autocorrelations of the model error (labelled as $AC(1)$ and $AC(2)$ respectively), and then transform the results to f and ϕ as the following:

$$
\phi = \sqrt{2AC(1)^2 - AC(2)}
$$
$$
f = \frac{\arccos\left(\frac{AC(1)}{\sqrt{2AC(1)^2 - AC(2)}}\right)}{2\pi}. \tag{4.45}
$$

In the second setting we use a decaying step length δ while using the untransformed parameters. However, Figure 4.14 shows that neither method is successful, suggesting that the problem is highly nonlinear and a more careful tuning of the minimization is required.

In the third setting we both transform problem to first estimating the autocorrelations and applying a decaying step length. However, with only 1 observation in the assimilation window the IEnKS failed, no matter what we tried. For a fully observed system with 10 observations per assimilation window successful results can be achieved. We show results for reducing the step length δ by 2% after each iteration (i.e. $c = 0.98$ in Eq. (4.37)). From the two top plots in Figure 4.15, we can see that the results are similar to the successful results for the linear model shown in Figure 4.10. The posterior shows improvement with more simulation windows, and the variance decreases with more windows as well. Figures 4.15 (C) and (D) show that the posterior mean moves towards the right values for both parameters, with slowly decaying uncertainty and not monotonically.

Even though the minimization is successful the convergence is extremely slow, even in this simple zero-dimensional model. This illustrates that estimating parameters in model errors that are related to temporal correlations is a very hard problem.
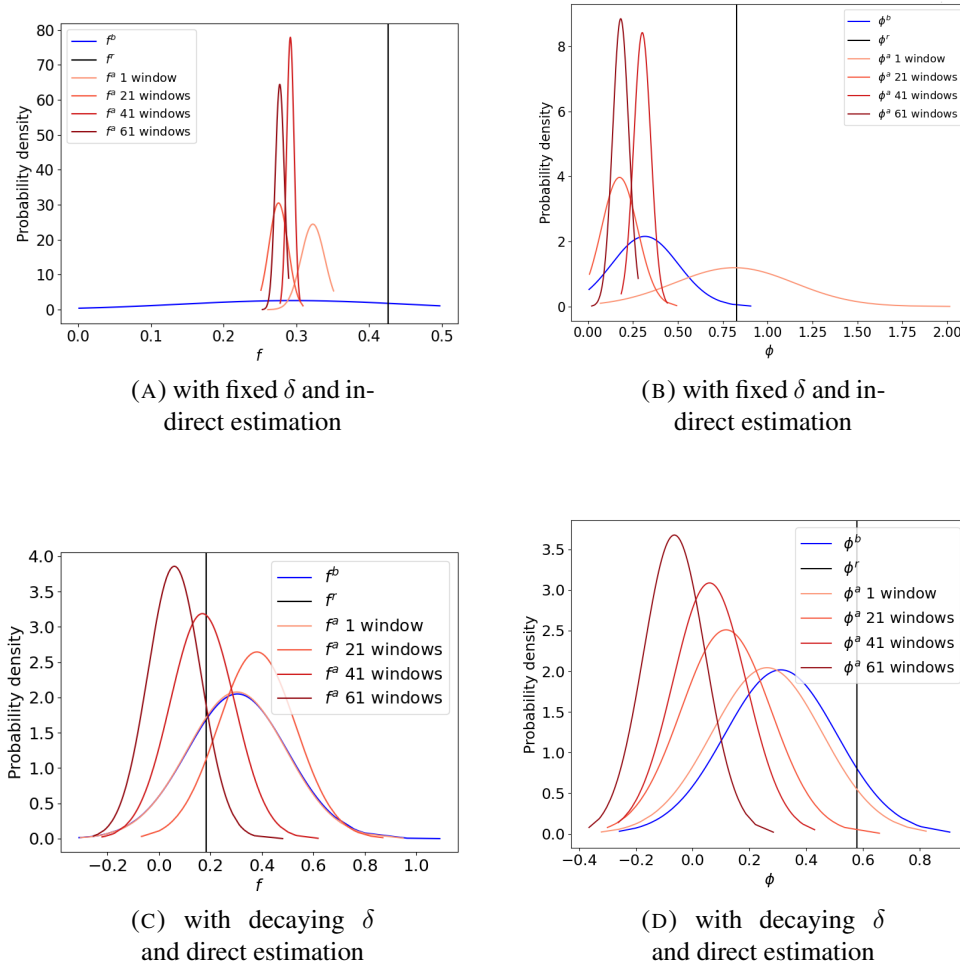
FIGURE 4.14. Two-parameter estimation using the IEnKS with 10 observations per window after 10 iterations with the logistic map. On the top panel, (A)∼(B) the iteration step-length is fixed (δ=0.3), and we estimate the lag-1 and lag-2 autocorrelation then transform them to f and ϕ. On the bottom panel, (C)∼(D) the parameters are estimated directly with a decaying δ.

4.5 Summary and conclusions

Including the model error in the DA process is a difficult task to tackle, especially when this model error possesses spatial and temporal correlations. The objective of this work is to test whether it is possible to include the estimation of time-related model error parameters in the DA process. We have shown that even in the case of linear dynamics,

(A) 10 observations per window

(B) 10 observations per window

(C) convergence of $\overline{f}^a$

(D) convergence of $\overline{\phi}^a$
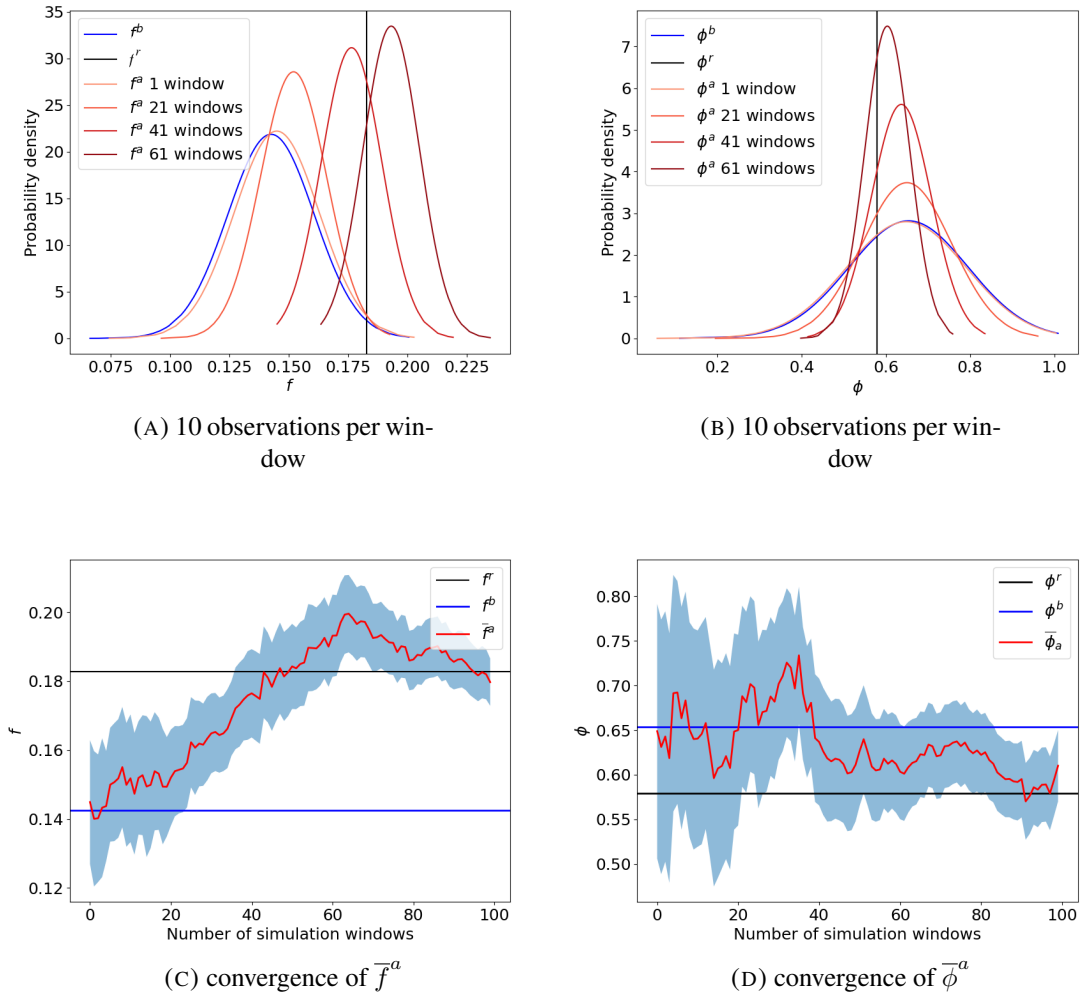
FIGURE 4.15. Lag-1 and lag-2 autocorrelation are estimated and transform to the parameters (A) f and (B) ϕ, using the IEnKS with a decaying iteration step length, 10 iterations per window, for different numbers of assimilation windows with the logistic map. The bottom two figures show the convergence of the posterior mean of (C) f and (D) ϕ over the number of simulation windows with observations every time-step.

the model error parameters are involved in an intricate and non-linear way in the model evolution. We have therefore focused in the use of the EnKS and the IEnKS combined with parameter augmentation to solve the joint state-parameter estimation problem.

With a simple linear-regressive model and exponential decaying memory in the model

error, the EnKS works well using the state augmentation method to estimate a single model error parameter, even with only one observation at the end of the assimilation window. The estimation results are improving with an increasing number of observations in each window, and by iterating over more assimilation windows. The IEnKS with 10 iterations gives slightly superior results compared to the EnKS in this case, as expected.

When the complexity of the temporal correlation in the model error is increased by including an oscillatory component state augmentation with the EnKS fails. Indeed, the cost functions involved show highly irregular shapes and convergence in one step is not possible. An IEnKS with a fixed iteration step length is shown to converge to the correct values when the number of observations in an assimilation window is high enough. The failure to converge with one observation at the end of the assimilation window can be traced back to the long valley with minimal gradients in the cost function, identified via different first guess values. There just isn't enough information to estimate both model error parameters. However, we do see that the uncertainty in the estimates remains large, so the IEnKS does show consistent results. Increasing the number of observations in the window solves this problem.

The nonlinear logistic map proves to be more challenging when estimating both model error parameters. The EnKS always fails, and the IEnKF fails with the standard fixed step length and direct estimation of the parameters. The combination of a step length that decreases with iteration and transforming the problem by first estimating the autocorrelations in the model error improves the results significantly, although we need a fully observed system, and convergence is very slow.

Overall, we conclude that estimating temporal-correlation related parameters in model errors is a highly nonlinear problem, and more difficult than expected even in zero-dimensional systems. The number of observations in an assimilation window needs to

be high enough for the smoother to have enough information on these parameters, and for nonlinear models careful tuning of an IEnKS is needed. The bottom line is that it is very well possible that the data-assimilation system does not have enough information to pinpoint the exact model error parameters, while at the same time we should realize that these parameterizations of model errors are approximate by their very nature.

