UNIVERSITÄT POTSDAM
INSTITUT FÜR MATHEMATIK
AG NUMERISCHE MATHEMATIK

Hybrid filters and multi-scale models

Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat)
in der Wissenschaftsdisziplin
"Numerische Mathematik"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von
Maria Reinhardt

Potsdam, den 30. September 2019

1. Gutachter: Prof. Dr. Sebastian Reich
2. Gutachter: Prof. Dr. Rupert Klein
3. Gutachter: Prof. Dr. Roland Potthast

https://doi.org/10.25932/publishup-47435
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-474356

Abstract

This thesis is concerned with Data Assimilation, the process of combining model predictions with observations. So called filters are of special interest. One is interested in computing the probability distribution of the state of a physical process in the future, given (possibly) imperfect measurements. This is done using Bayes' rule. The first part focuses on hybrid filters, that bridge between the two main groups of filters: ensemble Kalman filters (EnKF) and particle filters. The first are a group of very stable and computationally cheap algorithms, but they request certain strong assumptions. Particle filters on the other hand are more generally applicable, but computationally expensive and as such not always suitable for high dimensional systems. Therefore it exists a need to combine both groups to benefit from the advantages of each. This can be achieved by splitting the likelihood function, when assimilating a new observation and treating one part of it with an EnKF and the other part with a particle filter.

The second part of this thesis deals with the application of Data Assimilation to multi-scale models and the problems that arise from that. One of the main areas of application for Data Assimilation techniques is predicting the development of oceans and the atmosphere. These processes involve several scales and often balance relations between the state variables. The use of Data Assimilation procedures most often violates relations of that kind, which leads to unrealistic and non-physical predictions of the future development of the process eventually. This work discusses the inclusion of a post-processing step after each assimilation step, in which a minimisation problem is solved, which penalises the imbalance. This method is tested on four different models, two Hamiltonian systems and two spatially extended models, which adds even more difficulties.

Zusammenfassung

Diese Dissertation beschäftigt sich mit Daten Assimilation - die Kombination von Modellvorhersagen mit Beobachtungen. Sogenannte Filter sind dabei von besonderem Interesse. Diese Algorithmen berechnen die Wahrscheinlichkeitsverteilung des Zustandes eines physikalischen Prozesses in der Zukunft unter der Bedingung, dass wir (meist) fehlerbehaftete Messungen vorliegen haben. Der erste Teil bezieht sich auf Hybridfilter, welche eine Brücke zwischen den beiden Hauptgruppen von Filtern schlagen: Ensemble-Kalman-Filter (EnKF) und Teilchenfilter. Die erstgenannten sind sehr stabil und rechnerisch unaufwändig, aber basieren auf recht starken Voraussetzungen. Teilchenfilter hingegen sind allgemeiner aber rechenintensiv und daher nicht immer geeignet für höherdimensionale Systeme. Daher besteht die Notwendigkeit beide Gruppen zu kombinieren um von den Vorteilen beider Filter zu profitieren. Dies kann erreicht werden, indem man, wenn eine Beobachtung assimiliert werden soll, die Likelihood-Funktion in zwei Teile spaltet und auf den einen Teil einen EnKF und auf den anderen einen Teilchenfilter anwendet.

Der zweite Teil dieser Arbeit befasst sich mit der Anwendung von Daten Assimilation auf mehrskalige Modelle und die Probleme die daraus entstehen. Eines der Hauptanwendungsgebiete für Daten Assimilation ist die Vorhersage der Entwicklung von Ozeanen und der Atmosphäre. Diese Prozesse finden auf mehreren Skalen statt und häufig bestehen Balancerelationen zwischen den Zustandsvariablen. Die Nutzung von Daten Assimilationstechniken zerstört diese Beziehungen häufig, was schließlich zu unrealistischen und unphysikalischen Vorhersagen führt. In dieser Dissertation wird vorgeschlagen, nach jedem Assimilationsschritt ein Minimierungsproblem zu lösen, welches die Imbalance als Strafterm beinhaltet. Diese Methode wird an vier verschiedenen Modellen getestet, zwei Hamiltonische Systeme und zwei Modelle mit räumlicher Ausdehnung, was zusätzliche Schwierigkeiten schafft.

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Dr. Sebastian Reich for providing me with this intriguing and fascinating topic and for his support, help, guidance and inspiration. Without him, this work would have not been possible. I also thank my second supervisor, Prof. Dr. Rupert Klein for many interesting conversations and the insights that resulted from that. I owe special thanks to Gottfried Hastermann, who partly worked with me and who was always happy to help me out with his immense knowledge, comprehension, brilliant ideas, patience and last but not least his friendship.

I am very grateful to my colleague and dearest friend Jana de Wiljes, who was and is always there for me with support on all levels and endless friendship. You made the last years a fantastic and unforgettable time.

I am thankful for all the good times, helpful conversations and the exceptionally pleasant working atmosphere that I owe to my colleagues and friends Sahani Pathiraja, Jakiw Pidstrigach, Paul Rozdeba, Nik Nüsken, Márk Somogyvári, Noa Malem-Shinitski, Christian Molkenthin, Walter Acevedo, Nawinda Chutsagulprom, Yuan Cheng and Liv Heinecke.

I owe special thanks to two close friends of mine: Lashi Bandara and Kay Bergemann for proofreading this thesis and productive discussions. I would also like to thank Lashi for good times, fun, his patience and moral support.

I am deeply grateful for my family: my parents and my grandpa, who always supported me, loved me, advised me, guided me and believed in me. I hope to make you proud.

I want to dedicate my final words of gratitude to my husband Mijail Bley, who makes me feel happy, loved and confident every single day and who is always there for me.

This work was mostly financed by the SFB 1114 "Scaling Cascades in Complex Systems" and partly created in the context of the SFB 1294 "Data Assimilation". Therefore, I thank the "Deutsche Forschungsgemeinschaft" for the financial funding.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 The (discrete) filtering problem 2

1.3 Focus of this work 5

1.4 Outline 7

2 Basics **9**

2.1 Variational Data Assimilation 9

2.2 Kalman filters 11

2.2.1 Ensemble Kalman filters 13

2.2.2 Kalman-Bucy filter 17

2.3 Particle filters 18

2.3.1 SIR 21

2.3.2 ETPF 22

2.4 Localisation and Inflation 25

3 Hybrid filter **30**

3.1 Likelihood Splitting 30

3.1.1 Examples 33

3.1.2 Adaptive choice of α via KL-divergence 38

3.2 Gaussian mixture approach 43

3.2.1 Example: Lorenz-63 46

3.2.2 Example: Lorenz-96 46

4 Multi-scale models **48**

4.1 Highly-oscillatory Hamiltonian models 49

4.1.1 Scenario A 51

4.1.2 Scenario B 54

4.2 Spatially extended models 59

4.2.1 Slow-fast Lorenz-96 model 59

4.2.2 Shallow water equations 62

4.3 Existing methods 69

4.4 Post-processing approach 73

4.4.1 Modified method for scenario B 76

4.5 Experiments 76

4.5.1 Hamiltonian Scenario A 76

4.5.2 Hamiltonian Scenario B 80

4.5.3 Slow-fast Lorenz-96 82

4.5.4 Shallow water equations 84

5 Conclusion **92**

Bibliography **95**

List of Figures

1.1 Propagation of the posterior density to the prior density of the next time step . . . 3

1.2 Prior and posterior density . . . 4

2.1 Systematic representation of an ensemble based filter . . . 15

2.2 The weights of the sequential importance sampling after 1,4 and 7 time steps for $M = 50$ and $M = 150$. . . . 20

2.3 The time evolution of the effective ensemble size M_{eff}. . . . 21

2.4 Two different couplings between two Gaussian random variables . . . 23

2.5 Estimation of the covariance matrix with different ensemble sizes . . . 27

2.6 Localised empirical covariance matrix with different localisation radii 27

3.1 Attractor of the Lorenz-63 model . . . 34

3.2 Hybrid filter (ETPF-ESRF) applied to Lorenz-63 . . . 34

3.3 Hybrid filter (ESRF-ETPF) applied to Lorenz-63 . . . 35

3.4 Hybrid filter (ETPF-ESRF) applied to Lorenz-96 . . . 37

3.5 Hybrid filter (ETPF-ESRF) applied to Lorenz-96, more detailed . . . 37

3.6 Pdfs of bimodal distributions with increasing distance between the two modes. 38

3.7 Estimated KL-Divergence between a bi-modal distribution (with increasing distance between the two modes) and a Gaussian. 39

3.8 Effective sample size from bimodal distributions 40

3.9 Optimal α for the hybrid applied to bimodal distributions 41

3.10 Estimated and experimental curves $\tilde{\alpha}^M$ and α^M respectively for the hybrid filter applied to bimodal distributions. 43

3.11 Gaussian mixture hybrid applied to Lorenz-63 46

3.12 Gaussian mixture hybrid applied to Lorenz-96 47

4.1 Comparison of balanced and unbalanced double pendulum 54

4.2 Total and fast energy of double pendulum 55

4.3 Imbalances of the double pendulum with different stiffness parameters 55

4.4 Solutions of the thermally embedded model with two different integrators 58

4.5 Evolution of the action variable of the thermally embedded model . . 58

4.6 Evolution of the imbalance of the slow-fast Lorenz-96 model 61

4.7 Kernel function $f_{200}(x)$. 69

4.8 Localisation function and resulting imbalance in the 1.5D shallow water equation, for localisation radius $r_{\text{loc}} = 50$ 69

4.9 Localisation function and resulting imbalance in the 1.5D shallow water equation, for localisation radius $r_{\text{loc}} = 300$ 70

4.10 RMSE plotted against γ, double pendulum 77

4.11 Fast energy plotted against γ, double pendulum 78

4.12 RMSE plotted against time, double pendulum 79

4.13 RMSE plotted against α, double pendulum 80

4.14 RMSE plotted against γ, thermally embedded pendulum 81

4.15 Action plotted against γ, thermally embedded pendulum 81

4.16 RMSE plotted against time, thermally embedded pendulum 82

4.17 Imbalance plotted against γ, slow-fast Lorenz-96 83

4.18 RMSE plotted against time, slow-fast Lorenz-96 84

4.19 Initial state of shallow water model 85

4.20 Divergence and RMSE plotted against time, SWE, $r_{\text{loc}} = 1,\ p = 4,\ \rho = 10^{-8}$ 87

4.21 Divergence and RMSE plotted against time, SWE, $r_{\text{loc}} = 1,\ p = 8,\ \rho = 10^{-8}$ 88

4.22 Divergence and RMSE plotted against time, SWE, $r_{\text{loc}} = 1,\ p = 16,\ \rho = 10^{-8}$ 89

4.23 Divergence and RMSE plotted against time, SWE, $r_{\text{loc}} = 1,\ p = 4,\ \rho = 10^{-4}$ 90

4.24 Divergence and RMSE plotted against time, SWE, $r_{\text{loc}} = 3,\ p = 4,\ \rho = 10^{-8}$ 91

Chapter 1

Introduction

1.1 Motivation

Most physical processes can be modelled by differential equations. For example the flows in the atmosphere that determine the weather can be described by the Navier-Stokes equations. For a given initial state, a model allows us to predict the state of the system at a point in time that lies in the future. There are several problems though. First of all, the initial state is usually not known exactly in practical applications. For chaotic systems, the predicted state can differ drastically from the truth even if the error in the assumed initial state was very small. Another issue is that the differential equations might not be solvable analytically and therefore require a discretisation (and therefore approximation of the solution). The third problem is that the models that are used are usually prone to modelling errors and do not describe the system perfectly. Thus, predicting the future state of a physical process using the model alone seems to be impossible.

This painful experience was also encountered by Lewis Fry Richardson during what is today considered as the first attempt in numerical weather prediction (NWP). The Norwegian scientist Vilhelm Bjerknes stated in 1904 that the future state of the atmosphere is theoretically determined by the primitive equations of motion, mass, state and energy (Bjerknes (1904)). Bjerknes pointed out though, that these equations were too complex to be solved exactly and should be simplified and solved numerically. Richardson imagined huge halls, representing the globe, in which countless computers would solve Bjerknes equations to predict the state of the atmosphere all around the world. He came up with a numerical scheme to integrate the model of the atmosphere by hand in 1922 and tried to predict the change in surface pressure over six hours. Unfortunately, the forecast he got was profoundly wrong because of two reasons: the time step he used for the discretisation was too large and the initial conditions were too noisy.

Edward Lorenz stated in the 1960s, that even if the model describing the evolution of the atmosphere was perfect, due to its chaotic behaviour, there will always be an upper limit to the predictability of the weather (Lorenz (1963)).

Most often, one has some observations of the process whose state one would like to predict. For example, one can measure the pressure, the temperature, the humidity and such similar quantities in the atmosphere. These measurements have a dimension much smaller than the state space and therefore the estimation of the state from the data alone would be an ill-posed problem. Nevertheless, it is essential to use such additional information to successfully predict the state of a chaotic system like the weather, by reinitialising the model again and again as observations become available.

This process of combining model predictions with data, is called Data Assimilation. Despite the aforementioned difficulties, real-time NWP using Data Assimilation became operational in the 1950s and the forecasting skills improved drastically towards the end of the 20th century due to the expansion of computational power and availability of satellite data. The main Data Assimilation algorithms, namely variational Data Assimilation, Kalman filters and particle filters, will be described in detail in chapter 2.