It is important to keep in mind that these results were obtained with a very large ensemble size of 200 for a zero-dimensional system and two parameters. This choice made sense for the focus of this chapter, but in reality the limited ensemble size will give rise to extra noise in the estimates. The IEnKS, while being a very powerful method, does not converge to the posterior pdf with growing ensemble size for nonlinear data-assimilation problems (see e.g. Evensen, 2018). Especially when multiple modes are present in the posterior, as in our case, several of the modes can be missed. This is problematic when local modes are almost as high as the global mode, in which case the correct solution to the data-assimilation problem is this family of modes. To find these fully nonlinear methods like local particle filters or particle flows are needed (see e.g. Van Leeuwen et al., 2019; Hu and Van Leeuwen, 2021; Evensen, Vossepoel, and Van Leeuwen, 2022).

Our next step is to investigate how these results carry over to more realistic and complex models such as a two-layer quasi-geostrophic model. The challenge in this case is whether the spatiotemporal model error correlation is separable or not. The ultimate goal is to have a online update for the model error within an operational system.

Chapter 5

Parameter estimation of model error autocorrelation with the Quasi-geostrophic model

The final part of this project is to apply the online parameter estimation using data assimilation method on a more complex and realistic model, the Quasi-geostrophic (QG) model. In this chapter, we only experiment with the simplest model error autocorrelation. The program for the QG is provided by Prof. Peter Jan Van Leeuwen, and the experiments and chapter are finished by Haonan Ren with the supervision of Prof. Peter Jan Van Leeuwen and Dr. Javier Amezcua.

5.1 Introduction

In this chapter, we consider the model error in a more realistic and complex situation, in which the spatial and temporal correlations are both considered in the data assimilation process. We aim to estimate certain parameters in the model error autocorrelation using ensemble data assimilation schemes, so that we can have a better understanding of the model error statistics and improve the data assimilation results. We perform parameter estimation using the state augmentation method with the ensemble Kalman Smoother (EnKS) on the 2-layer QG model. The general structure of this chapter is as follows. In

section 2, we demonstrate the formulation of the 2-layer QG model, including the governing equations, the general solution procedure and the experiment settings. Then, we present the formulation of the model error we used in the QG model and the data assimilation scheme we employ in section 3. In section 4, we will show and discuss the results from the numerical experiments for both state update and parameter estimation. The conclusion of the experiment results on QG model will be presented in section 5.

5.2   The two layer Quasi-geostrophic model

To extend our experiments to more practical models instead of simple, low-dimensional linear and nonlinear models used in the previous chapters, we apply a two-layer quasi-geostrophic (QG) model. The QG model simulates mid-latitude synoptic and larger-scale flow. This flow is almost geostrophic because key features of synoptic weather systems require a non-geostrophic flow component (Shepherd, 2015). Using this model, we can more comprehensively understand the impact of the autocorrelated model error on the numerical forecasting system and data assimilation results, and how to deal with it in the operational systems in the future with same method applying to the operational system.

5.2.1   Governing equation

The 2-layer QG model solves the following equations:

$$\frac{\partial p^1}{\partial t} + \mathcal{J}(\psi^1, p^1) = A\Delta q^1, \tag{5.1}$$

$$\frac{\partial p^2}{\partial t} + \mathcal{J}(\psi^2, p^2) = A\Delta q^2, \tag{5.2}$$

in which ψ^i is the streamfunction related to the velocity in layer i, p^i represents the potential vorticity in layer i. The Jacobian $\mathcal{J}(\psi^i, p^i)$ is defined as:

$$\mathcal{J}(\psi^i, p^i) = \psi^{ix} p^{iy} - \psi^{iy} p^{ix}. \tag{5.3}$$

The potential vorticity p^i is defined as the sum of the relative vorticity q^i, the planetary vorticity f and a stretching term:

$$p^1 = q^1 + f - F^1(\psi^1 - \psi^2), \tag{5.4}$$

$$p^2 = q^2 + f + F^2(\psi^1 - \psi^2), \tag{5.5}$$

where q^i is the relative vorticity that represents the vertical component of the curl of the velocity

$$q^i = v^{ix} - u^{iy} = \Delta\psi^i, \tag{5.6}$$

and F^i is a constant related to so-called reduced gravity acceleration $g' = g(\rho^2 - \rho^1)/\rho^2$, and the layer thickness of layer i, H^i:

$$F^i = \frac{f^{0^2}}{g' H^i}, \tag{5.7}$$

Finally, planetary vorticity f is given by the β-plane approximation in which y is the meridional coordinate:

$$f = f^0 + \beta y. \tag{5.8}$$

5.2.2 Solving the model

The general procedure to solve the 2-layer QG model with the governing equations shown in Eq. (5.1) and (5.2) is as follow:

1. take the Laplacian to calculate relative vorticities from the streamfunction,

2. calculate the potential vorticity p by adding the planetary vorticity f and the stretching terms to the results in step 1,

3. solve the evolution equation for the potential vorticity over one time step ($3600s = 1h$),

4. determine the new streamfunction fields ψ^i, by solving the two Helmholtz equations in Eq. (5.4) and (5.5) with Eq. (5.6), go back to step 1.

The Helmholtz problems in Eq. (5.4) and (5.5) are coupled in this case due to the stretching term. Therefore, we need to reformulate the layer equations to model equations via the introduction of the barotropic and the baroclinic streamfunction fields, ψ^{bt} and ψ^{bc}:

$$\psi^{bt} = \frac{H^1\psi^1 + H^2\psi^2}{H^1 + H^2}, \tag{5.9}$$

$$\psi^{bc} = \psi^1 - \psi^2. \tag{5.10}$$

We can do the same for the potential vorticities since the they are linearly related to the streamfunction fields:

$$p^{bt} = q^{bt} + f = \Delta\psi^{bt} + f, \tag{5.11}$$

$$p^{bc} = q^{bc} - (F^2 + F^2)(\psi^1 - \psi^2) = \Delta\psi^{bt} - (F^1 + F^2)\psi^{bc}, \tag{5.12}$$

where q^{bt} and q^{bc} are the relative vorticity of the barotropic and the baroclinic streamfunction fields. Setting $F^1 + F^2$ as λ, then we have:

$$\Delta\psi^{bt} = p^{bt} - f, \tag{5.13}$$

$$\Delta\psi^{bc} - \lambda\psi^{bc} = p^{bc}. \tag{5.14}$$

Eq. (5.13) is a Poisson equation and Eq. (5.14) is a Helmholtz equation. Solving these elliptic equations requires boundary conditions.

5.2.3 Boundary conditions

The domain we use is rectangular with periodic boundaries in the zonal direction, which means for any variable x we have $x^i = x^{L+i}$ in which L is the dimension of the system in zonal direction. For the southern and northern boundaries, we decide that the normal component of the fluid velocity field is equal to zero and the tangential component is unrestricted along the boundaries (so-called free slip conditions). To avoid energy flow through the boundary we set tangential velocity u to zero at southern and northern boundaries. According the Eq. (5.6), we now have the relative vorticity $q = v^x - u^y = 0$, which sets the boundary conditions for the relative vorticity when calculating $q = \Delta\psi$. Since v is set to zero at the southern and northern boundaries, and $v = \psi^x$, the streamfunction ψ along these boundaries is constant. Furthermore, we can add an arbitrary constant to each streamfunction value as the dynamics only use its spatial and temporal derivatives, which allows us to choose the value of each along either the northern or the southern boundary. We choose $\psi^{bt} = \psi^{bc}$ for the southern boundary.

For the streamfunction along the northern boundary, it can be found by using a velocity constraint and integrating the zonal velocity along the northern boundary using the zonal momentum equation:

$$\frac{\partial u}{\partial t} + uu^x + vu^y - fv = A\Delta u, \tag{5.15}$$

where u is the tangential velocity. Because normal velocity v is zero (no mass flux through these boundaries), and the system is periodic in zonal direction:

$$\frac{d}{dt}\int_0^L u\,dx = A\int_0^L u^{yy}\,dx. \tag{5.16}$$

Since dissipation is small in the free atmosphere this is close to zonal momentum conservation along the northern boundary. In fact, in the program we set:

$$\frac{d}{dt} \int_0^L u\, dx = 0, \tag{5.17}$$

which is equivalent to a small momentum source that counteracts the dissipation to keep the flow going over longer periods of time. This can be considered the equivalent of the continuous heating at the equator and cooling at the North Pole by the differential solar insolation, which keeps a meridional temperature gradient that partly enforces the Jet Stream.

The boundaries can be incorporated via a two-step procedure. The Poisson or Helmholtz equation in Eq. (5.13) and (5.14) needs us to specify the actual ψ value at the northern boundary which is unknown beforehand. We solve this problem as:

$$\psi = \hat{\psi} + c\tilde{\psi}. \tag{5.18}$$

The first streamfunction $\hat{\psi}$ can be solved using the values: $\hat{\psi}^S = \hat{\psi}^N = 0$. For the east and west, we always have the periodic boundaries. As for the second streamfunction, we have the values: $\tilde{\psi}^N = 1$, $\tilde{\psi}^S = 0$. The constant c is to ensure we fulfill the northern boundary condition that the integrated zonal momentum does not change over time. For the barotropic case, we specifically have:

$$\Delta\hat{\psi} = p^{bt} - f, \tag{5.19}$$

$$\Delta\tilde{\psi} = 0. \tag{5.20}$$

Eq. (5.19) can be solved analytically, which indicates that $\tilde{\psi}$ doesn't depend on x, as $\tilde{\psi}^y = (y - y^S)(y^N - y^S)$. Eq. (5.20) is solved using a Poisson solver, in which the

constant c can be found as:

$$\frac{d}{dt}\int_0^L \mathbf{u}^N dx = -\frac{d}{dt}\int_0^L \psi^y dx \approx -\frac{d}{dt}\int_0^L \frac{\psi(y^N) - \psi(y^N - \Delta y)}{\Delta y} dx = 0. \quad (5.21)$$

Then, we plug in Eq. (5.18) and discretize the time derivative we find:

$$c_{t+1} = c_t - (\hat{s}_{t+1} + c_t\tilde{s}_{t+1})/\tilde{s}_t + (\hat{s}_t + c_t\tilde{s}_t)/\tilde{s}_t, \quad (5.22)$$

where t is the time index and:

$$\hat{s} = \frac{d}{dt}\int_0^L \frac{\hat{\psi}(y^N) - \hat{\psi}(y^N - \Delta y)}{\Delta y} dx, \quad (5.23)$$

$$\tilde{s} = \frac{d}{dt}\int_0^L \frac{\tilde{\psi}(y^N) - \tilde{\psi}(y^N - \Delta y)}{\Delta y} dx. \quad (5.24)$$

The baroclinic problem can be solved in a similar way, but the conditions used are

:

- the circulation condition, as for the barotropic case,

- the mean transport in the baroclinic field is zero.

Next, we will present the settings we use for the QG model in our experiments.