1.2 The (discrete) filtering problem

Let $x \in \mathbb{R}^d$ be a random vector that characterizes the state of a physical process. The vector x is distributed according to a known probability distribution with probability density function (pdf) $\pi_0(x)$. The temporal evolution of the process is described by the equation

$$x_{k+1} = \mathcal{M}(x_k) + \eta_k \tag{1.2.1}$$

with $x_k = x(t_k)$, $\mathcal{M} : \mathbb{R}^d \to \mathbb{R}^d$ and $\eta_k \in \mathbb{R}^d$ is a random vector that describes the model error.
Observations $y_k^{\text{obs}} \in \mathbb{R}^p$ become available at discrete time points t_k, $k = 1, 2, \ldots$ and are distributed according to the likelihood function $\pi(y_k^{\text{obs}}|x)$.
The filtering problem is to find the probability distribution $\pi_k(x|y_{1:k}^{\text{obs}})$ of the state at time t_k given observations up to time t_k, i.e. $y_{1:k}^{\text{obs}}$.
This can be done by applying an iterative procedure that consists of two stages. The first part is propagating the pdf from t_k to the next time point at which a measurement is available, namely t_{k+1}. Thus, we want to obtain $\pi_{k+1}(x|y_{1:k}^{\text{obs}})$, which we call the *prior* or *forecast* and this is achieved by using the Chapman-Kolmogorov equation

$$\pi_{k+1}(x|y_{1:k}^{\text{obs}}) = \int \pi(x|x')\pi_k(x'|y_{1:k}^{\text{obs}})dx'. \tag{1.2.2}$$

The so called transition kernel $\pi(x|x')$ describes the probability for ending up in state x at time t_{k+1} when starting at x' at time t_k under the model equation (1.2.1).

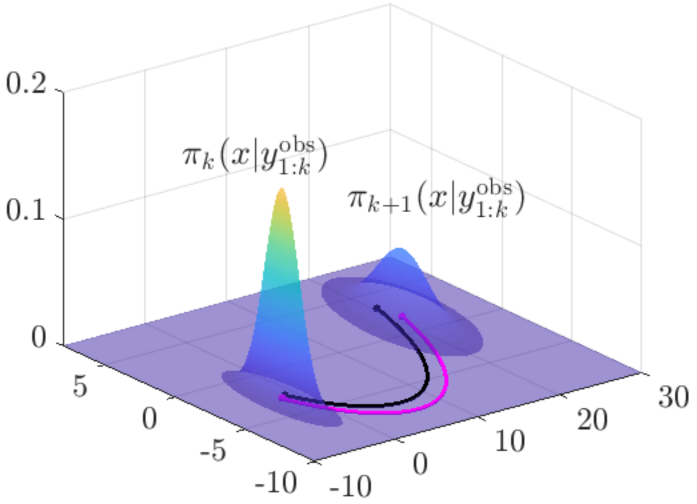


Figure 1.1: Propagation of the posterior density from the last time step t_k to obtain the prior density of the next time step t_{k+1}. The propagation of the mean (magenta) and of a reference solution (black) can also be seen.

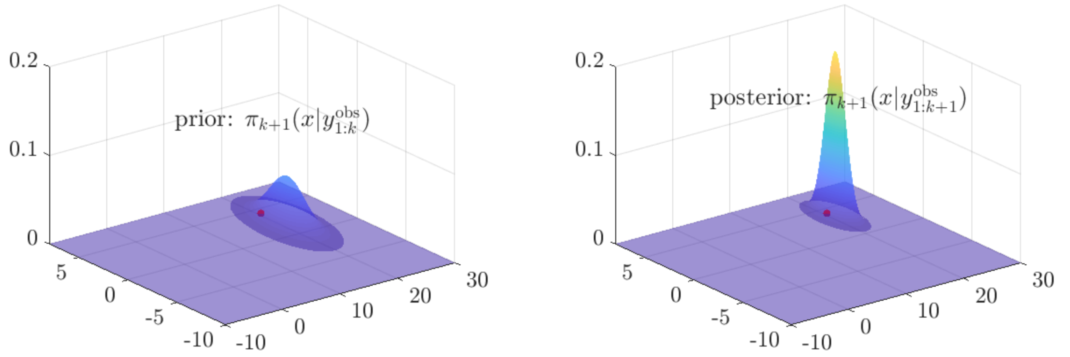
In Figure 1.1 a schematic depiction of the first stage, the propagation of the distribution via equation (1.2.2), is presented.

The second stage of a sequential Data Assimilation method is the application of Bayes' theorem to include the information about the new measurement y_{k+1}^{obs}. The distribution of interest is called the *posterior* or *analysis* and can be obtained via

$$\pi_{k+1}(x|y_{1:k+1}^{\text{obs}}) = \frac{\pi(y_{k+1}^{\text{obs}}|x)\pi_{k+1}(x|y_{1:k}^{\text{obs}})}{\int \pi(y_{k+1}^{\text{obs}}|x)\pi_{k+1}(x|y_{1:k}^{\text{obs}})dx}. \tag{1.2.3}$$

The denominator is just a normalisation constant such that the integral of the posterior equals one, which is a necessary property of a probability distribution. We can simplify and work with the proportionality relation

$$\pi_{k+1}(x|y_{1:k+1}^{\text{obs}}) \propto \pi(y_{k+1}^{\text{obs}}|x)\pi_{k+1}(x|y_{1:k}^{\text{obs}}). \tag{1.2.4}$$



(a) prior

(b) posterior

Figure 1.2: The prior density function with the observation on the left side and the posterior density function with the observation on the right side.

Figure 1.2 shows the second step of the iterative algorithm, namely the transformation of the prior distribution into the posterior distribution via equation (1.2.3). One can see how the mass is becoming more concentrated around the observation.

1.3 Focus of this work

If the initial distribution and the measurement errors are Gaussian and the model is linear, all prior and posterior distributions will be Gaussian and there exists an algorithm, called the Kalman filter (Swerling (1958), Kalman (1960), Kalman & Bucy (1961)), which is known for providing the exact solution to the filtering problem. However, these assumptions are rarely satisfied in practice. Another problem is that one has to store and update the covariance matrix which becomes computationally challenging for high-dimensional systems.

Therefore, in practice a Monte-Carlo approximation of the Kalman filter is often preferred. Instead of forwarding the probability distribution according to the model, one can consider an ensemble of solutions at a certain point in time $x_k^{1,\mathrm{a}}, ..., x_k^{M,\mathrm{a}}$ (the superscript 'a' will be explained later) and evolve them in time, obtaining the *forecast* ensemble $x_{k+1}^{1,\mathrm{f}}, ..., x_{k+1}^{M,\mathrm{f}}$. The superscript 'f' emphasizes that this is the forecast. There are several algorithms that use that strategy to obtain an approximation for the prior mean and covariance matrix by their empirical equivalents and then apply the classic Kalman formulas to them, obtaining the empirical posterior mean and covariance matrix and then produce samples from that, yielding the *analysis* ensemble $x_{k+1}^{1,\mathrm{a}}, ..., x_{k+1}^{M,\mathrm{a}}$. The superscript 'a' emphasizes that this is the analysis. These algorithms belong to the group of ensemble Kalman filters. For example, see Evensen (1994), Burgers et al. (1998), Tippett et al. (2003).

The second big group of sequential filter algorithms are particle filters, as for example described in Del Moral (1996), Gordon et al. (1993), Reich (2013). These methods use importance sampling and do not assume Gaussianity. Therefore they are consistent in the infinite ensemble size limit in contrast to the ensemble Kalman filters. On the other hand, particle filters suffer from the curse of dimensionality and require a much bigger ensemble size than Kalman filters (Doucet (1998), Snyder et al. (2008), Snyder et al. (2015)).

One part of this thesis explores hybrid approaches between an ensemble Kalman filter and a particle filter that exploits the advantages of both groups of filters. Existing approaches can be found in Bengtsson et al. (2003), Mandel & Beezley (2009), Sun et al. (2009), Papadakis et al. (2010), Dovera & Della Rossa (2011) and Frei & Künsch (2013). The method which is part of this thesis, has been developed by Chutsagulprom et al. (2016).

The second focus of this work is on imbalances. Many physical processes act on more than one scale. For example processes in the atmosphere include waves with different frequencies ranging from planetary waves to sound waves. If we want to forecast the behaviour of the whole system, we cannot just ignore the small-scale activities but we have to include them in the model. Often there exist balance relations between the large- and small-scale dynamics of a system, which are approximately preserved under the model evolution. When unbalanced model states are fed into the model, oscillations on the small scale arise and ultimately that leads to non-physical and unrealistic predictions of the system state. The general problem of unbalanced fields for numerical weather prediction is of great importance since a long time, as can be read in Lynch's book 'The Emergence of Numerical Weather Prediction: Richardson's Dream' (Lynch (2006)).

An easy example is the stiff spring pendulum, which is a Hamiltonian system with Hamiltonian

$$H^{\varepsilon}(q,p) = \frac{1}{2}\|p\|^2 + \frac{1}{2\varepsilon^2}(\|q\|-1)^2 + g^0 \begin{pmatrix} 0 & 1 \end{pmatrix} q \tag{1.3.1}$$

and equations of motion

$$\dot{q} = \nabla_p H = p \tag{1.3.2}$$

$$\dot{p} = -\nabla_q H = -\frac{1}{\varepsilon^2}\left(q - \frac{q}{\|q\|}\right) - g^0 \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1.3.3}$$

The state variable is $x = (q,p) \in \mathbb{R}^4$ containing the position q and the momenta p. The parameter ε determines the stiffness of the spring and g^0 is the gravitational constant. If the spring is balanced at the beginning the pendulum would approximately move on the circle with radius 1 around the origin. The balance equation can therefore be formulated as

$$g(q) = \|q\| - 1 \tag{1.3.4}$$

and the so-called slow manifold (a notion introduced by Leith in 1980) is the set

$$\{q \in \mathbb{R}^2 : \|q\| = 1\}. \tag{1.3.5}$$

If we start with an either elongated or compressed spring, or impulses that are not tangential to the slow manifold, high frequency oscillations are excited.

The filter algorithms that we consider are called linear ensemble transform filters (LETFs) and have the characteristic of producing an analysis ensemble, which is a linear combination of the forecast ensemble:

$$x_k^{j,\mathrm{a}} = \sum_{i=1}^{M} x_k^{i,\mathrm{f}} d_{ij}, \quad j = 1, ..., M. \tag{1.3.6}$$

That means, that non-linear balance relations are usually not preserved under an assimilation step. In case of our pendulum, the assimilation step would elongate (or compress) the spring, leading to high frequency oscillations of the spring that are artificially introduced by the Data Assimilation procedure. The phenomenon of Data Assimilation techniques producing unbalanced fields has been studied for example in Bloom et al. (1996), Lorenc (2003*a*) and Ourmières et al. (2006).

In spatio-temporal systems, problems can arise due to the finite ensemble size. It is therefore necessary to apply some kind of localisation technique as in Houtekamer & Mitchell (1998), Houtekamer & Mitchell (2001), Hamill et al. (2001), Ott et al. (2004), Szunyogh et al. (2005) and Hunt et al. (2007). Unfortunately, localisation leads to even more imbalances, which is, for example, discussed in Houtekamer & Mitchell (2005), Cohn et al. (1998), Lorenc (2003*b*), Mitchell et al. (2002), Oke et al. (2007), Kepert (2009) and Greybush et al. (2011).

There exist many approaches on how create more balanced analysis fields in the Data Assimilation context, for example Machenhauer (1977), Baer & Tribbia (1977), Thépaut & Courtier (1991), Lynch & Huang (1992), Bloom et al. (1996), Polavarapu et al. (2000), Gauthier & Thépaut (2001), Neef et al. (2006), Watkinson et al. (2007), Bergemann & Reich (2010), Cotter (2013) and Gottwald (2014). We developed a new strategy to overcome this issue, presented in Reinhardt et al. (2017).

1.4 Outline

We commence the scientific content of this thesis by providing an introduction to the broad area of Data Assimilation in chapter 2. We will give an explanation of variational Data Assimilation, followed by a description of Kalman filters, including ensemble Kalman filters. Afterwards, particle filters will be presented, in particular the sequential importance resampling and the ensemble transform particle filter. At the end of chapter 2, some common issues that arise when doing Data Assimilation, due to the finite sample sizes and at times high dimensions of the problems on hand, will be discussed. Two techniques to help deal with these issues, namely localisation and inflation, will be presented.

Chapter 3 addresses hybrid filters, i.e. possibilities to bridge between the two big groups of filters: Kalman based filters and importance sampling based filters. Two different approaches will be presented and the performance of these filters on two different models, namely the Lorenz-63 and the Lorenz-96 model, will be illustrated. Both are popular toy models in the field of Data Assimilation since they possess important key properties of models of the atmosphere and the oceans, the main application areas for Data Assimilation.

In chapter 4, multi-scale models and the difficulties they pose concerning Data Assimilation are explored. Namely, invalidation of balance relations between the variables that act on different scales and eventually unrealistic results. At first four multi-scale models will be introduced, two of which are spatially extended. After presenting some existing methods to prevent imbalances, our new approach will be explained and finally the results of our numerical experiments in which we applied that approach to the four models will be shown.