5.2.4 QG model settings

The general settings for the QG model are as follow:

Domain grid points zonal dimension of the system $n^{xx} = 257$, and meridional dimension of the system $n^{yy} = 129$, vertical levels $n^l = 2$, distance between grid points $dx = 5000m$,

Model thickness upper layer thickness $H^1 = 500m$ and lower layer thickness $H^2 = 4500m$,
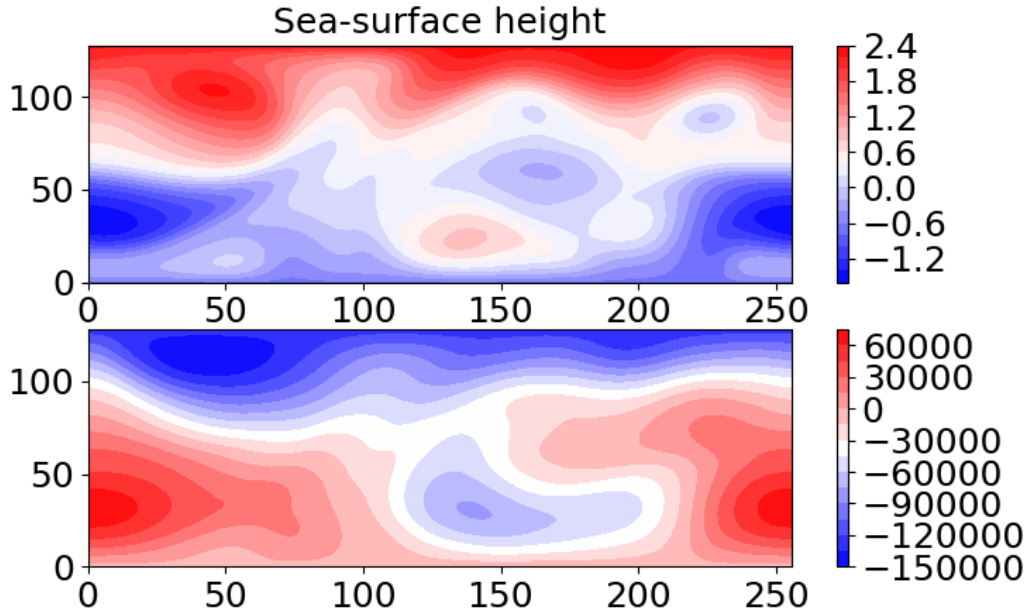
FIGURE 5.1. Plots of the QG model, the upper layer streamfunction is transformed to sea-surface height with the unit of m and lower layer has the unit of m^2/s.

Simulation time domain number of time-steps in each assimilation window $t = 10$, and model time-step $dt = 1h$,

Parameters Coriolis coefficient $f = -1.26 \times 10^{-4}$, and gravity acceleration $g = 9.81m/s^2$.

As an example of a model run for the 2-layer QG model is shown in Figure 5.1. For the streamfunction plots, we transform the upper layer streamfunction ψ_1 to sea-level height h with the unit of meter via:

$$h = \psi^1 f^0 / g \tag{5.25}$$

5.3 Model error and experimental settings

5.3.1 Model error

For this part of the project, we apply an additive model error at each time-step. In the program, we use the model error from the current time-step to create the model error for next time-step so that we can implement the model error autocorrelation:

$$\boldsymbol{\nu}_{t+1} = \sqrt{1 - \phi^{1^2}} \boldsymbol{\nu}_{t+1}^{random} + \phi^1 \boldsymbol{\nu}_t, \tag{5.26}$$

where $\boldsymbol{\nu}$ is the autocorrelated model error, ϕ^1 is the lag-1 autocorrelation and $\boldsymbol{\nu}^{random}$ is uncorrelated in time drawn from a normal distribution with zero mean and $\mathbf{Q}$ as the covariance matrix: $\mathcal{N}(\mathbf{0}, \mathbf{Q})$. The ϕ^1 is implemented via:

$$\phi^1 = e^{-\frac{1}{\omega}}, \tag{5.27}$$

where ω is the decaying correlation scale which is the parameter we need to estimate. Besides the model error autocorrelation in time, a Gaussian spatial correlation is assumed in which the correlation length is 50 grid points. Meanwhile, the model errors in different layers are correlated similar to how the model error is correlated in time:

$$\boldsymbol{\nu}^{l_2} = \sqrt{1 - \alpha^2} \boldsymbol{\nu}^{l_2} + \alpha \boldsymbol{\nu}^{l_1}, \tag{5.28}$$

in which α is the vertical correlation coefficient, l_1 and l_2 represent the levels in the QG model.

5.3.2 Experiment settings

To obtain better estimations of the state and parameter, we use as many ensemble members as we can considering the computational power in the forecast and data assimilation

scheme $N_e = 500$. With this large number, we don't need to consider localization. Based on the size of the model and the limited computational power, to have as many observations as we can, we decide to formulate the EnKS in the ensemble subspace (see chapter 8 in Evensen, Vossepoel, and Van Leeuwen, 2022), which avoid to the need of computing a huge matrix with the size of $\mathcal{R}^{N_{ob} \times N_{ob}}$ and N_{ob} is the number of observations we have in the field.

5.4   Results of the QG model experiments

The results presented here are preliminary, and it is realized that much more needs to be done to obtain scientifically sound conclusions. However, they do demonstrate that it is possible to estimate model error parameters in high-dimensional systems.

5.4.1   State update of the 2-layer QG model

We first run the model for 2000 time-steps (spin-up) for the QG model to reach a state of statistical equilibrium under the applied initial condition. Using the state at the end of the spin-up as the initial condition, the truth is generated with the real autocorrelation timescale. Each ensemble member is generated using the same model but perturbed initial condition and different timescales in model error. Before we start the experiments for parameter estimation, we first investigate the performance of the state estimation alone. For this part of the experiments, we observe every $3rd$ grid points in both x and y directions of the field away from the northern and southern boundaries (3 grid-points away), and only the state at the end of the simulation window is observed. The results for the state update are shown in Figure 5.2. From the state plots shown in Figure 5.2, for the upper layer, we can clear see update near the centre of the field and around the southern boundary. As for the lower layer, the difference between the truth and the prior is much
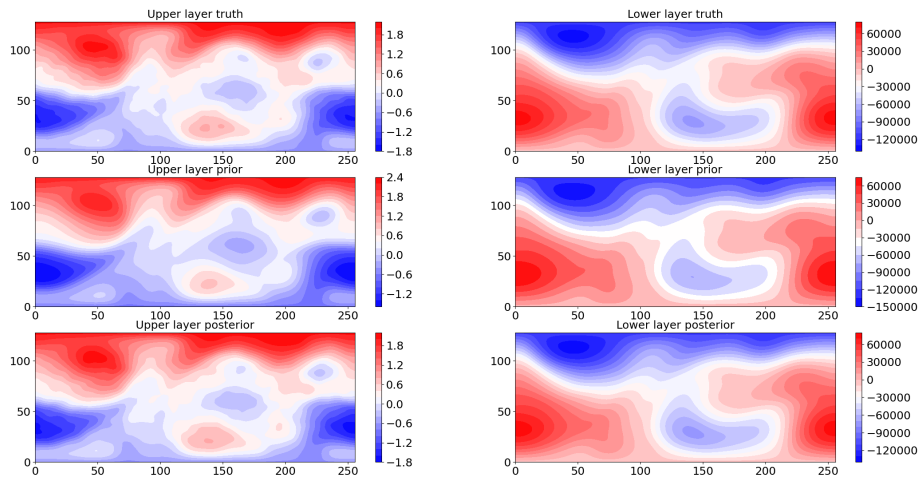
FIGURE 5.2. Plots for upper (left) and lower (right) layers (with the unit of m and m^2/s) truth (top) against the prior mean (middle) and posterior mean (bottom) at the observation time-step (the end of simulation window).

smaller. However, we can still see updates in the centre of the field.

Now, we plot the RMSE against the ensemble spread in Figure 5.3 to assess the overall quality/performance of the data assimilation results. Figure 5.3 shows that for both upper and lower layer, the posterior RMSE and ensemble spread are significantly smaller than prior RMSE and ensemble spread. At the observation time-step (the end of the simulation window), we can see that the posterior RMSE matches the value of the ensemble spread for both layers. However, it is noteworthy that the lower-layer spread is too low compared to the RMSE. Since both the truth and the ensemble are drawn from a long model run this is a random effect that should disappear after a few assimilation windows.

5.4.2 Parameter estimation with QG model

Parameter estimation is a complex problem that might require more information from the observations to successfully estimate the parameters. Therefore, we experiment with
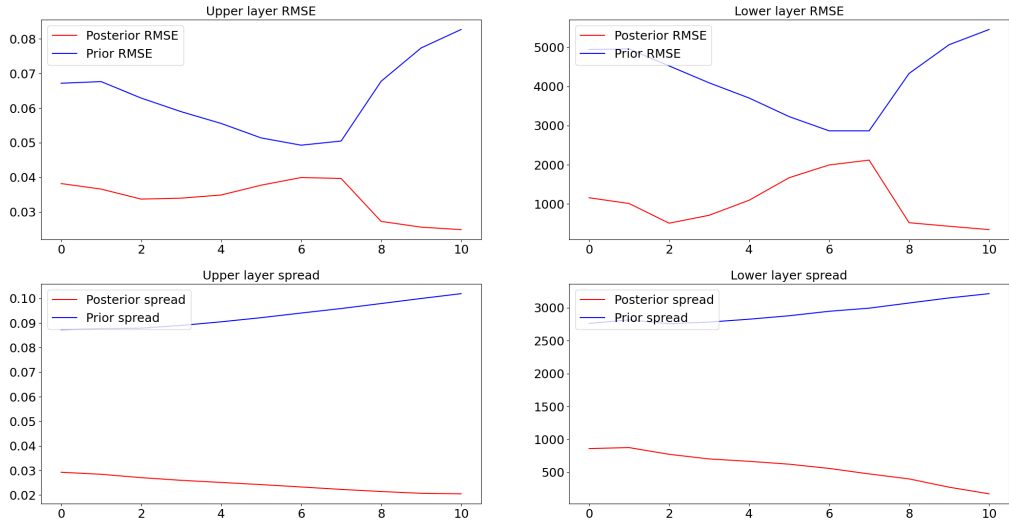
FIGURE 5.3. Comparison of the prior/posterior (blue/red) RMSE (top) with the ensemble spread (bottom) for both upper layer (left) with the unit of m and lower layer (right) with the unit of m^2/s over a single simulation window observing at the end of simulation window.

different numbers of observations in space. To try to keep some realism, we only use observations at the end of the 10-hour window. To achieve successful convergence of the parameter to the correct value, we use a cycling scheme, in which the posterior state as the end of the previous window is used as the initial condition for the next window simulation, and the parameter is updated in the previous window and used to generate the prior model error for the next window. As mentioned before, a single simulation window contains 10 time-steps, and the whole simulation period has 5 simulation windows (50 hours in total). As indicated in the previous chapters, instead of directly estimating the parameter, we estimate the lag-1 autocorrelation ϕ_1 in Eq. (5.27), then transform it back to the decaying timescale ω via:

$$\omega = -\frac{1}{\ln \phi^1}. \tag{5.29}$$

For this part of the experiments, we set the real value of the decaying coefficient ω to 1.0, and the guessed coefficient used for the ensemble forecast model and data assimilation

scheme $\omega^g \in \mathcal{N}(3.0, 1.0)$. We experiment with different observation densities in space: observing every $3rd$ grid points, every $2nd$ grid points and every grid point in both (x and y) directions. The Iterative EnKS is used, but, for computational efficiency, we do not update the linearization point of the observation operator as that would require to rerun the ensemble at every iteration.