Chapter 2

Basics

2.1 Variational Data Assimilation

The method of variational Data Assimilation has been developed in the context of numerical weather prediction. One seeks the minimiser of a cost functional. In three dimensional variational Data Assimilation (3D-Var), we assume we have obtained a *prior* estimate of the state of a model at time t_{k+1} by propagating the posterior estimate of the previous time point

$$x_{k+1}^{f} = \mathcal{M}(x_k^a). \tag{2.1.1}$$

We ignore the possible existence of a model error as in (1.2.1), but we model the general uncertainty in the forecasts by a fixed background error covariance matrix P^b, which is constant in time. We assume that the new incoming observation can be modelled by $y_{\text{obs}}(t_{k+1}) = Hx_k + \varepsilon_k$ with $\varepsilon \sim \mathcal{N}(0, R)$, i.e. the measurement error is assumed to be normally distributed with mean 0 and covariance matrix R. Now, the forecast x_{k+1}^{f} is updated to the analysis by minimising the cost functional

$$\begin{aligned}\mathcal{J}(x) = \frac{1}{2}\left(x_{k+1}^{\mathrm{f}} - x\right)^{\mathrm{T}}\left(P^b\right)^{-1}\left(x_{k+1}^{\mathrm{f}} - x\right) + \\ \frac{1}{2}\left(y_{\text{obs}}(t_{k+1}) - Hx\right)^{\mathrm{T}} R^{-1}\left(y_{\text{obs}}(t_{k+1}) - Hx\right).\end{aligned} \tag{2.1.2}$$

Solving this minimisation problem is essentially solving the ill-posed inverse problem $Hx = y_{\text{obs}}(t_{k+1})$ using a weighted L^2 norm $\|x - x^{\text{f}}_{k+1}\|_{\left(P^b\right)^{-1}}$ as regularisation.

The gradient of $\mathcal{J}$ is given by

$$\nabla_x \mathcal{J}(x) = \left(P^b\right)^{-1}\left(x^{\text{f}}_{k+1} - x\right) + H^{\text{T}} R^{-1}\left(y_{\text{obs}}(t_{k+1}) - Hx\right). \tag{2.1.3}$$

Therefore, the solution can be calculated explicitly as

$$x^{\text{a}}_{k+1} = \left(\left(P^b\right)^{-1} + H^{\text{T}} R^{-1} H\right)^{-1}\left(\left(P^b\right)^{-1} x^{\text{f}}_{k+1} + H^{\text{T}} R^{-1} y_{\text{obs}}(t_{k+1})\right). \tag{2.1.4}$$

Using the *Sherman-Morrison-Woodbury* formula

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \tag{2.1.5}$$

we get

$$\begin{aligned}
x^{\text{a}}_{k+1} &= \left(P^b - P^b H^{\text{T}}\left(R + HP^bH^{\text{T}}\right)^{-1} HP^b\right)\left(P^b\right)^{-1} x^{\text{f}}_{k+1} && (2.1.6)\\
&\quad + \left(\left(P^b\right)^{-1} + H^{\text{T}} R^{-1} H\right)^{-1} H^{\text{T}} R^{-1} y_{\text{obs}}(t_{k+1}) && (2.1.7)\\
&= x^{\text{f}}_{k+1} - P^b H^{\text{T}}\left(R + HP^bH^{\text{T}}\right)^{-1} Hx^{\text{f}}_{k+1} && (2.1.8)\\
&\quad + \left(\left(P^b\right)^{-1} + H^{\text{T}} R^{-1} H\right)^{-1} H^{\text{T}} R^{-1} y_{\text{obs}}(t_{k+1}). && (2.1.9)
\end{aligned}$$

Since

$$\left(\left(P^b\right)^{-1} + H^{\text{T}} R^{-1} H\right)^{-1} H^{\text{T}} R^{-1} = P^b H^{\text{T}}\left(R + HP^bH^{\text{T}}\right)^{-1}, \tag{2.1.10}$$

we finally obtain

$$x^{\text{a}}_{k+1} = x^{\text{f}}_{k+1} - P^b H^{\text{T}}\left(R + HP^bH^{\text{T}}\right)^{-1}\left(Hx^{\text{f}}_{k+1} - y_{\text{obs}}(t_{k+1})\right). \tag{2.1.11}$$

An extension of this method is the four dimensional variational Data Assimilation (4D-Var). In this procedure, a fourth dimension, namely time, is additionally taken into account. There, we consider a whole time window $t_{k-s}, ..., t_k$, which we shift w.l.o.g. to $t_0, ..., t_s$, and the cost functional becomes

$$\begin{aligned}\mathcal{J}(x) &= \frac{1}{2}\left(x_0^{\mathrm{f}} - x\right)^{\mathrm{T}} \left(P^b\right)^{-1} \left(x_0^{\mathrm{f}} - x\right) && (2.1.12)\\ &+ \frac{1}{2}\sum_{i=0}^{s} \left(y_{\mathrm{obs}}(t_i) - H\mathcal{M}^i(x)\right)^{\mathrm{T}} R^{-1} \left(y_{\mathrm{obs}}(t_i) - H\mathcal{M}^i(x)\right). && (2.1.13)\end{aligned}$$

Here, $\mathcal{M}^i(x)$ stands for the i-th fold application of the model $\mathcal{M}$ on x.

2.2 Kalman filters

The Kalman filter was invented in 1958 by Swerling (Swerling (1958)). Two years later, the Hungarian Rudolf E. Kálmán, gave a more complete and general description (Kalman (1960), Kalman & Bucy (1961)) and published it in a more prestigious journal. That's why it was named after him. A complaint by Swerling to the AIAA Journal was ignored.

Let $x_0 \sim \mathcal{N}(\mu_0, P_0)$, an initial Gaussian distribution π_0 with mean $\mu_0 \in \mathbb{R}^d$ and covariance matrix $P_0 \in \mathbb{R}^{d\times d}$. Let the model and the measurements be linear with normally distributed errors,

$$\begin{aligned} x_{k+1} &= Ax_k + \eta_k, \qquad \eta_k \sim \mathcal{N}(0, Q) && (2.2.1)\\ y_k^{\mathrm{obs}} &= Hx_k + \varepsilon_k, \qquad \varepsilon_k \sim \mathcal{N}(0, R), && (2.2.2)\end{aligned}$$

with $A, Q \in \mathbb{R}^{d\times d}, H \in \mathbb{R}^{p\times d}$ and $R \in \mathbb{R}^{p\times p}$. Again, the first part of the sequential procedure is the propagation of the probability distribution in time to obtain the prior $\pi_{k+1}(x|y_{1:k}^{\mathrm{obs}})$ at time t_{k+1}. We have given the posterior $\pi_k(x|y_{1:k}^{\mathrm{obs}}) \sim \mathcal{N}(\bar{x}_k^{\mathrm{a}}, P_k^{\mathrm{a}})$ at time t_k, where the superscript 'a' denotes the analysis mean and covariance. We use the Chapman-Kolmogorov equation (1.2.2) with transition kernel

$$\pi(x|x') \propto \exp\left(-\frac{1}{2}(x - Ax')^{\mathrm{T}}Q^{-1}(x - Ax')\right) \tag{2.2.3}$$

to obtain

$$\pi_{k+1}(x|y_{1:k}^{\mathrm{obs}}) \propto \exp\left(-\frac{1}{2}(x - A\bar{x}_k^{\mathrm{a}})^{\mathrm{T}}(AP_k^{\mathrm{a}}A^{\mathrm{T}} + Q)^{-1}(x - A\bar{x}_k^{\mathrm{a}})\right). \tag{2.2.4}$$

The prior distribution at time t_{k+1} is therefore Gaussian with

$$\bar{x}^{f}_{k+1} = A\bar{x}^{\mathrm{a}}_{k} \tag{2.2.5}$$

$$P^{\mathrm{f}}_{k+1} = AP^{\mathrm{a}}_{k}A^{\mathrm{T}} + Q, \tag{2.2.6}$$

where the superscript 'f' denotes the forecast mean and covariance.

For the assimilation step we have to apply Bayes' formula (1.2.4) with the likelihood function

$$\pi(y^{\mathrm{obs}}_{k+1}|x) \propto \exp\left(-\frac{1}{2}(y^{\mathrm{obs}}_{k+1} - Hx)^{\mathrm{T}}R^{-1}(y^{\mathrm{obs}}_{k+1} - Hx)\right). \tag{2.2.7}$$

This leads to

$$\begin{aligned}\pi_{k+1}(x|y^{\mathrm{obs}}_{1:k+1}) &\propto \exp\Bigg(-\frac{1}{2}\Big[(y^{\mathrm{obs}}_{k+1} - Hx)^{\mathrm{T}}R^{-1}(y^{\mathrm{obs}}_{k+1} - Hx)+ \qquad (2.2.8)\\ &\qquad (x - \bar{x}^{\mathrm{f}}_{k+1})^{\mathrm{T}}(P^{\mathrm{f}}_{k+1})^{-1}(x - \bar{x}^{\mathrm{f}}_{k+1})\Big]\Bigg)\\ &\propto \exp\left(-\frac{1}{2}(x - \bar{x}^{\mathrm{a}}_{k+1})^{\mathrm{T}}(P^{\mathrm{a}}_{k+1})^{-1}(x - \bar{x}^{\mathrm{a}}_{k+1})\right). \qquad (2.2.9)\end{aligned}$$

When we expand the argument of the exponential in (2.2.8) and complete the square, we deduce

$$P^{\mathrm{a}}_{k+1} = \left((P^{\mathrm{f}}_{k+1})^{-1} + H^{\mathrm{T}}R^{-1}H\right)^{-1} \tag{2.2.10}$$

$$\bar{x}^{\mathrm{a}}_{k+1} = \bar{x}^{\mathrm{f}}_{k+1} - P^{\mathrm{a}}_{k+1}H^{\mathrm{T}}R^{-1}(H\bar{x}^{\mathrm{f}}_{k+1} - y^{\mathrm{obs}}_{k}). \tag{2.2.11}$$

For details see Reich & Cotter (2015).

We see, that $\bar{x}^{\mathrm{a}}_{k+1}$ agrees with the solution (2.1.11) of the minimising problem from the 3D-Var method, for $P^{b} = P^{\mathrm{f}}_{k+1}$. This is no coincidence. The exponent in (2.2.8) agrees with the negative cost functional in 3D-Var. Therefore the 3D-Var solution maximises the posterior probability and gives the so called MAP (maximum a-posteriori probability) estimate. However, that this estimate agrees with the mean is a special feature of Gaussian distributions.

The prior covariance matrix P^{f}_{k+1} is usually very large and therefore computationally expensive to invert. Using the *Sherman-Morrison-Woodbury* formula (2.1.5) we rewrite

$$P^{\mathrm{a}}_{k+1} = P^{\mathrm{f}}_{k+1} - P^{\mathrm{f}}_{k+1}H^{\mathrm{T}}(HP^{\mathrm{f}}_{k+1}H^{\mathrm{T}} + R)^{-1}HP^{\mathrm{f}}_{k+1}. \tag{2.2.12}$$

We define the so called *Kalman gain* matrix

$$K = P^{\mathrm{f}}_{k+1}H^{\mathrm{T}}(HP^{\mathrm{f}}_{k+1}H^{\mathrm{T}} + R)^{-1} \tag{2.2.13}$$

and finally obtain the Kalman formulas

$$\bar{x}^{\mathrm{a}}_{k+1} = \bar{x}^{\mathrm{f}}_{k+1} - K(H\bar{x}^{\mathrm{f}}_{k+1} - y^{\mathrm{obs}}_{k+1}) \tag{2.2.14}$$

$$P^{\mathrm{a}}_{k+1} = P^{\mathrm{f}}_{k+1} - KHP^{\mathrm{f}}_{k+1}. \tag{2.2.15}$$

The term $H\bar{x}^{\mathrm{f}}_{k+1} - y^{\mathrm{obs}}_{k+1}$ is called innovation and the Kalman gain matrix K determines the weight we give the innovation. If the observation error is large compared to the uncertainty in our forecast, K will be relatively small and we stay close to the forecast. If R is relatively small compared to P^{f}_{k+1}, K will be large and we therefore include more information about the measurements.

Since all involved distributions in this case are Gaussian for all time, it is sufficient to keep track of the mean and covariance matrix and therefore formulas (2.2.5), (2.2.6), (2.2.14) and (2.2.15) yield the whole algorithm.

If all the assumptions of having a linear model, a linear observation operator, a Gaussian initial distribution and a Gaussian measurement error are met, the Kalman filter indeed yields the exact solution to the filtering problem. However, these assumptions are very restrictive and rarely met in practical applications. Another issue is that storing and updating the covariance matrix becomes excessively computationally expensive for high dimensional systems. This can, in fact, be remedied via ensemble Kalman filters which we introduce in the following section.

2.2.1 Ensemble Kalman filters

Ensemble Kalman filters are essentially Monte Carlo approximations to the original Kalman filter. If the assumptions for the Kalman filter hold, the ensemble Kalman filter converges to the true solution of the filtering problem with increasing ensemble size and is the *best linear unbiased estimator* (BLUE). But even if the system is nonlinear and the distributions are not exactly Gaussian, ensemble Kalman filters still produce reasonable results. They are sufficiently robust.

The idea is that one starts with an ensemble of initial states $x_0^1, ..., x_0^M$ such that the empirical mean and covariance

$$\bar{x}_{0,M} = \sum_{i=1}^{M} x_0^i \tag{2.2.16}$$

$$P_{0,M} = \sum_{i=1}^{M} (x_0^i - \bar{x}_{0,M})(x_0^i - \bar{x}_{0,M})^{\mathrm{T}} \tag{2.2.17}$$

coincide with the mean and covariance of the initial distribution $\pi_0(x)$. Each ensemble member at time t_k is propagated until time t_{k+1} using the model. By that we obtain the forecast ensemble at time t_{k+1} and we can deduce the empirical mean and covariance of the prior distribution by

$$\bar{x}^{\mathrm{f}}_{k+1,M} = \sum_{i=1}^{M} x_{k+1}^i \tag{2.2.18}$$

$$P^{\mathrm{f}}_{k+1,M} = \sum_{i=1}^{M} (x_{k+1}^i - \bar{x}_{k+1,M})(x_{k+1}^i - \bar{x}_{k+1,M})^{\mathrm{T}}. \tag{2.2.19}$$

Now, one can apply the Kalman formulas (2.2.14) and (2.2.15) to obtain the approximations of the posterior mean and covariance $\bar{x}^{\mathrm{a}}_{k+1,M}$ and $P^{\mathrm{a}}_{k+1,M}$.

However, we need an analysis ensemble instead of just the mean and covariance for iterating the algorithm further. Therefore we have to create samples $x_{k+1}^1, ..., x_{k+1}^M$, such that their empirical mean and covariance coincide with $\bar{x}^{\mathrm{a}}_{k+1,M}$ and $P^{\mathrm{a}}_{k+1,M}$ respectively.

In what follows we drop the time index for simplicity. The simplest approach for creating a suitable analysis ensemble is the *Ensemble Kalman filter with perturbed observations* (EnKF) proposed in Burgers et al. (1998). The update of the ensemble members is carried out as

$$x_i^{\mathrm{a}} = x_i^{\mathrm{f}} - K(Hx_i^{\mathrm{f}} - y^{\mathrm{obs}} - \xi_i), \quad i = 1, ..., M \tag{2.2.20}$$

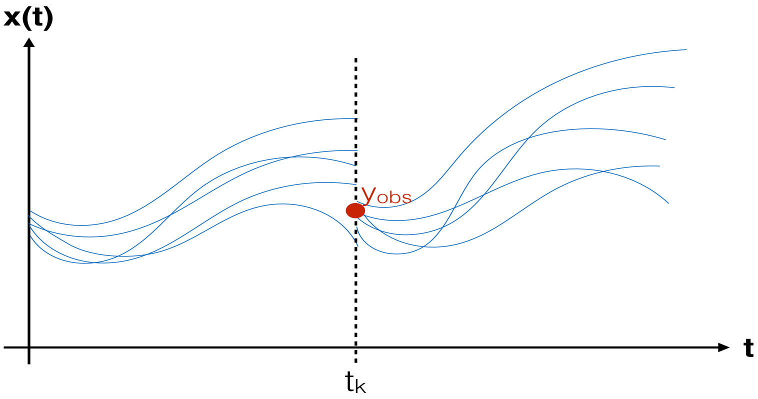


Figure 2.1: Systematic representation of an ensemble based filter

with the Kalman gain matrix

$$K = P_M^{\mathrm{f}} H^{\mathrm{T}} (H P_M^{\mathrm{f}} H^{\mathrm{T}} + R)^{-1}. \tag{2.2.21}$$

Here, $\xi_i \sim \mathcal{N}(0, R),\ i = 1, ..., M$ are independent identically distributed Gaussian random variables with the same covariance as the measurements. It is necessary to ensure that the empirical covariance matrix of the analysis ensemble coincides with the posterior covariance matrix that is obtained by applying the Kalman formula to the empirical covariance matrix of the forecast ensemble. In the first formulation of an ensemble Kalman filter in Evensen (1994), this perturbation of the observations was missing and had to be corrected. In Figure (2.1) one can see a systematic representation of an ensemble based filter.

This approach involves stochasticity and therefore introduces sampling issues. A simple deterministic way to create a suitable analysis ensemble would be

$$x_i^{\mathrm{a}} = \bar{x}_i^{\mathrm{a}} + (P_M^{\mathrm{a}})^{\frac{1}{2}} (P_M^{\mathrm{f}})^{-\frac{1}{2}} (x_i^{\mathrm{f}} - \bar{x}_M^{\mathrm{f}}). \tag{2.2.22}$$

Since P_M^{f} and P_M^{a} are usually very large matrices the computation of their square roots is computationally not affordable in practice.

A less expensive way to create an analysis ensemble deterministically is the *Ensemble square root filter* (ESRF), see for example Tippett et al. (2003). In order to derive that algorithm we write the ensemble covariance matrix in terms of the ensemble anomalies matrix

$$P_M^{\mathrm{f}} = \frac{1}{M-1} A^{\mathrm{f}} (A^{\mathrm{f}})^{\mathrm{T}} \tag{2.2.23}$$

with

$$A^{\mathrm{f}} := \begin{bmatrix} (x_1^{\mathrm{f}} - \bar{x}_M^{\mathrm{f}}) & (x_2^{\mathrm{f}} - \bar{x}_M^{\mathrm{f}}) & \cdots & (x_M^{\mathrm{f}} - \bar{x}_M^{\mathrm{f}}) \end{bmatrix} \in \mathbb{R}^{d \times M}. \tag{2.2.24}$$

The Kalman formula for the covariance matrix (2.2.15) then becomes

$$\begin{align} P_M^{\mathrm{a}} &= P_M^{\mathrm{f}} - KHP_M^{\mathrm{f}} \tag{2.2.25} \\ &= P_M^{\mathrm{f}} - \left(P_M^{\mathrm{f}} H^{\mathrm{T}} \left(HP_M^{\mathrm{f}} H + R \right)^{-1} \right) HP_M^{\mathrm{f}} \tag{2.2.26} \\ &= \frac{1}{M-1} A^{\mathrm{f}} (A^{\mathrm{f}})^{\mathrm{T}} - \frac{1}{M-1} A^{\mathrm{f}} (A^{\mathrm{f}})^{\mathrm{T}} H^{\mathrm{T}} \left(HP_M^{\mathrm{f}} H + R \right)^{-1} H \frac{1}{M-1} A^{\mathrm{f}} (A^{\mathrm{f}})^{\mathrm{T}} \tag{2.2.27} \\ &= \frac{1}{M-1} A^{\mathrm{f}} \left(I - \frac{1}{M-1} (HA^{\mathrm{f}})^{\mathrm{T}} \left(HP_M^{\mathrm{f}} H + R \right)^{-1} HA^{\mathrm{f}} \right) (A^{\mathrm{f}})^{\mathrm{T}} \tag{2.2.28} \\ &= \frac{1}{M-1} A^{\mathrm{a}} (A^{\mathrm{a}})^{\mathrm{T}}. \tag{2.2.29} \end{align}$$

Now we seek a matrix $S \in \mathbb{R}^{M \times M}$ that satisfies

$$SS^{\mathrm{T}} = I - \frac{1}{M-1} (HA^{\mathrm{f}})^{\mathrm{T}} \left(HP_M^{\mathrm{f}} H + R \right)^{-1} HA^{\mathrm{f}} \tag{2.2.30}$$

such that

$$A^{\mathrm{a}} = A^{\mathrm{f}} S. \tag{2.2.31}$$

We use the *Sherman-Morrison-Woodbury* formula (2.1.5) again and get

$$S = \left(I + \frac{1}{M-1} (HA^{\mathrm{f}})^{\mathrm{T}} R^{-1} HA^{\mathrm{f}} \right)^{-\frac{1}{2}}. \tag{2.2.32}$$

For taking the square root of a matrix one can use singular value decomposition and therefore the algorithm has a complexity of $\mathcal{O}(M^3)$. With that, the update for the mean (2.2.14) becomes

$$\bar{x}^{\mathrm{a}}_M = \bar{x}^{\mathrm{f}}_M - \frac{1}{M-1} A^{\mathrm{f}} S^2 (A^{\mathrm{f}})^{\mathrm{T}} H^{\mathrm{T}} R^{-1} (H\bar{x}^{\mathrm{f}}_M - y^{\mathrm{obs}}) \tag{2.2.33}$$

$$= \sum_{i=1}^{M} x^{\mathrm{f}}_i w_i \tag{2.2.34}$$

with weights w_i being the i-th entry of the column vector

$$w = \frac{1}{M}\mathbb{1} - \frac{1}{M-1} S^2 (A^{\mathrm{f}})^{\mathrm{T}} H^{\mathrm{T}} R^{-1} (H\bar{x}^{\mathrm{f}}_M - y^{\mathrm{obs}}). \tag{2.2.35}$$

The ensemble update is then defined as follows:

$$x^{\mathrm{a}}_j = \sum_{i=1}^{M} w_i x^{\mathrm{f}}_i + \sum_{i=1}^{M} (x^{\mathrm{f}}_i - \bar{x}^{\mathrm{f}}_M) s_{ij} \tag{2.2.36}$$

$$= \sum_{i=1}^{M} x^{\mathrm{f}}_i d^{\mathrm{KF}}_{ij} \tag{2.2.37}$$

where

$$d^{\mathrm{KF}}_{ij} = w_i + s_{ij} - \frac{1}{M}. \tag{2.2.38}$$

This is the form of a general *linear ensemble transform filter* (LETF). All the filters we consider in this thesis belong to that group. This depiction emphasizes that the posterior ensemble is a linear transformation of the prior ensemble.

2.2.2 Kalman-Bucy filter

In the classical Kalman filter, we assume that the observations become available at discrete points in time. However, if they are coming in continuously, we need a continuous formulation of the Kalman filter. We look at a linear model, which is not discretised and given as a Itô stochastic differential equation

$$\mathrm{d}x = Ax \; \mathrm{d}t + Q^{\frac{1}{2}} \mathrm{d}w(t). \tag{2.2.39}$$

The observations are given by

$$\mathrm{d}y^{\mathrm{obs}} = Hx\ \mathrm{d}t + R^{\frac{1}{2}}\mathrm{d}r(t). \tag{2.2.40}$$

A and Q are again $d \times d$ matrices, $H \in \mathbb{R}^{p\times d}, R \in \mathbb{R}^{p\times p}$ and $w(t)$ and $r(t)$ denote Brownian motion.

A continuous formulation of the classical Kalman filter is the *Kalman-Bucy filter*, see Jazwinski (1970). It describes the propagation of the mean and the covariance matrix by the two differential equations

$$\mathrm{d}\bar{x} = A\bar{x}\mathrm{d}t + PH^{\mathrm{T}}R^{-1}(\mathrm{d}y^{\mathrm{obs}} - H\bar{x}\mathrm{d}t) \tag{2.2.41}$$

$$\frac{\mathrm{d}P}{dt} = AP + PA^{\mathrm{T}} + Q - PH^{\mathrm{T}}R^{-1}HP. \tag{2.2.42}$$

If we want to only look at the assimilation part of the procedure, the equations simplify to

$$d\bar{x} = -PH^{\mathrm{T}}R^{-1}(H\bar{x}\mathrm{d}t - \mathrm{d}y_{\mathrm{obs}}) \tag{2.2.43}$$

$$\frac{P}{dt} = -PH^{\mathrm{T}}R^{-1}HP, \tag{2.2.44}$$

by just leaving out the evolution of the mean and covariance due to the model. An ensemble formulation of the Kalman-Bucy filter has been derived in Bergemann & Reich (2012) and is given by

$$dz_i = -\frac{1}{2}P_M H^{\mathrm{T}}R^{-1}(Hz_i\mathrm{d}t + H\bar{z}\mathrm{d}t - 2\mathrm{d}y_{\mathrm{obs}}). \tag{2.2.45}$$

We will see later on, that this formulation of the EnKF will be useful when dealing with the problem of imbalances.

2.3 Particle filters

Although ensemble Kalman filters provide relatively good approximations of the solution of the filtering problem, particle filters may be better suited when the assumptions for the Kalman filter are strongly violated, i.e. when the model or the observation operator become very non-linear or the initial distribution or the measurement error are highly non-Gaussian. These filters do not assume a certain family of distributions but approximate the filtering distribution by using the empirical measure determined by the ensemble members. It is a non-parametric approach, whereas the Kalman filter based filters are parametric approaches. However, the disadvantage of particle filters is that they require a significantly bigger ensemble size. In fact, the number of required particles increases exponentially with the dimension of the model. This is colloquially known as the *curse of dimensionality* and prohibits the use of particle filters for large scale Data Assimilation problems.

All the particle filters are based on *importance sampling.* The development of this sampling technique began in the 1950's with Kahn (1950) and Kahn (1956). Let X and X' be random variables with densities π_X and $\pi_{X'}$ respectively such that the measure μ_X is absolutely continuous with respect to the measure $\mu_{X'}$. Assume we want to approximate the expectation value of X but have samples $x'_1, ..., x'_M$ from X'. We use

$$\mathbb{E}[X] \approx \sum_{i=1}^{M} w_i x'_i \tag{2.3.1}$$

with importance weights

$$w_i \propto \frac{\pi_X(x'_i)}{\pi_{X'}(x'_i)}, \tag{2.3.2}$$

with a normalisation constant chosen such that $\sum_{i=1}^{M} w_i = 1$.