We start with observing every $3rd$ grid points of the streamfunction field in both layers, at the end of the 10 hour assimilation window. The pdfs of the model error parameter are shown in Figure 5.4. The pdf moves towards the true value but its variance decreases
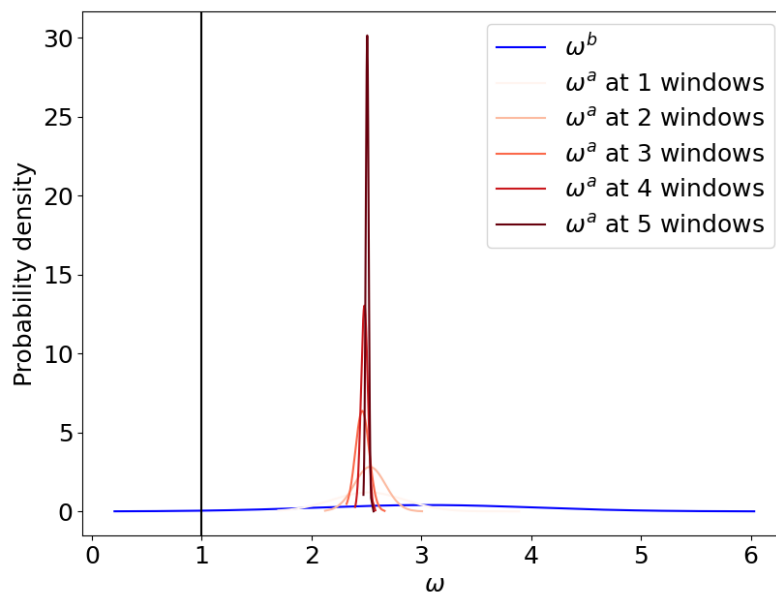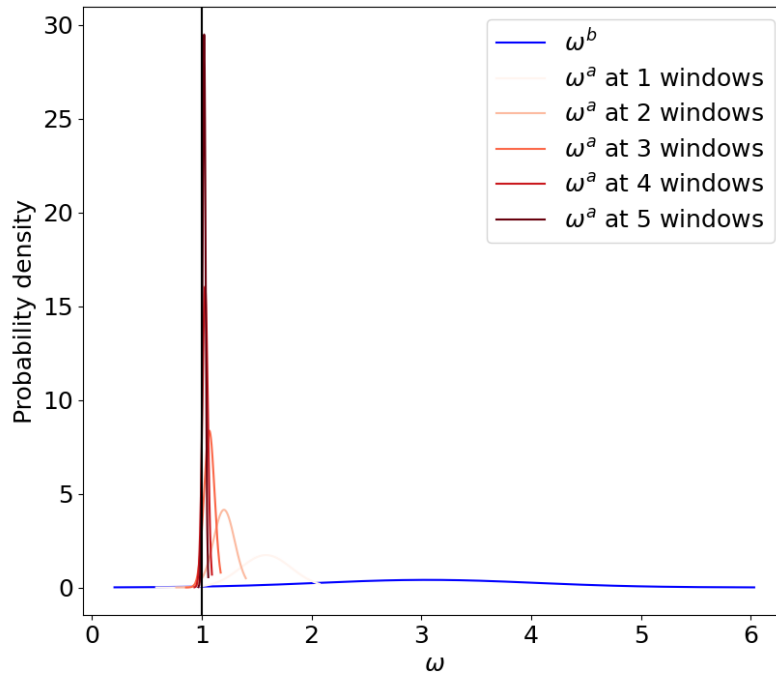

FIGURE 5.4. PDF of the estimated parameter within 5 simulation windows and with observation every $3rd$ grid points, the black vertical line is the value of the real ω of the model error.

too rapidly, and the estimates get stuck in a local minimum around $\omega = 2.5$. To solve this problem we inflated the posterior pdf after each iteration, but that did not help. We then increased the observation density at the end of the 10-hour window to every $2nd$ grid point. As can be seen in Figure 5.5, the estimated pdfs move closer to the truth, but
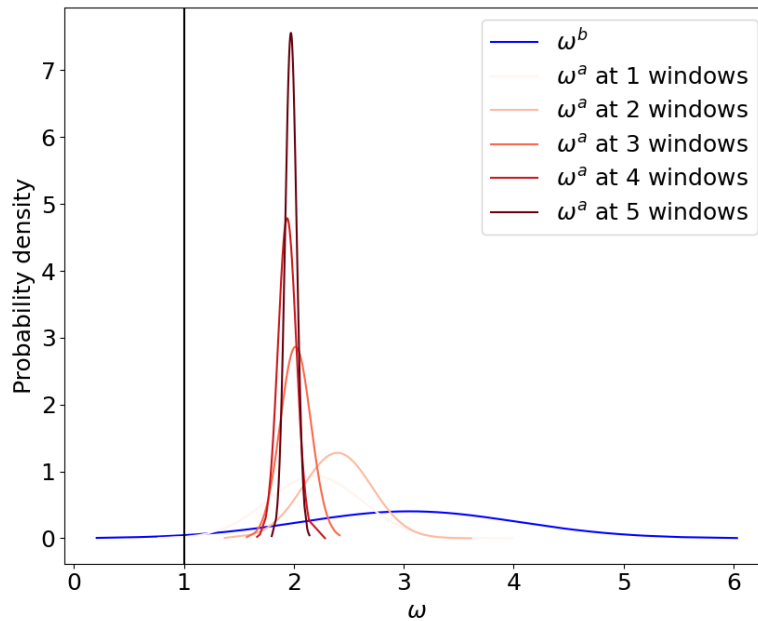
FIGURE 5.5. PDF of the estimated parameter within 5 simulation windows
and with observation every $2nd$ grid points, the black vertical line is the
value of the real ω of the model error.

the estimations get stuck in local minima around $\omega = 2.0$. A detailed look at the figure

shows that the pdfs wanders around the $\omega = 2.0$ value, with the estimate after 4 windows

closer to the truth value than after 5 windows. This suggests that there is indeed a local

minimum for this parameter estimation problem. We then pushed to observing every grid

point at the end of the 10-hour window, and the result is depicted in Figure 5.6. The

convergence is rapid, and after 3 assimilation windows the parameter pdf has converged

around the true value. This suggests that we need such a high observation density to avoid

local minima in the costfunction of each ensemble member.

Although we managed to obtain a good centering of the posterior pdf around the cor-

rect value of ω, its spread is likely too narrow, especially as it is still reducing with each

assimilation window. This is a well-known problem in sequential parameter estimation,

and is caused by the fact that each assimilation will reduce the spread, but there is no error

FIGURE 5.6. PDF of the estimated parameter within 5 simulation windows and with observation every grid point, the black vertical line is the value of the real ω of the model error.

growth in the parameter pdf between assimilations. This can be solved by artificially inflating the parameter pdf between assimilation windows. We have not experimented with that any further.

5.5 Conclusion

With high-dimensional and nonlinear models, the experiment results can be more realistic and representative for real numerical forecasting systems. More complex models indeed bring more challenges that requires more time and higher computational power for the data assimilation problem. In this chapter, we extend our experiments of model error parameter estimation to the 2-layer quasi-geostrophic model. Due to the high-computational

cost of very dense observation in space, we use a smoother in the ensemble subspace (described in Evensen, Vossepoel, and Van Leeuwen, 2022). In the experiments, the parameter encoded in the model error autocorrelation is updated simultaneously with the state variables using the state augmentation methods.

The update of the state variables is simpler compared with the parameter estimation. Even with sparse observations in both time and space, the ensemble smoother manages to have a decent update for the variable. The ensemble smoother indeed minimize the ensemble spread (ensemble variance) quite well, and the RMSE matches the results for the ensemble spread at the observation time-step which indicates the performance of the smoother is working well. On the other hand, parameter estimation with the QG model is much more difficult than performing parameter estimation on the linear model and logistic map. The variance of the ensemble is minimized too fast for the smoother to find the real ω, and only with very dense observations in space will bring the smoother to the correct value of the parameter.

Although we managed to create an experimental situation in which estimating the model error parameter was mostly successful, we are not in a position to say that is now a solved problem. First, even in the successful case the uncertainty in the posterior parameter values is underestimated, which is, as discussed, a general problem in sequential parameter estimation. Second, while we managed to estimate the parameter by only using observations at the end of a 10-hour window, we did need to observe the full state. It is not clear why such high observation density is needed, and it is indeed counter-intuitive given the scales of the features in the QG model.

On the other hand, we can come up with some interesting observations. It is perhaps remarkable that we can estimate model error parameters using only observations every 10 time-steps, while the real decorrelation time scale of the model errors is only 1 time-step.

It has to be related to the quality of the smoother, and we speculate that this will be hard to achieve with a filer. This brings us automatically to the fact that we could use 500 ensemble members in this still low-dimensional model compared to e.g. NWP. This high ensemble size allowed us to forgo localization. While spatial localization is well studied and for many cases a tuning problem, time localization is still in its infancy, after at least 20 years of active research. We anticipate that if the ensemble size was much smaller and especially time localization would be needed the results might be less favorable as advection will project observation influences to the wrong positions.

Chapter 6

Conclusion and future work

6.1 Conclusion

The aim of this thesis was to evaluate the influence of misspecified time correlations in model errors on data assimilation results and devise ways to estimate the statistics of these time correlations. This is particularly important for next-generation weather forecasting since much more powerful computers were built that can process much more complex algorithms now, and the model error has shown its impact on the accuracy of the weather forecasts. We first investigated the performance of the EnKS with autocorrelated model error and a simple linear model. We tested the behavior of the EnKS for updating the state of the system when we overestimated or underestimated the desired parameter in the model error autocorrelation. We next performed some simple experiments to estimate a single parameter in the model error autocorrelation. After experimenting with the linear system, we then extended our investigations to a relatively more complex circumstance, a simple nonlinear system. Meanwhile, we considered more realistic formulations of the model error autocorrelation structure. Finally, we ended our project with a relatively realistic model, the 2-layer quasi-geostrophic model, implemented with the model error autocorrelation we proposed. In section 1.2, we proposed several research questions in this thesis, and the main conclusions of the research questions are listed as follows:

1. **What impact does the temporal autocorrelated model error have on the data assimilation results?**

In our experiments, we proposed a first guess of the estimated parameter, which can be underestimated or overestimated with respect to the true values of the parameters. With incorrect autocorrelation of the model error using an ensemble smoother, we found that in the low-dimensional linear system, the actual error (MSE) of the posterior is larger than the estimated error (ensemble variance) from the data assimilation scheme when the temporal parameter in the model error autocorrelation is overestimated, and vice versa. The results of a higher-dimensional linear system with the EnKS agreed with the results in the low-dimensional case. In general, the influence of a misspecified autocorrelated model error can be significant for the data assimilation results.

2. **Is it possible to establish a way to update the parameters encoded in the model error autocorrelation online during the data assimilation procedure?**

   In the experiment with a simple linear system when only a decaying parameter is considered in the model error correlation, parameter estimation can be successfully performed online using the state augmentation method. However, even in this most straightforward case, to ensure a sufficiently quick convergence of the estimated parameter to the correct value, dense observations are required. For more complex formulations of the model error autocorrelations when more parameters are considered, and applied to nonlinear models, parameter estimation confronts more complications. When more than one parameter is considered, even with the low-dimensional linear system, the EnKS cannot provide a satisfying estimation of the parameters. The cost functions show highly irregular shapes, so that the convergence cannot be achieved in one step. Therefore, the iterative version of the EnKS (IEnKS) was applied to fix the problem and turned out to be successful for the linear system. However, with extra nonlinearity provided by the nonlinear systems, there are more challenges when estimating multiple parameters. We have to carefully tune the IEnKS, and transform the problem by first estimating the lag-1 autocorrelation in the model error, from which the actual parameter values are then

recovered. Furthermore, we find that the initial guess of the parameters can also influence the parameter estimation results.

3. **How does our estimation method perform in a more complex and higher-dimensional system, such as the Quasi-geostrophic (QG) model?**

    A higher-dimensional nonlinear system makes the parameter estimation problem significantly more difficult finding the right answer, but the results are also more realistic. With a simple setting for the model error autocorrelation, we manage to estimate the parameter using the state augmentation method successfully. However, the uncertainty in the posterior parameter values is underestimated, and in the cases of successful estimation, we do need to observe the entire streamfunction field, which can be unaffordable in time and computational power within a real forecasting system. With less dense observations of the field, the posterior of the parameter finds the right direction but fails to converge to the correct value and gets stuck with local minima. On the other hand, with only observations every 10 time-steps, the smoother can still find the correct value of the parameter, even when the real autocorrelation timescale of the model error is 1. This is perhaps a remarkable result and has to be related to the quality of the smoother. There are still many aspects of the experiments on the QG model, and perhaps some other high-dimensional nonlinear systems, that can be further investigated.

The answers to those questions in section 1.2 summarise the main results of my four-year PhD research project. We will shortly discuss possible future work in the next section.

6.2   Future work

The experiment of parameter estimation with the QG model is just the start of a much larger topic. There are still many aspects of this topic that can be investigated. For instance, higher temporal and lower spatial resolution in the IEnKS for the QG model can be

implemented to understand the temporal-spatial interplay. A possible, more challenging experiment is to include not only the misspecified temporal autocorrelation in the model error and perform the parameter estimation online, which is already a challenging problem, but also consider an incorrect spatial correlation in the model error. Furthermore, as mentioned in chapter 6, the ensemble size we use in our experiments is quite large and avoids the need for localization. Thus, in the cases of smaller ensemble sizes, we can investigate the influence of localization. As the parameter estimation results shown in chapter 6, we can see that the posterior pdf variances of the parameter are minimized very small and fast. An alternative way to improve the chance of successful parameter estimation is by using the perturbation method from parameters between assimilation windows to broaden the posterior pdf.

For the longer term, to implement online parameter estimation on numerical forecasting systems, we can apply some more realistic models. We can also use different data assimilation methods that may be candidates for the next-generation operational data assimilation schemes, perhaps fully nonlinear data assimilation methods. With higher-computational power and state-of-the-art data assimilation schemes, model error parameter estimation within the operational data assimilation scheme will be improved in the foreseeable future. With this online-parameter-estimation scheme, we will be able to calibrate the weather forecasting system to a higher accuracy. On the other hand, it will also benefit the data assimilation results.

Appendix A

Appendix A

A.1 The prior variance, prior covariance and prior correlation

TABLE I. Expressions of the prior variance, prior covariance and prior correlation in different scenarios.

		$\boldsymbol{\alpha} = 0.0$	$0 < \boldsymbol{\alpha} < 1.0$	$\boldsymbol{\alpha} = 1.0$
$\boldsymbol{\omega} = 0.0$	**Var**(x_n)	q^2	$\alpha^{2n}b^2 + q^2\frac{\alpha^{2n}-1}{\alpha^2-1}$	$b^2 + q^2 n$
	Cov(x_n, x_m)	$q^2\delta_{nm}$	$\alpha^{m+n}b^2 + q^2\frac{\alpha^{m+n}-\alpha^{m-n}}{\alpha^2-1}$	$b^2 + q^2 n$
	Corr(x_n, x_m)	δ_{nm}	$\frac{Cov(x_n,x_m)}{\sqrt{Var(x_n)Var(x_m)}}$	$\sqrt{\frac{b^2+q^2 n}{b^2+q^2 m}}$
$\boldsymbol{\omega} > 0.0$, but finite	**Var**(x_n)	q^2	$\alpha^{2n} + q^2\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha^{2n-i-j}e^{\frac{-|i-j|}{\omega}}$	$b^2 + q^2\sum_{i=1}^{n}\sum_{j=1}^{n}e^{\frac{-|i-j|}{\omega}}$
	Cov(x_n, x_m)	$q^2 e^{\frac{-(m-n)}{\omega}}$	$\alpha^{m+n} + q^2\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha^{m+n-i-j}e^{\frac{-|i-j|}{\omega}}$	$b^2 + q^2\sum_{i=1}^{n}\sum_{j=1}^{m}e^{\frac{-|i-j|}{\omega}}$
	Corr(x_n, x_m)	$e^{\frac{-(m-n)}{\omega}}$	$\frac{Cov(x_n,x_m)}{\sqrt{Var(x_n)Var(x_m)}}$	$\frac{Cov(x_n,x_m)}{\sqrt{Var(x_n)Var(x_m)}}$
$\boldsymbol{\omega} \to \infty$	**Var**(x_n)	q^2	$\alpha^{2n} + q^2\left(\frac{\alpha^n-1}{\alpha-1}\right)^2$	$b^2 + q^2 n^2$
	Cov(x_n, x_m)	q^2	$\alpha^{m+n}b^2 + q^2\frac{(\alpha^m-1)(\alpha^n-1)}{(\alpha-1)^2}$	$b^2 + q^2 nm$
	Corr(x_n, x_m)	1	$\frac{Cov(x_n,x_m)}{\sqrt{Var(x_n)Var(x_m)}}$	$\frac{b^2+q^2 nm}{\sqrt{(b^2+q^2 n^2)(b^2+q^2 m^2)}}$

A.2　The posterior variance with a single observation at the end of the window

TABLE II. Expressions of the posterior variance in different scenarios.

	$\boldsymbol{\alpha} = 0.0$	$0 < \boldsymbol{\alpha} < 1.0$	$\boldsymbol{\alpha} = 1.0$
$\boldsymbol{\omega} = 0.0$	$q^2 - \frac{q^4 \delta_{t\tau}}{q^2 + r^2}$	$\lim\limits_{\omega \to 0.0} Var(x_t, x_\tau) - \frac{Cov(x_t, x_\tau)^2}{Var(x_\tau) + r^2}$	$(b^2 + q^2 t) - \frac{(b^2 + q^2 t)^2}{(b^2 + q^2 \tau) + r^2}$
$0.0 < \boldsymbol{\omega} \ll \infty$	$q^2 - \frac{q^4 e^{\frac{-2(t-\tau)}{\omega}}}{q^2 + r^2}$	$Var(x_t, x_\tau) - \frac{Cov(x_t, x_\tau)^2}{Var(x_\tau) + r^2}$	$\lim\limits_{\alpha \to 1.0} Var(x_t, x_\tau) - \frac{Cov(x_t, x_\tau)^2}{Var(x_\tau) + r^2}$
$\boldsymbol{\omega} \to \infty$	$\frac{q^2 r^2}{q^2 + r^2}$	$\lim\limits_{\omega \to \infty} Var(x_t, x_\tau) - \frac{Cov(x_t, x_\tau)^2}{Var(x_\tau) + r^2}$	$(b^2 + q^2 t^2) - \frac{(b^2 + q^2 t\tau)^2}{b^2 + q^2 \tau^2 + r^2}$

A.3　The Mean-Square Error in the finite ensemble and scalar case

Let us start with the simplest case for the finite ensemble member size with only one observation at $t = \tau$, with the ensemble members having the same distribution as the truth and hence the same model-error memory. The ensemble size is N_e, and the ensemble members $\{x_t^{b,1}, x_t^{b,2}, ..., x_t^{b,N_e}\}$ and the truth are drawn from $\mathcal{N}(\mu_t^b, B^{t2})$. The sample mean, $\bar{x}_t^{b,N_e}$, has the distribution $\sim \mathcal{N}(\mu_t^b, \frac{B^{t2}}{N_e})$, and the MSE of the prior sample mean is given by:

$$
\begin{aligned}
E_{x_t^r}[(\bar{x}_t^{b,N_e} - x_t^r)^2] &= E_{x_t^r}[\{(\bar{x}_t^{b,N_e} - \nu_t^b) - (x_t^r - \nu_t^b)\}^2] \\
&= E_{x_t^r}[(\bar{x}_t^{b,N_e} - \nu_t^b)^2] + E_{x_t^r}[(x_t^r - \nu_t^b)^2] = \frac{B^{t2}}{N_e} + B^{t2}.
\end{aligned}
\tag{A.1}
$$

As we can see, the prior MSE under the perfect assumption is just the variance of the truth if $N_e \to \infty$. In this case, the posterior MSE of the ensemble mean can be computed as:

$$
\begin{aligned}
E_{x_t^r}[(\bar{x}_t^{a,N_e} - x_t^r)^2] &= E_{x_t^r}[\{\bar{x}_t^{b,N_e} + K_t^{x,\omega_r}(\bar{y} - H\bar{x}_\tau^{N_e}) - x_t^r\}^2] \\
&= B^{t2} - K_t^{x,\omega_r} Cov(x_t^r, x_\tau^r) + \frac{B^{t2}}{N_e} - K_t^{x,\omega_r} Cov(\bar{x}_t^{b,N_e}, \bar{x}_\tau^{b,N_e}).
\end{aligned}
\tag{A.2}
$$

Even with the ideal assumptions, the posterior MSE for the finite ensemble case is not as simple as the prior MSE. The first two terms on the RHS represents the real MSE of the posterior, and the rest is the sampling error. Note that the Kalman Gain is optimal since $\omega^g = \omega^r$.

Now, let's take the different memory scales in the model error into account. These lead to different variances of the prior ensemble mean and the truth: $\bar{x}_t^{b,N_e} \sim \mathcal{N}(\mu_t^b, \frac{\beta^{t2}}{N_e})$, and $x_t^b \sim \mathcal{N}(\mu_t^b, B_t^{\,2})$. Thus, we have a slightly different expression for the prior MSE:

$$
E_{x_t^r}[(\bar{x}_t^{b,N_e} - x_t^b)^2] = \frac{\beta^{t2}}{N_e} + B^{t2}.
\tag{A.3}
$$

Comparing with the expression shown in Equation (A.1), if we increase the ensemble size to infinity, the prior MSE is the same; it is just the variance of the truth. But the sampling error part is different. As for the posterior MSE in this scenario:

$$
\begin{aligned}
E_{x_t^r}[(\bar{x}_t^{a,N_e} - x_t^r)^2] &= E_{x_t^r}[\{\bar{x}_t^{b,N_e} + K_t^{x,N_e}(\bar{y} - H\bar{x}_\tau^{r,N_e}) - x_t^r\}^2] \\
&= B_t^{\,2} + \left(K_t^{x,\omega_g}\right)^2 (B_\tau^{\,2} + r^2) - 2K_t^{x,\omega_g} Cov(x_t^r, x_\tau^r) + \frac{\beta^{t2}}{N_e} \\
&\quad - K_t^{x,\omega_g} Cov(\bar{x}_t^{b,N_e}, \bar{x}_t^{b,N_e}).
\end{aligned}
\tag{A.4}
$$

In this case the Kalman Gain K_t^{x,ω_g} is not optimal.