We can use that sampling method for numerically solving the filtering problem, since we have given samples of the prior distribution π^{f} but want to find the mean of the posterior π^{a}. At time t_1, the first point in time where observations become available, the pdf of the posterior is given as $\pi_1(x)\pi(y_1^{\mathrm{obs}}|x)$ and the weights therefore become

$$w_{i,1} \propto \frac{\pi_1(x_{i,1}^{\mathrm{f}})\pi(y_1^{\mathrm{obs}}|x_{i,1}^{\mathrm{f}})}{\pi_1(x_{i,1}^{\mathrm{f}})} = \pi(y_1^{\mathrm{obs}}|x_{i,1}^{\mathrm{f}}), \quad i = 1, ..., M. \tag{2.3.3}$$

Applying that sequentially and propagating the ensemble members according to the model equations, we get

$$w_{i,k+1} \propto w_{i,k}\pi(y_k^{\mathrm{obs}}|x_{i,k}^{\mathrm{f}}), \tag{2.3.4}$$

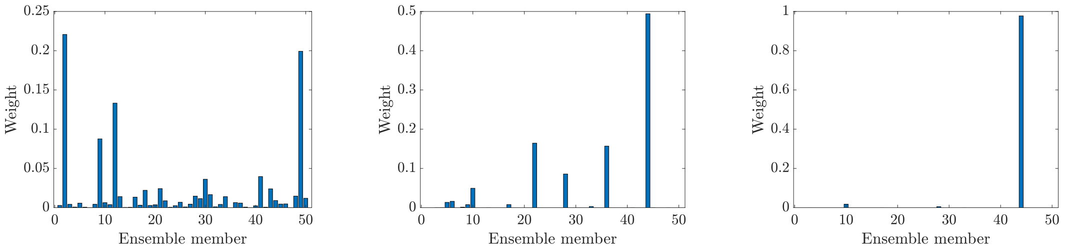
with $w_i^0 = \frac{1}{M}$. The posterior distribution is then estimated via the weighted empirical distribution

$$\pi_{k+1}(x|y^{\text{obs}}_{1:k+1}) = \sum_{i=1}^{M} w_i^{k+1}\delta(x - x^{\text{f}}_{i,k+1}), \tag{2.3.5}$$

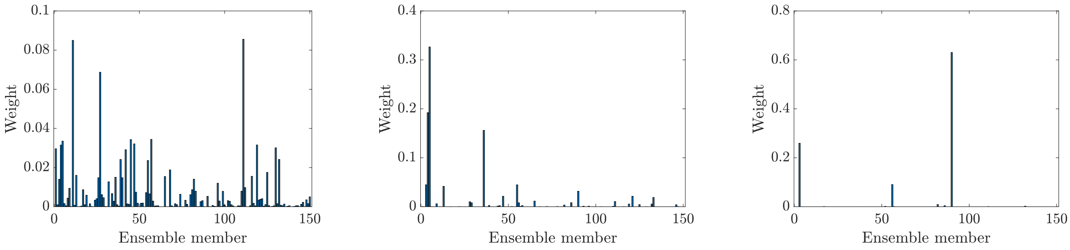
where $\delta(x - x^{\text{f}}_{i,k+1})$ denotes the Dirac measure centred at $x^{\text{f}}_{i,k+1}$. This procedure is called *sequential importance sampling* (SIS) first coined by in Del Moral (1996). For $M \to \infty$ formula (2.3.5) should converge to the true solution of the filtering problem but due to the finite ensemble size a phenomenon called filter degeneracy occurs (Doucet (1998), Bickel et al. (2008), Snyder et al. (2008)). That means that after a while the weights will become extremely non-uniform, i.e., one weight tends to 1 whereas the others converge to 0. This can be also described by the *effective ensemble size*:

$$M_{\text{eff}} := \frac{1}{\sum_{i=1}^{M} w_i^2}. \tag{2.3.6}$$

If the weights become highly non-uniform M_{eff} tends to 1, whereas for the optimal case of uniform weights, M_{eff} is equal to the actual ensemble size M.



(a) $M = 50$, step 1 (b) $M = 50$, step 4 (c) $M = 50$, step 7



(d) $M = 150$, step 1 (e) $M = 150$, step 4 (f) $M = 150$, step 7

Figure 2.2: The weights of the sequential importance sampling after 1,4 and 7 time steps for $M = 50$ and $M = 150$.

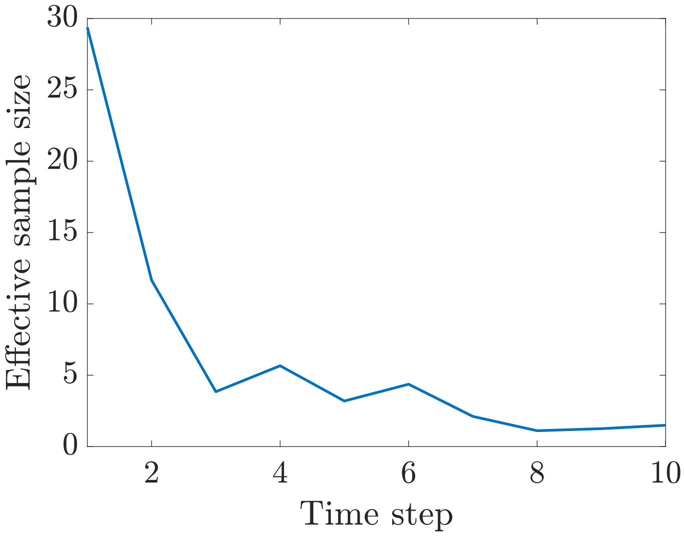


Figure 2.3: The time evolution of the effective ensemble size M_{eff}.

To illustrate that problem, we conducted an experiment and applied the SIS procedure to a model (the slow-fast Lorenz96 model, which will be described and used later in this thesis again), with the ensemble sizes $M = 50$ and $M = 150$. In Figure 2.2 the weights are depicted and it is noticeable that they become more non-uniform from step 1 to step 4 and finally step 7. In Figure 2.3 we can see the effective sample size (2.3.6) plotted against time steps for $M = 150$. It is already significantly smaller than the actual M in the first step and it rapidly converges towards 1.

2.3.1 SIR

A way to deal with this issue is including a resampling step in the algorithm if the effective sample size drops below some tolerance value. Often, $\frac{M}{2}$ is used as tolerance value. In that case a new ensemble of M equally weighted members is created from the forecast ensemble x_i^{f}. This is called *sequential importance resampling* (SIR), which is the standard particle filter and has been invented by Gordon et al. (1993).

One way to resample is by applying *multinomial resampling*. It works by drawing M times from a multinomial distribution $Mult(M; w_1, ..., w_M)$ with outcomes $1, ..., M$ and respective probabilities $w_1, ..., w_M$, which are our importance weights. N_i denotes the amount of times the index i was drawn, and our new ensemble will then consist of N_i copies of particle i for $i = 1, ..., M$. Algorithm 1 shows how to implement the calculation of the quantities N_i.

Algorithm 1 Multinomial Resampling

```
$N_i := 0 \quad \forall\, i = 1, ..., M$
for $l = 1, ..., M$ do
    draw $u$ from uniform distribution $\mathcal{U}[0,1]$
    solve $i^* = \arg\min_{i \geq 1} \sum_{j=1}^{i} w_j \geq u$
    $N_{i^*} = N_{i^*} + 1$
end for
```

However, the resampling step basically creates copies of ensemble members with high weights. In case the underlying model is deterministic, those particles will remain equal. Therefore the ensemble size still has to be very high. See for example Snyder et al. (2008) and Snyder et al. (2015).

In this filter, a posterior particle x_i^{a} is drawn from the prior ensemble $\{x_1^{\mathrm{f}}, ..., x_M^{\mathrm{f}}\}$ and thus can be written as

$$x_j^{\mathrm{a}} = \sum_{i=1}^{M} M t_{ij} x_i^{\mathrm{f}}, \tag{2.3.7}$$

with $\sum_{i=1}^{M} t_{ij} = \frac{1}{M}$ and $\mathbb{E}[\sum_{j=1}^{M} t_{ij}] = w_i$.

2.3.2 ETPF

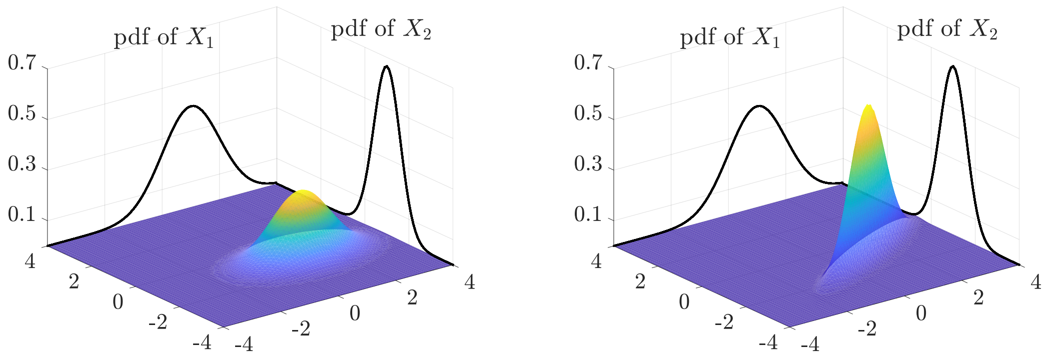
Another approach is the *ensemble transform particle filter* (ETPF), which was developed by Reich (2013). We first define a coupling of two probability measures μ_{X_1} and μ_{X_2} as the tuple $Z = (X_1, X_2)$ with the marginal distributions $X_1 \sim \mu_{X_1}$, $X_2 \sim \mu_{X_2}$ and the joint distribution $Z \sim \mu_Z$. The set of all possible joint measures is denoted by $\Pi(\mu_{X_1}, \mu_{X_2})$.

As an example we consider two univariate Gaussians $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$.

The possible joint distributions are of the form

$$Z \sim \mathcal{N}\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right), \tag{2.3.8}$$

where $\rho \in [-1,1]$ is the correlation coefficient. In Figure 2.4 we can see two different couplings of X_1 and X_2 for $\mu_1 = 1, \mu_2 = -1, \sigma_1 = 1, \sigma_2 = 0.6$ for choices of $\rho = 0.1$ and $\rho = 0.9$.



(a) $\rho = 0.1$ (b) $\rho = 0.9$

Figure 2.4: Two different couplings between two Gaussian random variables $X_1 \sim \mathcal{N}(1,1),\ X_2 \sim \mathcal{N}(-1,0.6)$.

As $\rho \to 1$, the variance $\sigma_c^2 = (1-\rho^2)\sigma_2^2$ of the conditional distribution $\pi(x_2|x_1)$ converges to zero and therefore, the conditional distribution converges to a Dirac delta distribution. This kind of coupling is called a deterministic coupling and it possesses a transport map T, such that $X_2 = T(X_1)$ and $\mu_Z(dx_1, dx_2) = \delta(x_2 - T(x_1))\mu_{X_1}(dx_1)dx_2$. The pdf of the conditional distribution can be written formally as $\pi(x_2|x_1) = \delta(x_2 - T(x_1))$. Such a coupling always produces a non-zero covariance $\mathrm{cov}(X_1, X_2)$. In certain situations, for example when looking for a coupling between the prior and posterior distribution, it can be necessary to find a coupling that maximises the covariance between two measures. It can be shown, that this simultaneously minimises the so called L^2-Wasserstein distance

$$W(\mu_{X_1}, \mu_{X_2}) = \sqrt{\mathbb{E}[(\|X_1 - X_2\|)^2]}, \tag{2.3.9}$$

which is known as the Monge-Kantorovitch problem.

The ETPF is based on this optimisation problem. Having the prior ensemble and the importance weights at hand, instead of the classical resampling step of the SIR, we solve the following optimal transport problem:

$$T^* = \underset{T\in\mathbb{R}^{M\times M},\ t_{ij}=(T)_{ij}\geq 0}{\arg\min} \left\{ \sum_{i,j=1}^{M} t_{ij} \|x_i^{\mathrm{f}} - x_j^{\mathrm{f}}\|^2 \right\} \tag{2.3.10}$$

with respect to

$$\sum_{i=1}^{M} t_{ij} = \frac{1}{M}, \quad \sum_{j=1}^{M} t_{ij} = w_i. \tag{2.3.11}$$

The solution of that problem is a stochastic matrix which presents the optimal coupling of the prior and posterior distribution in the sense that it minimises the Euclidean distance

$$\mathbb{E}[\|X_M^{\mathrm{f}} - X_M^{\mathrm{a}}\|^2]. \tag{2.3.12}$$

The two marginals are the prior and the posterior distribution. The analysis ensemble is given by

$$x_j^{\mathrm{a}} = \sum_{i=1}^{M} M t_{ij}^* x_i^{\mathrm{f}} \tag{2.3.13}$$

$$=: \sum_{i=1}^{M} d_{ij}^{\mathrm{PF}} x_i^{\mathrm{f}} \tag{2.3.14}$$

When we compare this to (2.3.7), we see that this procedure is a special case of resampling, in fact it is a resampling that maximises the correlation between the prior and posterior by creating an equally weighted posterior ensemble whose members differ as little as possible from the prior ensemble members. It can be shown that the ETPF is consistent in the sense that $\frac{1}{M}\sum_{j=1}^{M} f(x_j^{\mathrm{a}}) \to \mathbb{E}\left[f(X^{\mathrm{a}})\right]$ for $M \to \infty$, see Reich (2013).

However, example 8.11 in Reich & Cotter (2015) shows that even in the ETPF, the covariance matrix is permanently underestimated and rejuvenation (explained in the next section) is necessary. In Acevedo et al. (2017), a second-order accurate version is presented. This means both, the first and second moment are estimated correctly by the proposed algorithm.

Solving problem (2.3.10) with common algorithms makes the ETPF more expensive than the ESRF since it would be of complexity $M^3 \log(M)$. The *FastEMD* algorithm by Pele & Werman (2009) reduced this complexity. However, for a one-dimensional optimal transport problem, the solution can be computed with linear complexity and therefore we consider an approximation by updating the components of x_i^{f} sequentially and then for each $k = 1, ..., d$ solve

$$\sum_{i,j=1}^{M} t_{ij} |x_{i,k}^{\mathrm{f}} - x_{j,k}^{\mathrm{f}}|^2. \tag{2.3.15}$$

2.4 Localisation and Inflation

In real world applications of Data Assimilation techniques, the dimension of the state space d is usually much larger than the affordable ensemble size M. The empirical covariance matrix only has rank $M - 1$ and is consequently not representing the real covariance properly. Therefore, all ensemble based filters have the major drawback of underestimating the error covariance matrix. This was first mentioned in Houtekamer & Mitchell (1998). The systematic underestimation leads to overconfidence in the forecast and therefore the observations become less influential which can ultimately cause filter divergence. This issue can be remedied with the so called multiplicative *ensemble inflation*, which was first introduced in Anderson & Anderson (1999). The idea is to artificially increase the ensemble spread by redefining the update for the samples via

$$\hat{x}_i^{\mathrm{f}} := \bar{x}_M^{\mathrm{f}} + \alpha(x_i^{\mathrm{f}} - \bar{x}_M^{\mathrm{f}}) \tag{2.4.1}$$

with $\alpha > 1$, before the assimilation step. Another approach is additive inflation, also called *particle rejuvenation* as in Pham (2001). In this method, one just adds random values

$$\xi_i \sim \mathcal{N}(0, \tau^2 P^{\mathrm{a}}), \ i = 1, ..., M \tag{2.4.2}$$

to the posterior ensemble members after the assimilation step, with P^{a} being the ensemble covariance matrix and $\tau > 0$ a parameter, controlling how much the particles are changed. One mostly uses multiplicative inflation for ensemble Kalman filters and rejuvenation for particle filters.

In particular, spatio-temporal systems exhibit problems due to the finite sample size effects of filtering algorithms. Let the state space be $\mathbb{R}^d$ and we introduce the notation of $z_i(x_k)$ as the value of the i-th ensemble member, $i = 1, ..., M$, at grid point x_k, $k = 1, ..., d$. Spurious correlations between grid points that are physically distant from each other can occur in the empirical covariance matrix. E.g. the covariance matrix might suggest that there is a high correlation between the temperature in Tokyo and Berlin, which is not true in reality. A technique called *localisation* can help to overcome this issue. There are two approaches, *B-localisation* and *R-localisation*. The first one was invented in Houtekamer & Mitchell (1998) and replaces the ensemble covariance matrix by

$$\hat{P}^{\mathrm{f}}_M = C \circ P^{\mathrm{f}}_M. \tag{2.4.3}$$

Here, $\circ$ denotes the Schur product and C is a suitable matrix. One possibility is to choose

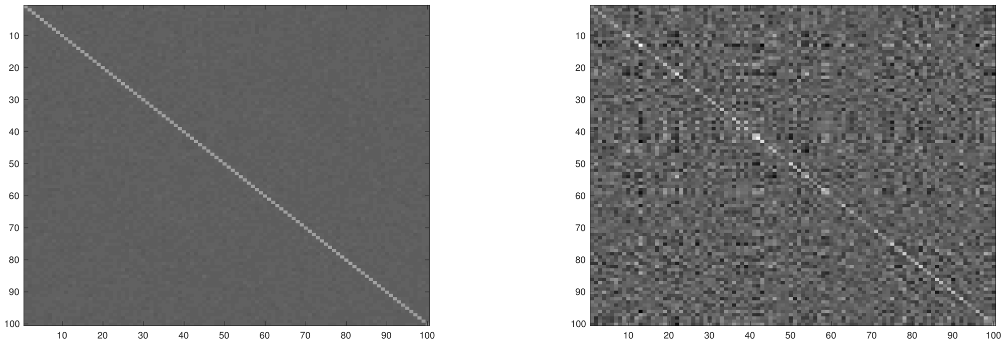
$$C_{st} = \rho\left(\frac{\|x_s - x_t\|}{r_{\mathrm{loc}}}\right), \tag{2.4.4}$$

with localisation radius r_{loc}, $\|x_s - x_t\|$ being the distance between grid points x_s and x_t in physical space and ρ denoting the Gaspari-Cohn-function (Gaspari & Cohn (1999))

$$\rho(z) = \begin{cases} -\frac{1}{4}z^5 + \frac{1}{2}z^4 + \frac{5}{8}z^3 - \frac{5}{3}z^2 + 1, & z < 1, \\ \frac{1}{12}z^5 - \frac{1}{2}z^4 + \frac{5}{8}z^3 + \frac{5}{3}z^2 - 5z + 4 - \frac{2}{3z}, & 1 \leq z \leq 2, \\ 0, & z > 2. \end{cases} \tag{2.4.5}$$

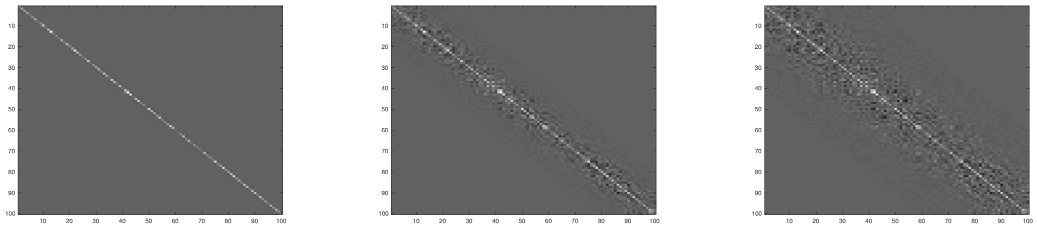
Example 2.4.1. We draw samples from the distribution $\mathcal{N}(0_{100}, I_{100})$, i.e. the samples are in $\mathbb{R}^{100}$ and are normally distributed with mean 0 and the covariance matrix being the identity. In Figure (2.5) we can see drastic spurious correlations in the ensemble covariance matrix, using 10 samples compared to using 1000 samples. We apply B-localisation with localisation radii 1, 10 and 20 and in Figure (2.6), we can see that it significantly reduces the spurious correlations.

This straight forward approach cannot be applied when the ensemble covariance matrix is not explicitly stored and updated, as in the case of the ESRF. Therefore, Hunt et al came up with the idea of R-localisation (Hunt et al. (2007)), inventing the *local ensemble transform Kalman filter* (LETKF). We write the update step of



(a) P_{1000} (b) P_{10}

Figure 2.5: Estimation of the covariance matrix with $M = 1000$ ensemble members (left) and $M = 10$ ensemble members (right).



(a) $C_1 \circ P_{10}$ (b) $C_{10} \circ P_{10}$ (c) $C_{20} \circ P_{10}$

Figure 2.6: Localised empirical covariance matrix with localisation radius 1 (left), 10 (middle) and 20 (right).

a LETF separately for each grid point x_k as

$$z_j^{\mathrm{a}}(x_k) = \sum_{i=1}^{M} z_i^{\mathrm{f}}(x_k) d_{ij}(x_k). \tag{2.4.6}$$

With B-localisation, the components of the prior covariance matrix, that represent long distance correlations between the grid points are artificially reduced. In contrast, with R-localisation, the components of the observation error covariance matrix R_k that is used for the update of $z_j(x_k)$ are altered. This is done in such a way, that the observation error is artificially increased for those components, that are spatially far away from the grid point x_k. For each $k = 1, ..., d$, we define a diagonal localisation matrix $C_k \in \mathbb{R}^{p \times p}$ by

$$(C_k)_{ll} = \rho\left(\frac{\|x_l - x_k\|}{r_{\mathrm{loc}}}\right), \tag{2.4.7}$$

with $l = 1, ..., p$ being the indices of the grid points on which we have observations. With that we define

$$\tilde{R}_k^{-1} = C_k \circ R^{-1}. \tag{2.4.8}$$

When applying the ESRF separately for every $k = 1, ..., d$, we replace R by $\tilde{R}_k$ in formulas (2.2.30) and (2.2.35), yielding S_k and w_k respectively.

Therefore, when updating the value of z_i at grid point x_k the observation error for observations that are outside twice the localisation radius are formally set to infinity. This leads to not taking those observations into account.

Other works about localisation in Kalman based filters include Ott et al. (2004) and Szunyogh et al. (2005), where the ESRF is also applied in every gridpoint separately but in contrast to the LETKF, the matrices S, A and P_M^{f} in formula (2.2.30) are built using only grid points in the vicinity of the currently updated grid point. This also makes the routine suitable for parallelisation.

van Leeuwen (2017) explains that the motivation for localisation in particle filters is different from Kalman based filters, since these methods do not rely on the estimation of the covariance. However, one wants to reduce the number of observations at each grid point to avoid filter degeneracy. One cannot apply the concept of R- or B-localisation to a standard SIR since the resampling step would destroy any spatial regularity. However it is possible with the ETPF, which has been done by Cheng & Reich (2015).

Publications dealing with the theoretical need for localisation of particle filters include Bengtsson et al. (2003), van Leeuwen (2003) and Rebeschini & van Handel (2015), while practical methods (besides the aforementioned ETPF) are presented in Poterjoy (2016) and Peny & Miyoshi (2016).

Chapter 3

Hybrid filter

3.1 Likelihood Splitting

As aforementioned, ensemble Kalman filters are not consistent in the non-Gaussian and non-linear case and particle filters suffer from the curse of dimensionality. To exploit the advantages of both classes, there have been several studies about combining the two families of filters. Those include modified particle filters in which the EnKF posterior is used as a proposal distribution instead of the prior (Papadakis et al. (2010), Mandel & Beezley (2009)), several approaches of approximating the forecast distribution as a sum of Gaussians (Bengtsson et al. (2003), Sun et al. (2009), Dovera & Della Rossa (2011)), the EnKPF by Frei & Künsch (2013), which bridges between the EnKF and the SIR and a couple of hybrid variants suggested by Nerger (2018) that bridge between the ETKF by Bishop et al. (2001) (another variant of an ensemble Kalman filter) and the NETF by Tödter & Ahrens (2015) (another variant of a particle filter).

We introduced a hybrid filter that bridges between the ESRF and the ETPF in Chutsagulprom et al. (2016). In contrast to the EnKPF by Frei and Künsch from 2013, in this hybrid filter, the concept of localisation can be applied. However, Robert & Künsch (2017) provide a way to apply localisation to the EnKPF.

The basic idea is to split the likelihood as follows given $\alpha \in [0,1]$:

$$\pi(y^{\text{obs}}|x) \propto \exp\left(-\frac{1-\alpha}{2}(Hx - y^{\text{obs}})^{\text{T}} R^{-1}(Hx - y^{\text{obs}})\right) \times \\ \exp\left(-\frac{\alpha}{2}(Hx - y^{\text{obs}})^{\text{T}} R^{-1}(Hx - y^{\text{obs}})\right). \tag{3.1.1}$$

We have seen that both the ESRF and the ETPF belong to the class of linear ensemble transform filters and therefore yield coefficients d_{ij}^{KF} and d_{ij}^{PF} determined by (2.2.38) and (2.3.14) respectively.
For the prior ensemble $x_1^{\text{f}}, ..., x_M^{\text{f}}$, we can now treat both factors of the likelihood (3.1.1) separately. We apply an ESRF to the first factor and obtain

$$d_{ij}^{\text{KF}}(\alpha) = w_i^{\text{KF}}(\alpha) - \frac{1}{M} + s_{ij}(\alpha) \tag{3.1.2}$$

with $w_i^{\text{KF}}(\alpha)$ being the i-th entry of

$$w^{\text{KF}}(\alpha) := \frac{1}{M}1 - \frac{1-\alpha}{M-1} S(\alpha)^2 (HA^{\text{f}})^{\text{T}} R^{-1}(H\bar{x}_M^{\text{f}} - y^{\text{obs}}) \tag{3.1.3}$$

and $(s_{ij}(\alpha)) = \{S(\alpha)\}_{ij}$ with

$$S(\alpha) := \left(I + \frac{1-\alpha}{M-1}(HA^{\text{f}})^{\text{T}} R^{-1} HA^{\text{f}}\right)^{-\frac{1}{2}}. \tag{3.1.4}$$

Then, we apply an ETPF to the second one by calculating the importance weights

$$w_i^{\text{PF}}(\alpha) \propto \exp\left(-\frac{\alpha}{2}(Hx_i^{\text{f}} - y^{\text{obs}})^{\text{T}} R^{-1}(Hx_i^{\text{f}} - y^{\text{obs}})\right) \tag{3.1.5}$$

and solve the optimal transport problem (2.3.10) obtaining $t_{ij}^*(\alpha)$ and therefore ultimately $d_{ij}^{\text{PF}}(\alpha)$. This leads to two different implementations of the hybrid filter:

(A) ETPF-ESRF

$$x_j^{\text{h}} = \sum_{i=1}^{M} d_{ij}^{\text{PF}}(\alpha) x_i^{\text{f}} \tag{3.1.6}$$

$$x_j^{\text{a}} = \sum_{i=1}^{M} d_{ij}^{\text{KF}}(\alpha) x_i^{\text{h}} \tag{3.1.7}$$

(B) ESRF-ETPF

$$x_j^{\mathrm{h}} = \sum_{i=1}^{M} d_{ij}^{\mathrm{KF}}(\alpha) x_i^{\mathrm{f}} \tag{3.1.8}$$

$$x_j^{\mathrm{a}} = \sum_{i=1}^{M} d_{ij}^{\mathrm{PF}}(\alpha) x_i^{\mathrm{h}}. \tag{3.1.9}$$