At last, let's consider the most different case, when we have no knowledge in both the model error memory and its mean. In this case, we also have a bias in the mean of the prior ensemble members besides incorrect variance: $\bar{x}_t^{b,N_e} \sim \mathcal{N}(\tilde{\mu}_t^b, \frac{\beta^{t2}}{N_e})$. Then, the prior

MSE of the ensemble mean has a similar expression as in Equation 3.28:

$$E_{x_t^r}[(\bar{x}_t^{b,N_e} - x_t^r)^2] = \frac{\beta^{t2}}{N_e} + B_t^2 + (\mu_t^b - \tilde{\mu}_t^b)^2. \qquad (A.5)$$

In this scenario extra errors come from the bias in the mean. The posterior MSE becomes:

$$\begin{aligned} E_{x_t^r}[(\bar{x}_t^{a,N_e} - x_t^r)^2] &= E_{x_t^r}[\{\bar{x}_t^{b,N_e} + K_t^{x,\omega_g}(\bar{y} - H\bar{x}_\tau^{b,N_e}) - x_t^r\}^2] \\ &= B_t^2 + \left(K_t^{x,\omega_g}\right)^2 \left(B_\tau^2 + r^2\right) - 2K_t^{x,\omega_g} \, Cov(x_t^r, x_\tau^r) + \frac{\beta^{t2}}{N_e} - \quad (A.6) \\ &\quad K_t^{x,\omega_g} Cov(\bar{x}_t^{b,N_e}, \bar{x}_\tau^{b,N_e}) + \left(K_t^{x,\omega_g}\right)^2 (\nu_t - \tilde{\mu}_t^b)^2. \end{aligned}$$

The posterior MSE in this scenario contains the errors that are introduced by the sampling error, incorrect auto-correlation timescale, and the bias term.

Bibliography

Aksoy, Altuğ, Fuqing Zhang, and John W Nielsen-Gammon (2006). "Ensemble-based si-multaneous state and parameter estimation with MM5". *Geophysical Research Letters* 33.12. DOI: `10.1175/MWR3224.1`.

Alley, Richard B, Kerry A Emanuel, and Fuqing Zhang (2019). "Advances in weather prediction". *Science* 363.6425, pp. 342–344. DOI: `10.1126/science.aav7274`.

Amezcua, Javier, Haonan Ren, and Peter Jan Van Leeuwen (2023). "Using the (Iterative) Ensemble Kalman Smoother to Estimate the Time Correlation in Model Error". *Tellus A: Dynamic Meteorology and Oceanography* 75.1, pp. 208–128. DOI: `10.16993/tellusa.55`.

Amezcua, Javier and Peter Jan Van Leeuwen (2018). "Time-correlated model error in the (ensemble) Kalman smoother". *Quarterly Journal of the Royal Meteorological Society* 144.717, pp. 2650–2665. DOI: `10.1002/qj.3378`.

Anderson, Jeffrey L (2001). "An ensemble adjustment Kalman filter for data assimila-tion". *Monthly weather review* 129.12, pp. 2884–2903. DOI: `10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2`.

Anderson, Jeffrey L and Stephen L Anderson (1999). "A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts". *Monthly weather review* 127.12, pp. 2741–2758. DOI: `10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2`.

Arnold, HM, IM Moroz, and TN Palmer (2013). "Stochastic parametrizations and model uncertainty in the Lorenz'96 system". *Philosophical Transactions of the Royal Society*

A: Mathematical, Physical and Engineering Sciences 371.1991, p. 20110479. DOI: 10.1098/rsta.2011.0479.

Asch, Mark, Marc Bocquet, and Maëlle Nodet (2016). *Data assimilation: methods, algorithms, and applications*. SIAM.

Axell, Lars and Ye Liu (2016). "Application of 3-D ensemble variational data assimilation to a Baltic Sea reanalysis 1989–2013". *Tellus A: Dynamic Meteorology and Oceanography* 68.1, p. 24220. DOI: 10.3402/tellusa.v68.24220.

Bannister, RN (2017). "A review of operational methods of variational and ensemble-variational data assimilation". *Quarterly Journal of the Royal Meteorological Society* 143.703, pp. 607–633. DOI: 10.1002/qj.2982.

Bauer, Peter, Alan Thorpe, and Gilbert Brunet (2015). "The quiet revolution of numerical weather prediction". *Nature* 525.7567, pp. 47–55. DOI: 10.1038/nature14956.

Bayes, T. (1763). "An essay towards solving a problem in the doctrine of chances". *Philosophical Transactions of the Royal Society of London* 53, pp. 370–418. DOI: 10.1098/rstl.1763.0053.

Bell, Michael J, MJ Martin, and NK Nichols (2004). "Assimilation of data into an ocean model with systematic errors near the equator". *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 130.598, pp. 873–893. DOI: 10.1256/qj.02.109.

Bennett, Andrew F (1992). *Inverse methods in physical oceanography*. Cambridge university press.

Bertino, Laurent, Geir Evensen, and Hans Wackernagel (2003). "Sequential data assimilation techniques in oceanography". *International Statistical Review* 71.2, pp. 223–241. DOI: 10.1111/j.1751-5823.2003.tb00194.x.

Bocquet, Marc and Pavel Sakov (2014). "An iterative ensemble Kalman smoother". *Quarterly Journal of the Royal Meteorological Society* 140.682, pp. 1521–1535. DOI: `10.1002/qj.2236`.

Bonavita, M. and P. Laloyaux (2020). "Machine learning for model error inference and correction". *Journal of Advances in Modeling Earth Systems* 12.12, pp. 1–22. DOI: `10.1029/2020MS002232`.

Bonavita, Massimo (2021). "Exploring the structure of time-correlated model errors in the ECMWF data assimilation system". *Quarterly Journal of the Royal Meteorological Society* 147.739, pp. 3454–3471. DOI: `10.1002/qj.4137`.

Bonavita, Massimo, Elias Hólm, Lars Isaksen, and Mike Fisher (2016). "The evolution of the ECMWF hybrid data assimilation system". *Quarterly Journal of the Royal Meteorological Society* 142.694, pp. 287–303. DOI: `10.1002/qj.2652`.

Bony, S., B. Stevens, D.and Ch. Jacob Frierson, M Kageyama, R Pincus, Th G Sheperd, S C Sherwood, A P Siebesma, A H Sobel, M Watanabe, and M Webb (2015). "Clouds, circulation and climate sensitivity". *Nature Geoscience* 8, pp. 261–268. DOI: `10.1038/ngeo2398`.

Brajard, J., A. Carrassi, M. Bocquet, and L. Bertino (2021). "Combining data assimilation and machine learning to infer unresolved scale parametrization". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379, p. 20200086. DOI: `10.1098/rsta.2020.0086`.

Brasseur, P., P. Bahurel, L. Bertino, F. Birol, J.-M. Brankart, N. Ferry, and et al. (2005). "Data assimilation for marine monitoring and prediction: The MERCATOR operational assimilation systems and the MERSEA developments". *Quarterly Journal of the Royal Meteorological Society* 131, pp. 3561–3582. DOI: `10.1256/qj.05.142`.

Carrassi, Alberto and Stéphane Vannitsem (2010). "Accounting for model error in variational data assimilation: A deterministic formulation". *Monthly Weather Review* 138.9, pp. 3369–3386. DOI: `10.1175/2010MWR3192.1`.

Carrassi, Alberto and Stéphane Vannitsem (2011). "State and parameter estimation with the extended Kalman filter: an alternative formulation of the model error dynamics". *Quarterly Journal of the Royal Meteorological Society* 137.655, pp. 435–451. DOI: `10.1002/qj.762`.

Carton, James A, Gennady Chepurin, Xianhe Cao, and Benjamin Giese (2000). "A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part I: Methodology". *Journal of Physical Oceanography* 30.2, pp. 294–309. DOI: `10.1175/1520-0485(2000)030<0294:ASODAA>2.0.CO;2`.

Cohn, Stephen E and David F Parrish (1991). "The behavior of forecast error covariances for a Kalman filter in two dimensions". *Monthly Weather Review* 119.8, pp. 1757–1785. DOI: `10.1175/1520-0493(1991)119<1757:TBOFEC>2.0.CO;2`.

Comin, Alcimoni Nelci, Vanucia Schumacher, Flávio Justino, and Alfonso Fernandez (2018). "Impact of different microphysical parameterizations on extreme snowfall events in the Southern Andes". *Weather and climate extremes* 21, pp. 65–75. DOI: `10.1016/j.wace.2018.07.001`.

Courtier, Philippe, E Andersson, W Heckley, D Vasiljevic, M Hamrud, A Hollingsworth, F Rabier, M Fisher, and J Pailleux (1998). "The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation". *Quarterly Journal of the Royal Meteorological Society* 124.550, pp. 1783–1807. DOI: `10.1002/qj.49712455002`.

Courtier, Philippe and Olivier Talagrand (1987). "Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results". *Quarterly Journal of the Royal Meteorological Society* 113.478, pp. 1329–1347. DOI: 10.1002/qj.49711347813.

Couvreux, Fleur, Frédéric Hourdin, Daniel Williamson, Romain Roehrig, Victoria Volodina, Najda Villefranque, Catherine Rio, Olivier Audouin, James Salter, Eric Bazile, et al. (2021). "Process-based climate model development harnessing machine learning: I. A calibration tool for parameterization improvement". *Journal of Advances in Modeling Earth Systems* 13.3, e2020MS002217. DOI: 10.1029/2020MS002217.

Crommelin, D. and E. Vanden-Eijnden (2008). "Subgrid-scale parameterization with conditional Markov chains". *Journal of the Atmospheric Sciences* 65, pp. 2661–2675. DOI: 10.1175/2008JAS2566.1.

Dalcher, Amnon and Eugenia Kalnay (1987). "Error growth and predictability in operational ECMWF forecasts". *Tellus A: Dynamic Meteorology and Oceanography* 39.5, pp. 474–491. DOI: 10.3402/tellusa.v39i5.11774.

Daum, Fred, Jim Huang, and Arjang Noushin (2010). "Exact particle flow for nonlinear filters: Seventeen dubious solutions to a first order linear underdetermined PDE". *Signal processing, sensor fusion, and target recognition XIX*. Vol. 7697. SPIE. IEEE, pp. 92–110. DOI: 10.1109/ACSSC.2010.5757468.