The bridging parameter α can be chosen as a fixed value in $[0, 1]$ with $\alpha = 0$ corresponding to a pure ESRF and $\alpha = 1$ to a pure ETPF. It can also be chosen adaptively at every time step. For that, we recall the effective sample size of an ETPF,

$$M_{\text{eff}}(\alpha) = \frac{1}{\sum_{i=1}^{M} w_i^{\mathrm{PF}}(\alpha)^2} \tag{3.1.10}$$

and consider the ratio

$$\frac{M_{\text{eff}}(\alpha)}{M} =: \theta. \tag{3.1.11}$$

We now choose α such that θ is equal to a fixed reference value in $\left(\frac{1}{M}, 1\right]$ and note that $\theta \to 1$ implies the weights converging towards uniformity, corresponding to the ESRF ($\alpha = 0$), whereas $\theta \to \frac{1}{M}$ means, that the hybrid filter converges towards the ETPF ($\alpha = 1$).
However, this approach still leaves us with a tuning parameter. Another idea uses the rule of thumb, that the spread of a good ensemble should be close to the RMSE

$$\sqrt{\frac{1}{d}\sum_{k=1}^{d}(\bar{x}_k - x_k^{\mathrm{ref}})^2}, \tag{3.1.12}$$

with d being the dimension of the state space (Stephenson & Doblas-Reyes (2000)). A measure of the spread is $\text{spread}(t) := \sqrt{\frac{1}{d}\mathrm{Tr}(\mathrm{P_M(t)})}$ where $P_M(t)$ is the empirical covariance matrix at time t. Since the reference solution is obviously not available, we have to use the RMSE in the observation space

$$\mathrm{RMSE}_{obs} = \sqrt{\frac{1}{p}\sum_{k=1}^{p}\left((H\bar{x})_k - y_k^{\mathrm{obs}}\right)^2}, \tag{3.1.13}$$

with p being the dimension of the observation space. Then we can choose α at each time point t_n where we apply the filter, such that the difference $|\text{spread}(t_n) - \text{RMSE}_{obs}(t_n)|$ is minimal. This can simply be done, by calculating the ensemble mean $\bar{x}$ for each $\alpha \in \{0, 0.1, 0.2, ..., 1\}$ and with these ten values at hand, one can choose the α, that produced the smallest RMSE_{obs}. Admittedly, this means, calculating the analysis mean ten times, but since only the mean and not the whole ensemble is needed to determine the RMSE, one does not have to solve the optimal transport problem, which is part of the ETPF, when computing the optimal α.

3.1.1 Examples

Let us consider the well known Lorenz-63 model (Lorenz (1963)):

$$\dot{x} = \sigma(y - x) \tag{3.1.14}$$

$$\dot{y} = x(\rho - z) - y \tag{3.1.15}$$

$$\dot{z} = xy - \beta z. \tag{3.1.16}$$

The equations are derived from Saltzman's model for atmospheric convection and describe a 2-dimensional fluid layer being uniformly heated from below and cooled down from above. Hereby, x describes the intensity of convective motion, y the temperature difference between ascending and descending currents and z is proportional to the deviation from linearity of the vertical temperature profile. The parameters σ, ρ, β represent physical parameters as for example the Prandtl number and the Rayleigh number. For the standard parameter setting of $\sigma = 10, \rho = 28$ and $\beta = \frac{8}{3}$ the system is chaotic with the famous butterfly shaped attractor, which can be seen in Figure 3.1.

We apply the hybrid filter to it, observing only the x-component of the state vector with an observation interval of $\Delta t_{\text{obs}} = 0.12$ and an observation error variance of $R = 8$. We use a rejuvenation of 0.2, an initial variance of the ensemble of 1. We conduct the experiment for 10^5 assimilation cycles for different ensemble sizes. The results in terms of the time averaged RMSE (see 3.1.12), with

$$\text{RMSE}(t) = \sqrt{\frac{1}{3}\left\| \begin{pmatrix} \bar{x}(t) \\ \bar{y}(t) \\ \bar{z}(t) \end{pmatrix} - \begin{pmatrix} x^{\text{ref}}(t) \\ y^{\text{ref}}(t) \\ z^{\text{ref}}(t) \end{pmatrix} \right\|^2} \tag{3.1.17}$$

plotted against the bridging parameters α and θ are presented in Figure 3.2. We can

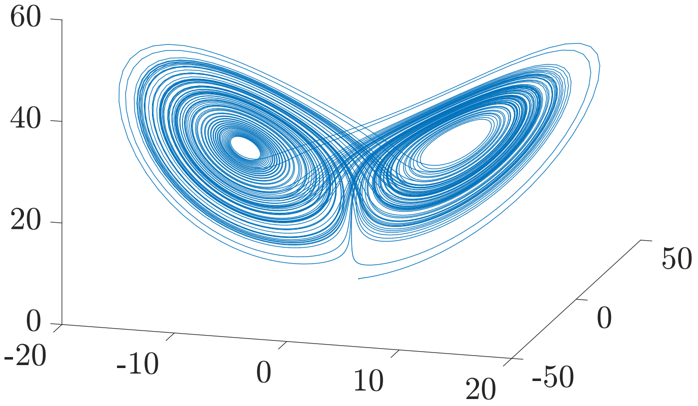
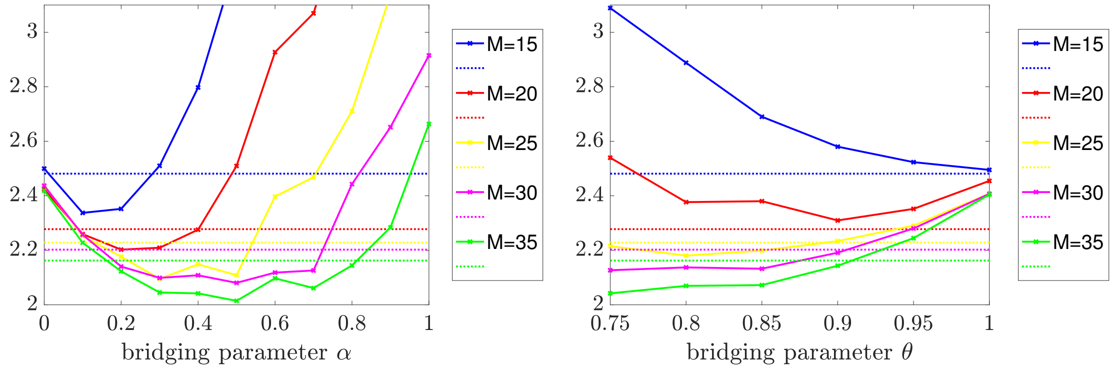


Figure 3.1: Attractor of the Lorenz-63 model with parameters $\sigma = 10,\ \rho = 28,\ \beta = \frac{8}{3}$.



(a) fixed α

(b) adaptive α

Figure 3.2: RMSE values for the hybrid filter (ETPF-ESRF) applied to the Lorenz-63 model plotted against the bridging parameter α on the left and against θ on the right. Both plots also contain the result of the error-vs-spread criterion to choose α as dotted lines for every ensemble size.

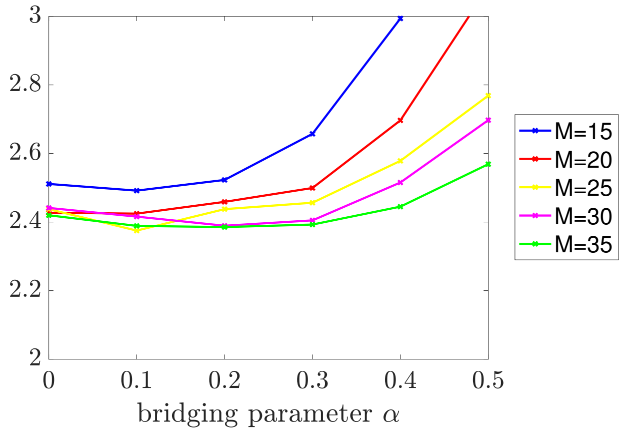


Figure 3.3: RMSE values for the hybrid filter (ESRF-ETPF) applied to the Lorenz-63 model plotted against the bridging parameter α.

clearly see that even for a small ensemble size of $M = 15$, the bridging parameter α being larger than zero (which would correspond to the pure ESRF) is beneficial. The larger the ensemble size, the larger is the optimal value of α. For the adaptive case, $\theta = 1$ corresponds to the pure ESRF and for all ensemble sizes larger than $M = 15$ we can see that a choice of $\theta < 1$ yields a smaller error. Overall, we obtain worse results with the adaptive approaches than for fixed α and it seems like the error-vs-spread criterion works slightly better than using the effective sample size for smaller ensemble sizes but worse for larger ones. In fact, the optimal results achieved by the adaptive approach which is based on the effective ensemble size become comparable to the optimal results of the hybrid using a fixed bridging parameter for larger ensemble sizes.

In Figure 3.3, we show the results for using version (B), the ESRF-ETPF implementation for fixed α. This approach does not work that well, presumably because the prior distributions are more non-Gaussian than the posterior distributions. Therefore it is beneficial to treat the prior ensemble at first with an ETPF and only apply the ESRF afterwards.

As we have seen, R-localisation is applicable to both, the ESRF and the ETPF. We can therefore use a localised hybrid filter for spatially extended systems. Let us consider the Lorenz-96 model (Lorenz (1996)):

$$\dot{z}_l = (z_{l+1} - z_{l-2})z_{l-1} - z_l + F, \quad l = 1, ..., 40, \tag{3.1.18}$$

which is another chaotic toy model for dynamics in the atmosphere. Here, the z_l can be thought of as values of some atmospheric quantity on a latitude circle. As in equations describing processes in the atmosphere, this model contains external forcing (the constant F), internal dissipation (described by the linear term) and some kind of advection (simulated by the quadratic terms). Another important property of this model is, that the total energy

$$E = \frac{1}{2}\sum_{l=1}^{40} z_l^2 \tag{3.1.19}$$

is preserved under the advective part $(z_{l+1} - z_{l-2})z_{l-1}$ of the model evolution. This model becomes chaotic for somewhat large values of F and we will use $F = 8$ for all experiments with this model throughout this thesis.
We apply the ETPF-ESRF hybrid filter to that model with observing every other grid point in time intervals of $\Delta t_{\text{obs}} = 0.11$ and an error variance of $R = 8$. We use a rejuvenation parameter (see formula (2.4.2)) of $\tau = 0.2$ and an initial variance of 1 again and we choose the localisation radius to be 4. We perform 50,000 assimilation cycles and show the results for different ensemble sizes from $M = 20$ up to $M = 40$ in Figure 3.4. Figure 3.5 shows the same results for only $M = 20, 25$ and $M = 30$ on a smaller window of the bridging parameter and compares it to the results of the adaptive hybrid. As in the experiment with the Lorenz-63 model, we plot the results of the error-vs-spread criterion for the adaptive choice of α as dotted lines.

Again, it is clearly visible that a value of α greater than 0 but smaller than 1 is better than choosing $\alpha = 0$ or $\alpha = 1$. The optimal α increases with the ensemble size again, which is to be expected. Again, we obtain poorer results with the adaptive approach than with the fixed bridging parameter. The error-vs-spread criterion yields again comparable results to the choice of α via the effective sample size, which are again worse than the results for fixed α.

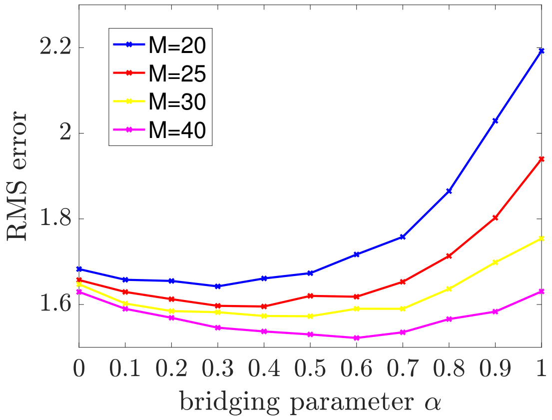
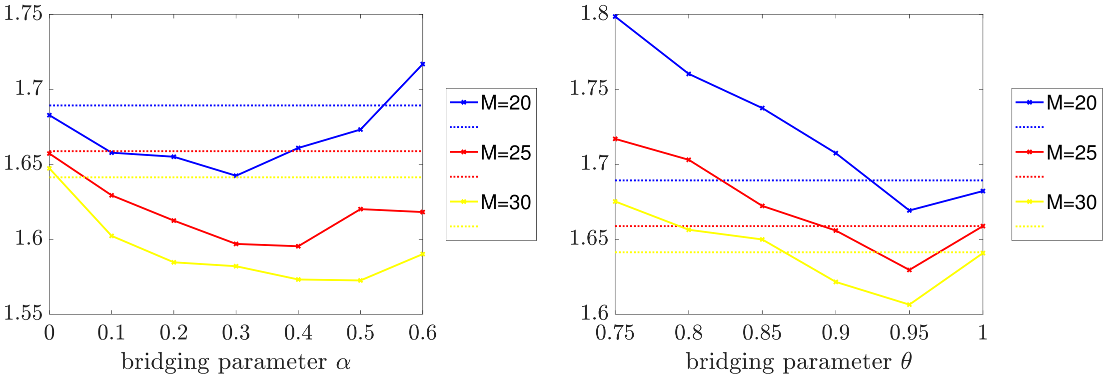


Figure 3.4: RMSE values for the hybrid filter (ETPF-ESRF) applied to the Lorenz-96 model plotted against the bridging parameter $\alpha \in [0, 1]$.



(a) fixed α

(b) adaptive α

Figure 3.5: RMSE values for the hybrid filter (ETPF-ESRF) applied to the Lorenz-96 model plotted against the bridging parameter $\alpha \in [0, 0.6]$ on the left and against $\theta \in [0.75, 1]$ on the right.