Daum, Frederick and Jim Huang (2009). "Nonlinear filters with particle flow induced by log-homotopy". *Signal Processing, Sensor Fusion, and Target Recognition XVIII*. Vol. 7336. SPIE, pp. 76–87. DOI: 10.1117/12.814241.

Dee, Dick P and Arlindo M Da Silva (1998). "Data assimilation in the presence of forecast bias". *Quarterly Journal of the Royal Meteorological Society* 124.545, pp. 269–295. DOI: 10.1256/smsqj.54511.

Dee, Dick P and Ricardo Todling (2000). "Data assimilation in the presence of forecast bias: The GEOS moisture analysis". *Monthly Weather Review* 128.9, pp. 3268–3282. DOI: `10.1175/1520-0493(2000)128<3268:DAITPO>2.0.CO;2`.

Ebi, Kristie L, Jennifer Vanos, Jane W Baldwin, Jesse E Bell, David M Hondula, Nicole A Errett, Katie Hayes, Colleen E Reid, Shubhayu Saha, June Spector, et al. (2021). "Extreme weather and climate change: population health and health system implications". *Annual review of public health* 42.1, pp. 293–315. DOI: `10.1146/annurev-publhealth-012420-105026`.

Errico, Ronald M (1997). "What is an adjoint model?" *Bulletin of the American Meteorological Society* 78.11, pp. 2577–2592. DOI: `10.1175/1520-0477(1997)078<2577:WIAAM>2.0.CO;2`.

Errico, Ronald M, Tomislava Vukićević, and Kevin Raeder (1993). "Examination of the accuracy of a tangent linear model". *Tellus A: Dynamic Meteorology and Oceanography* 45.5, pp. 462–477. DOI: `10.3402/tellusa.v45i5.15046`.

Evensen, Geir (1994). "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics". *Journal of Geophysical Research: Oceans* 99.C5, pp. 10143–10162. DOI: `10.1029/94JC00572`.

— (2009). "The ensemble Kalman filter for combined state and parameter estimation". *IEEE Control Systems Magazine* 29.3, pp. 83–104. DOI: `10.1109/MCS.2009.932223`.

— (2018). "Analysis of iterative ensemble smoothers for solving inverse problems". *Computational Geosciences* 22.3, pp. 885–908. DOI: `10.1007/s10596-018-9731-y`.

— (2021). "Formulating the history matching problem with consistent error statistics". *Computational Geosciences* 25.3, pp. 945–970. DOI: `10.1007/s10596-021-10032-7`.

Evensen, Geir and Nabil Fario (1997). "Solving for the Generalized Inverse of the Lorenz Model (gtSpecial IssueltData Assimilation in Meteology and Oceanography: Theory and Practice)". *Journal of the Meteorological Society of Japan. Ser. II* 75.1B, pp. 229–243. DOI: `10.2151/JMSJ`.

Evensen, Geir, Patrick N Raanes, Andreas S Stordal, and Joakim Hove (2019). "Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching". *Frontiers in Applied Mathematics and Statistics* 5, p. 47. DOI: `10.3389/fams.2019.00047`.

Evensen, Geir and Peter Jan Van Leeuwen (1996). "Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model". *Monthly Weather Review* 124.1, pp. 85–96. DOI: `10.1175/1520-0493(1996)124<0085:AOGADF>2.0.CO;2`.

— (2000). "An ensemble Kalman smoother for nonlinear dynamics". *Monthly Weather Review* 128.6, pp. 1852–1867. DOI: `10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2`.

Evensen, Geir, Femke C Vossepoel, and Peter Jan Van Leeuwen (2022). *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. Springer Nature.

Farchi, Alban, Patrick Laloyaux, Massimo Bonavita, and Marc Bocquet (2021). "Using machine learning to correct model error in data assimilation and forecast applications". *Quarterly Journal of the Royal Meteorological Society* 147.739, pp. 3067–3084. DOI: `10.1002/qj.4116`.

Fennel, Katja, Marion Gehlen, Pierre Brasseur, Christopher W. Brown, Stefano Ciavatta, Gianpiero Cossarini, Alessandro Crise, Christopher A. Edwards, David Ford, Marjorie A. M. Friedrichs, Marilaure Gregoire, Emlyn Jones, Hae-Cheol Kim, Julien Lamouroux, Raghu Murtugudde, Coralie Perruche, the GODAE OceanView Marine

Ecosystem Analysis, and Prediction Task Team (2019). "Advancing Marine Biogeo-chemical and Ecosystem Reanalyses and Forecasts as Tools for Monitoring and Managing Ecosystem Health". *Frontiers in Marine Science* 6, p. 89. DOI: `10.3389/fmars.2019.00089`.

Fisher, R and C Koven (2020). "Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems". *Journal of Advances in Modeling Earth Systems* 12.4. DOI: `10.1029/2018ms001453`.

Fortin, V, M Abaza, F Anctil, and R Turcotte (2014). "Why should ensemble spread match the RMSE of the ensemble mean?" *Journal of Hydrometeorology* 15.4, pp. 1708–1713. DOI: `10.1175/JHM-D-14-0008.1`.

Fox-Kemper, Baylor, Alistair Adcroft, Claus W Böning, Eric P Chassignet, Enrique Curchitser, Gokhan Danabasoglu, Carsten Eden, Matthew H England, Rüdiger Gerdes, Richard J Greatbatch, et al. (2019). "Challenges and prospects in ocean circulation models". *Frontiers in Marine Science* 6, p. 65. DOI: `10.3389/fmars.2019.00065`.

Ghil, M (1989). "Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework". *Dynamics of Atmospheres and Oceans* 13.3-4, pp. 171–218. DOI: `10.1016/0377-0265(89)90040-7`.

Ghil, Michael, S Cohn, John Tavantzis, K Bube, and Eugene Isaacson (1981). "Applications of estimation theory to numerical weather prediction". *Dynamic meteorology: Data assimilation methods*, pp. 139–224. DOI: `10.1007/978-1-4612-5970-1_5`.

Goodliff, Michael, Javier Amezcua, and Peter Jan Van Leeuwen (2015). "Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity". *Tellus A: Dynamic Meteorology and Oceanography* 67.1, p. 26928. DOI: `10.1007/978-1-4612-5970-1_5`.

Gordon, Neil J, David J Salmond, and Adrian FM Smith (1993). "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". *IEEE proceedings F (radar and signal processing)*. Vol. 140. 2. IET, pp. 107–113. DOI: `10.1049/ip-f-2.1993.0015`.

Gratton, Serge, Amos S Lawless, and Nancy K Nichols (2007). "Approximate Gauss–Newton methods for nonlinear least squares problems". *SIAM Journal on Optimization* 18.1, pp. 106–132. DOI: `10.1137/050624935`.

Griffith, Anne K and Nancy K Nichols (2000). "Adjoint methods in data assimilation for estimating model error". *Flow, turbulence and combustion* 65, pp. 469–488. DOI: `10.1023/A:1011454109203`.

Hamill, Thomas M and Chris Snyder (2000). "A hybrid ensemble Kalman filter–3D variational analysis scheme". *Monthly Weather Review* 128.8, pp. 2905–2919. DOI: `10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2`.

Hanlon, Helen M, Dan Bernie, Giulia Carigi, and Jason A Lowe (2021). "Future changes to high impact weather in the UK". *Climatic Change* 166.3, pp. 1–23. DOI: `10.1007/s10584-021-03100-5`.

Hansen, James A and Cécile Penland (2007). "On stochastic parameter estimation using data assimilation". *Physica D: Nonlinear Phenomena* 230.1-2, pp. 88–98. DOI: `10.1016/j.physd.2006.11.006`.

Heemink, Arnold Willem, Martinus Verlaan, and Adrianus Johannes Segers (2001). "Variance reduced ensemble Kalman filtering". *Monthly Weather Review* 129.7, pp. 1718–1728. DOI: `10.1175/1520-0493(2001)129<1718:VREKF>2.0.CO;2`.

Houtekamer, Peter L and Herschel L Mitchell (1998). "Data assimilation using an ensemble Kalman filter technique". *Monthly Weather Review* 126.3, pp. 796–811. DOI: `10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2`.

Howes, KE, Alison M Fowler, and AS Lawless (2017). "Accounting for model error in strong-constraint 4D-Var data assimilation". *Quarterly Journal of the Royal Meteorological Society* 143.704, pp. 1227–1240. DOI: `10.1002/qj.2996`.

Hu, Chih-Chi and Peter Jan Van Leeuwen (2021). "A particle flow filter for high-dimensional system applications". *Quarterly Journal of the Royal Meteorological Society* 147.737, pp. 2352–2374. DOI: `10.1002/qj.4028`.

Isard, Michael and Andrew Blake (1998). "Condensation—conditional density propagation for visual tracking". *International journal of computer vision* 29.1, pp. 5–28. DOI: `10.1023/A:1008078328650`.

Jazwinski, Andrew H (2007). *Stochastic processes and filtering theory*. Courier Corporation.

Kalman, Rudolph E and Richard S Bucy (1960). "New results in linear filtering and prediction theory". *Journal of Basic Engineering* 83, pp. 95–107. DOI: `10.1115/1.3658902`.

Kalman, Rudolph Emil (1960). "A new approach to linear filtering and prediction problems". *Transaction of the AMSE, Journal of Basic Engineering* 82, pp. 35–45. DOI: `10.1109/9780470544334.ch9`.

Khare, Shree P, Jeffrey L Anderson, Timothy J Hoar, and Douglas Nychka (2008). "An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems". *Tellus A: Dynamic Meteorology and Oceanography* 60.1, pp. 97–112. DOI: `10.1111/j.1600-0870.2007.00281.x`.

Kotsuki, Shunji, Koji Terasaki, Hisashi Yashiro, Hirofumi Tomita, Masaki Satoh, and Takemasa Miyoshi (2018). "Online model parameter estimation with ensemble data assimilation in the real global atmosphere: A case with the nonhydrostatic icosahedral atmospheric model (NICAM) and the global satellite mapping of precipitation data".

Journal of Geophysical Research: Atmospheres 123.14, pp. 7375–7392. DOI: 10 . 1029/2017JD028092.

Krishnamurthy, V (2019). "Predictability of weather and climate". *Earth and Space Science* 6.7, pp. 1043–1056. DOI: 10.1029/2019EA000586.

Kuma, P., A. J. McDonald, O. Morgenstern, S. P. Alexander, J. J. Cassano, S. Garrett, and J. Williams (2018). "Evaluation of southern ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations. Atmospheric Chemistry and Physics". *Atmospheric Chemistry and Physics* 20, pp. 6607–6630. DOI: 10.5194/acp-20-6607-2020.

Kuznetsov, L., K. Ide, and C. K. R. T. Jones (2003). "A Method for Assimilation of Lagrangian Data". *Monthly Weather Review* 131.10, pp. 2247–2260. DOI: 10.1175/ 1520-0493(2003)131<2247:AMFAOL>2.0.CO;2.

Lee, Joshua Chun Kwang, Javier Amezcua, and Ross Noel Bannister (2022). "Hybrid ensemble-variational data assimilation in ABC-DA within a tropical framework". *Geoscientific Model Development* 15.15, pp. 6197–6219. DOI: 10.5194/gmd-15-6197-2022.

Li, Hong, Eugenia Kalnay, Takemasa Miyoshi, and Christopher M Danforth (2009). "Accounting for model errors in ensemble data assimilation". *Monthly Weather Review* 137.10, pp. 3407–3419. DOI: 10.1175/2009MWR2766.1.

Liu, Gaisheng, Yan Chen, and Dongxiao Zhang (2008). "Investigation of flow and transport processes at the MADE site using ensemble Kalman filter". *Advances in Water Resources* 31.7, pp. 975–986. DOI: 10.1016/j.advwatres.2008.03.006.

Lorenc, Andrew C (1986). "Analysis methods for numerical weather prediction". *Quarterly Journal of the Royal Meteorological Society* 112.474, pp. 1177–1194. DOI: 10. 1002/qj.49711247414.

Lorenc, Andrew C, Neill E Bowler, Adam M Clayton, Stephen R Pring, and David Fairbairn (2015). "Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP". *Monthly Weather Review* 143.1, pp. 212–229. DOI: 10.1175/MWR-D-14-00195.1.

Lorenc, Andrew C and Mohamed Jardak (2018). "A comparison of hybrid variational data assimilation methods for global NWP". *Quarterly Journal of the Royal Meteorological Society* 144.717, pp. 2748–2760. DOI: 10.1002/qj.3401.

Lorenz, Edward (1972). *Predictability: does the flap of a butterfly's wing in brazil set off a tornado in texas?* American Association for the Advancement of Sciences; 139th meeting.

Lorenz, Edward N (1963). "Deterministic nonperiodic flow". *Journal of the atmospheric sciences* 20.2, pp. 130–141. DOI: 10.1175/1520-0469(1963)020<0130: DNF>2.0.CO;2.

Lucini, María Magdalena, Peter Jan Van Leeuwen, and Manuel Pulido (2021). "Model error estimation using the expectation maximization algorithm and a particle flow filter". *SIAM/ASA Journal on Uncertainty Quantification* 9.2, pp. 681–707. DOI: 10.1137/19M1297300.

Mao, Fuping, Shuwen Zhang, Dan Ye, and Xixi Yang (2015). "Impact of time correlated model errors on soil moisture estimates with the ensemble square root filter". *Advances in Earth Science* 30.6, p. 700. DOI: 10.11867/j.issn.1001-8166. 2015.06.0700.

Maraun, Douglas (2016). "Bias correcting climate change simulations-a critical review". *Current Climate Change Reports* 2.4, pp. 211–220. DOI: 10.1007/s40641-016-0050-x.

Martin, M. J., N. K. Nichols, and M. J. Bell (1999). *Treatment of Systematic Errors in Sequential Data Assimilation*. Tech. rep. Note 21.

May, Robert (1976). "Simple mathematical models with very complicated dynamics". *Nature* 261, pp. 459–467. DOI: `10.1007/978-0-387-21830-4_7`.

Mazzarella, Vincenzo, Ida Maiello, Vincenzo Capozzi, Giorgio Budillon, and Rossella Ferretti (2017). "Comparison between 3D-Var and 4D-Var data assimilation methods for the simulation of a heavy rainfall case in central Italy". *Advances in Science and Research* 14, pp. 271–278. DOI: `10.5194/asr-14-271-2017`.

Mitchell, Herschel L and PL Houtekamer (2009). "Ensemble Kalman filter configurations and their performance with the logistic map". *Monthly Weather Review* 137.12, pp. 4325–4343. DOI: `10.1175/2009MWR2823.1`.

Muelmenstadt, J and G Feingold (2018). "The Radiative Forcing of Aerosol–Cloud Interactions in Liquid Clouds: Wrestling and Embracing Uncertainty". *Current Climate Change Reports* 4, pp. 23–40. DOI: `10.1007/s40641-018-0089-y`.

Mukhopadhyay, P, VS Prasad, R Phani Murali Krishna, Medha Deshpande, Malay Ganai, Snehlata Tirkey, Sahadat Sarkar, Tanmoy Goswami, CJ Johny, Kumar Roy, et al. (2019). "Performance of a very high-resolution global forecast system model (GFS T1534) at 12.5 km over the Indian region during the 2016–2017 monsoon seasons". *Journal of Earth System Science* 128, pp. 1–18. DOI: `10.1007/s12040-019-1186-6`.

Nerger, Lars, Svenja Schulte, and Angelika Bunse-Gerstner (2014). "On the influence of model nonlinearity and localization on ensemble Kalman smoothing". *Quarterly Journal of the Royal Meteorological Society* 140.684, pp. 2249–2259. DOI: `10.1002/qj.2293`.

Orrell, David, Leonhard Smith, Jan Barkmeijer, and Tim N Palmer (2001). "Model error in weather forecasting". *Nonlinear processes in geophysics* 8.6, pp. 357–371. DOI: `10.5194/npg-8-357-2001`.

Pachauri, Rajendra K, Myles R Allen, Vicente R Barros, John Broome, Wolfgang Cramer, Renate Christ, John A Church, Leon Clarke, Qin Dahe, Purnamita Dasgupta, et al. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC. ISBN: 9789291691432.

Palmer, Tim N (2001). "A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models". *Quarterly Journal of the Royal Meteorological Society* 127.572, pp. 279–304. DOI: `10.1002/qj.49712757202`.

Palmer, TN (2019). "Stochastic weather and climate models". *Nature Reviews Physics* 1.7, pp. 463–471. DOI: `10.1038/s42254-019-0062-2`.

Pathiraja, Sahani Darshika and Peter Jan van Leeuwen (2017). "Model Uncertainty Quantification Methods For Data Assimilation In Partially Observed Multi-Scale Systems". *AGU Fall Meeting Abstracts*. Vol. 2017, NG31A–0154. DOI: `10.48550/arXiv.1807.09621`.

Ren, Haonan, Javier Amezcua, and Peter Jan Van Leeuwen (2021). "Effects of misspecified time-correlated model error in the (ensemble) Kalman Smoother". *Quarterly Journal of the Royal Meteorological Society* 147.734, pp. 573–588. DOI: `10.1002/qj.3934`.

Ruiz, Juan Jose, Manuel Pulido, and Takemasa Miyoshi (2013). "Estimating model parameters with ensemble-based data assimilation: A review". *Journal of the Meteorological Society of Japan. Ser. II* 91.2, pp. 79–99. DOI: `10.2151/jmsj.2013-201`.

Sakov, Pavel, Dean S Oliver, and Laurent Bertino (2012). "An iterative EnKF for strongly nonlinear systems". *Monthly Weather Review* 140.6, pp. 1988–2004. DOI: `10.1175/MWR-D-11-00176.1`.

Särkkä, Simo (2013). *Bayesian filtering and smoothing*. 3. Cambridge university press.

Shepherd, T.G. (2015). "DYNAMICAL METEOROLOGY | Hamiltonian Dynamics". *Encyclopedia of Atmospheric Sciences (Second Edition)*. Ed. by Gerald R. North, John Pyle, and Fuqing Zhang. Second Edition. Oxford: Academic Press, pp. 324–331. ISBN: 978-0-12-382225-3. DOI: https://doi.org/10.1016/B978-0-12-382225-3.00008-6.

Smith, Leonard A and A Mees (2000). *Nonlinear dynamics and statistics*. Springer Science & Business Media.

Szunyogh, Istvan, Eric J Kostelich, G Gyarmati, DJ Patil, Brian R Hunt, Eugenia Kalnay, Edward Ott, and James A Yorke (2005). "Assessing a local ensemble Kalman filter: Perfect model experiments with the National Centers for Environmental Prediction global model". *Tellus A: Dynamic Meteorology and Oceanography* 57.4, pp. 528–545. DOI: 10.3402/tellusa.v57i4.14721.

Tonani, Marina, Peter Sykes, Robert R King, Niall McConnell, Anne-Christine Péquignet, Enda O'Dea, Jennifer A Graham, Jeff Polton, and John Siddorn (2019). "The impact of a new high-resolution ocean model on the Met Office North-West European Shelf forecasting system". *Ocean Science* 15.4, pp. 1133–1158. DOI: 10.5194/os-15-1133-2019.

Tong, Mingjing and Ming Xue (2008). "Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part II: Parameter estimation experiments". *Monthly weather review* 136.5, pp. 1649–1668. DOI: 10.1175/2007MWR2071.1.

Toth, Zoltan and Roberto Buizza (2019). "Weather forecasting: What sets the forecast skill horizon?" *Sub-Seasonal to Seasonal Prediction*. Elsevier, pp. 17–45. DOI: 10.1016/B978-0-12-811714-9.00002-4.

Tremolet, Yannick (2006). "Accounting for an imperfect model in 4D-Var". *Quarterly Journal of the Royal Meteorological Society* 132.621, pp. 2483–2504. DOI: `10.1256/qj.05.224`.

Trudinger, Cathy M, Michael R Raupach, Peter J Rayner, and Ian G Enting (2008). "Using the Kalman filter for parameter estimation in biogeochemical models". *Environmetrics: The official journal of the International Environmetrics Society* 19.8, pp. 849–870. DOI: `10.1002/env.910`.

Van Leeuwen, Peter Jan (2020). "A consistent interpretation of the stochastic version of the Ensemble Kalman Filter". *Quarterly Journal of the Royal Meteorological Society* 146.731, pp. 2815–2825. DOI: `10.1002/qj.3819`.

Van Leeuwen, Peter Jan and Geir Evensen (1996). "Data assimilation and inverse methods in terms of a probabilistic formulation". *Monthly weather review* 124.12, pp. 2898–2913. DOI: `10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2`.

Van Leeuwen, Peter Jan, Hans R Künsch, Lars Nerger, Roland Potthast, and Sebastian Reich (2019). "Particle filters for high-dimensional geoscience applications: A review". *Quarterly Journal of the Royal Meteorological Society* 145.723, pp. 2335–2365. DOI: `10.1002/qj.3551`.

Whitaker, Jeffrey S and Thomas M Hamill (2002). "Ensemble data assimilation without perturbed observations". *Monthly weather review* 130.7, pp. 1913–1924. DOI: `10.1175/1520-0493(2002)130<1913:EDAWPO>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". *Monthly Weather Review* 132.5, pp. 1238–1253. DOI: `10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2`.

Zheng, Yue, Kiran Alapaty, Jerold A Herwehe, Anthony D Del Genio, and Dev Niyogi (2016). "Improving high-resolution weather forecasts using the Weather Research and

Forecasting (WRF) Model with an updated Kain–Fritsch scheme". *Monthly Weather Review* 144.3, pp. 833–860. DOI: 10.1175/MWR-D-15-0005.1.

Zhu, Mengbin, Peter J Van Leeuwen, and Weimin Zhang (2018). "Estimating model error covariances using particle filters". *Quarterly Journal of the Royal Meteorological Society* 144.713, pp. 1310–1320. DOI: 10.1002/qj.3132.

Zupanski, Dusanka and Milija Zupanski (2006). "Model error estimation employing an ensemble data assimilation approach". *Monthly Weather Review* 134.5, pp. 1337–1354. DOI: 10.1175/MWR3125.1.