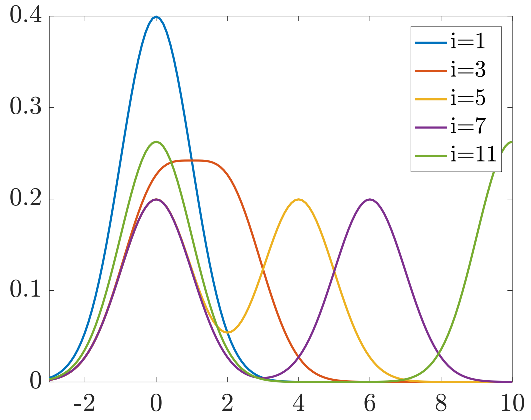


Figure 3.6: Pdfs of bimodal distributions with increasing distance between the two modes.

3.1.2 Adaptive choice of α via KL-divergence

In this section we present a different idea on how to choose the bridging parameter α adaptively for a very specific case. Therefore we created M samples $X_i^1, ..., X_i^M$ for $i = 1, ..., 11$ of eleven different distributions. The first distribution P_1 is Gaussian with mean $\mu_1 = 0$ and variance $\sigma^2 = 1$. The other distributions $P_i,\ i = 2, ..., 11$ are consisting of two Gaussians. One being P_1 and the second one being a Gaussian with mean $\mu_i = i - 1$ and variance $\sigma^2 = 1$ for $i = 2, ..., 11$. Therefore the two humps of the two Gaussians are increasingly far apart with increasing i. Figure 3.6 shows the pdfs of some of the distributions.

The Kullback-Leibler-Divergence (KL-Divergence) (Kullback & Leibler (1951))

$$\mathrm{KL}(p||q) = \int_{-\infty}^{\infty} p(x)\log\frac{p(x)}{q(x)}dx \tag{3.1.20}$$

measures the 'inefficiency' that occurs when a true underlying distribution with pdf p is approximated by another distribution with pdf q. It can, therefore, be seen as a kind of distance measure between two distributions. This is only a heuristic, since it is not symmetric and it does not satisfy the triangle inequality. We now estimate the KL-Divergence of the distribution represented by our sample, that is the distribution with pdf p, and a theoretical Gaussian distribution with mean $\mu = \bar{X}_i$ and variance $\sigma^2 = \frac{1}{M-1}\sum_{k=1}^{M}(X_i^k - \bar{X}_i)^2$, being the distribution with pdf q. I.e. we are taking the

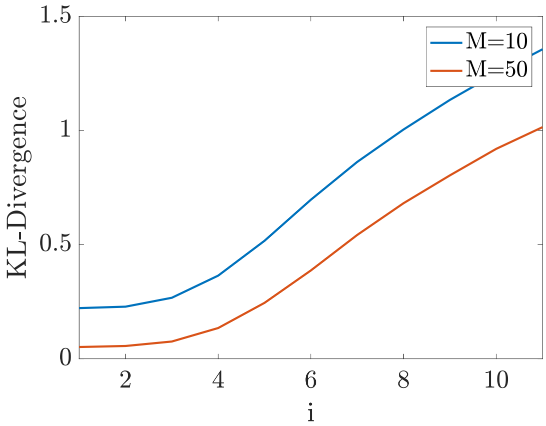


Figure 3.7: Estimated KL-Divergence between a bi-modal distribution (with increasing distance between the two modes) and a Gaussian.

empirical mean and variance as the mean and variance for the Gaussian we want to compare the ensemble to. To get an estimate for (3.1.20) we need to approximate $p(x)$ ($q(x)$ is given, since it is the pdf of a Gaussian). In order to do so, for each $i = 1, \dots, 11$ we first create the empirical distribution function from the samples $X_i^1, \dots, X_i^M$

$$\hat{F}_M(t) = \frac{1}{M} \sum_{j=1}^{M} \mathbb{1}_{X_i^j < t} \tag{3.1.21}$$

and afterwards we approximate its derivative, by first smoothing it via convolution with a mollifier $\phi(x) = \frac{1}{\varepsilon} c\exp\left(-\frac{1}{1-|x|^2}\right)$, with c being a normalisation constant and ε a tuning parameter, influencing how intense the smoothing is going to be. Afterwards, we apply the difference quotient. The KL-Divergence is estimated in that way for every distribution P_i 1000 times (every time drawing a new ensemble of size M from P_i) and then the average KL-Divergence over those 1000 repeats is taken. In Figure 3.7 you can see the results for $M = 10$ and 50.

Those results are exactly what we would expect, the larger the distance of the two Gaussians forming P_i, the larger the KL-Divergence becomes. Clearly, for $i = 1$, meaning $\mu_1 = 0$, the KL-Divergence between P_1 and the reference Gaussian must be 0, but due to our finite samplesize, the estimate of the KL-Divergence is not. Therefore we normalise the estimated KL curve by subtracting the first entry. Hence,

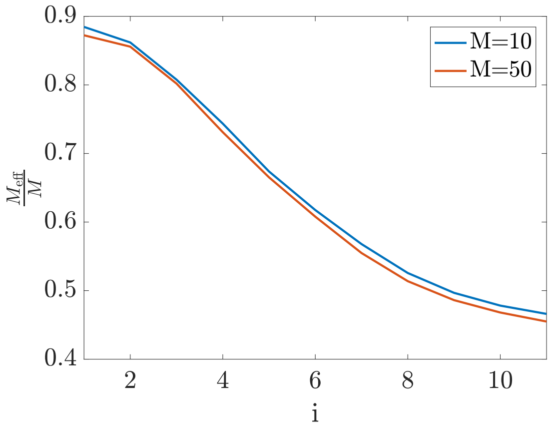


Figure 3.8: Effective sample size normalised by actual sample size of samples from bimodal distributions (with increasing distance between the two modes).

we define a parameter

$$p_1(i) := KL(i) - KL(1), \qquad i = 1, ..., 11. \tag{3.1.22}$$

We repeat the same procedure with the ratio of the effective sample size $M_{\text{eff}} = \sum_{i=1}^{M} \frac{1}{w_i^2}$ and M, and set a second parameter

$$p_2(i) := \frac{M_{\text{eff}}(i)}{M}, \qquad i = 1, ..., 11. \tag{3.1.23}$$

In Figure 3.8, p_2 is plotted against the distance of the two modes of the distributions.

Now, for every distribution P_i, $i = 1, ..., 11$, we set $x_i^{\text{ref}} = X_i^*$ with X_i^* being a randomly picked sample from the ensemble $X_i^1, ..., X_i^M$ and $y_i^{\text{obs}} = x_i^{\text{ref}} + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \rho)$. From the definition of y_i^{obs}, it follows that the observation operator is $H = 1$ and the measurement error variance is ρ. A hybrid filter is applied 1000 times to every distribution P_i (every time to another ensemble drawn from P_i) with $\alpha = 0, 0.1, ..., 1$. Each time the α that yields the minimal RMSE after applying the Hybrid filter is chosen, where $\alpha = 0$ means applying a pure EnKF and $\alpha = 1$ means applying a pure ETPF. After the 1000 repeats we take the average of the chosen α. In Figure 3.9 you can see the results for $\rho = 5$, for $M = 10$ and 50. Note that we only applied a single assimilation step and not a sequential procedure in time. Therefore the phenomenon of weight degeneracy in the ETPF part of the hybrid filter does not

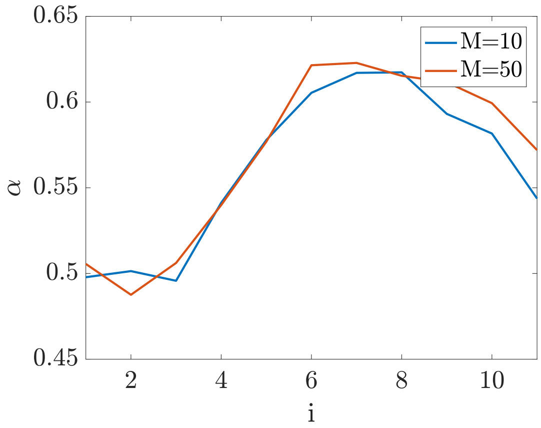


Figure 3.9: Chosen 'optimal' α for the hybrid filter plotted against the distance of the modes of the bimodal distribution that the ensemble represents.

come into effect strongly and the 'optimal' α are therefore much larger than they would be in a time dependent sequential application. This is especially noticeable in the fact that the 'optimal' α for $i = 1$ (meaning the underlying distribution is purely Gaussian) is approximately 0.5, which makes sense since in that case both, the ESRF and the ETPF, should be unbiased estimators.

Now we want to learn how to choose the optimal α, knowing p_1 and p_2 of the ensemble. On taking the ansatz

$$\alpha(p_1, p_2) = \beta_1 + \beta_2\, p_1 + \beta_3\, p_2 + \beta_4\, p_1^2 + \beta_5\, p_2^2 + \beta_6\, p_1 p_2, \tag{3.1.24}$$

define the matrix

$$A^M = \begin{pmatrix} 1 & p_1^M(1) & p_2^M(1) & p_1^M(1)^2 & p_2^M(1)^2 & p_1^M(1)p_2^M(1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & p_1^M(11) & p_2^M(11) & p_1^M(11)^2 & p_2^M(11)^2 & p_1^M(11)p_2^M(11) \end{pmatrix}. \tag{3.1.25}$$

We also have the vector

$$\left(\alpha^M(1) \cdots \alpha^M(11)\right) \tag{3.1.26}$$

at hand, as we just explained how to experimentally find the 'optimal' α. We fit the parameters with the results from $M = 10$ and $M = 50$ by solving the linear system

$$\begin{pmatrix} A^{10} \\ \hline A^{50} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_6 \end{pmatrix} = \begin{pmatrix} \alpha^{10}(1) \\ \vdots \\ \alpha^{10}(11) \\ \hline \alpha^{50}(1) \\ \vdots \\ \alpha^{50}(11) \end{pmatrix} \tag{3.1.27}$$

to obtain

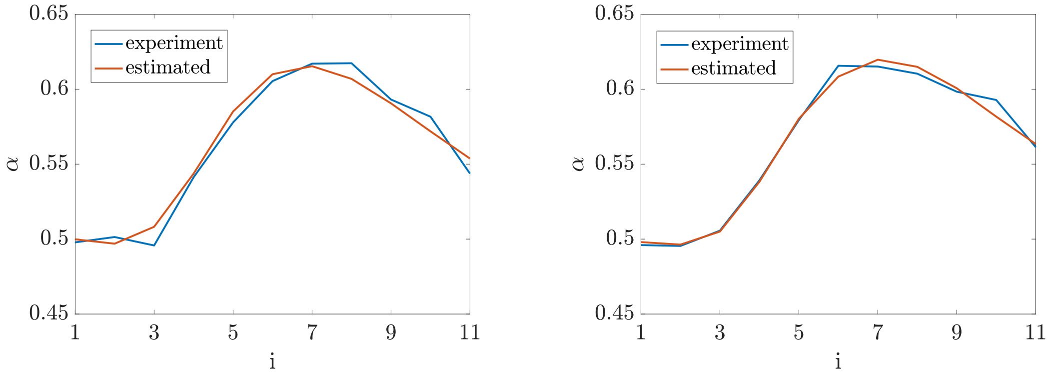
$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{pmatrix} = \begin{pmatrix} 2.47 \\ -1.53 \\ -4.79 \\ 0.14 \\ 2.89 \\ 2.34 \end{pmatrix}. \tag{3.1.28}$$

Now this β can be used to estimate the curves α^M for different M, by creating the matrix A^M and calculating

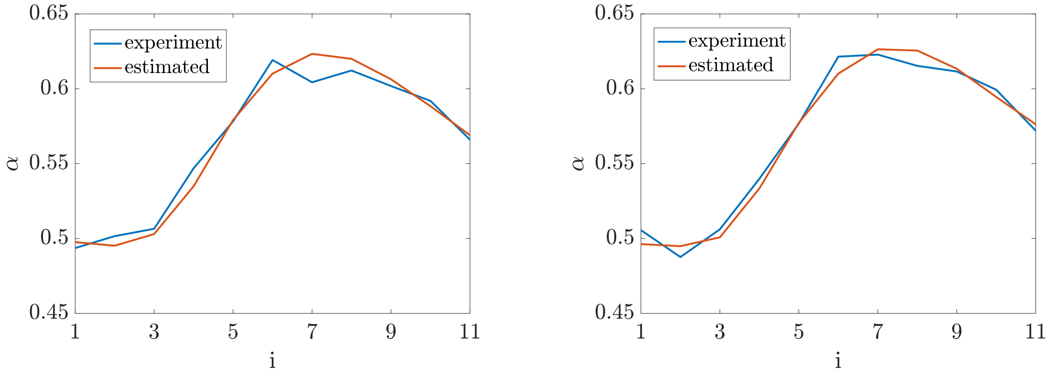
$$\tilde{\alpha}^M = A^M \beta. \tag{3.1.29}$$

We can see in Figure 3.10 the experimental curves α^M for $M = 10, 20, 30, 50$ and the estimated curves $\tilde{\alpha}^M$ for the same sample sizes in comparison. Note that the experimental results for $M = 20$ and 30 have not been used for fitting the parameters $\beta_1, ..., \beta_6$.

One can see upon inspection that the estimated choice of α is very close to the α that was actually chosen in the experiment in order to minimise the RMSE.



(a) $M = 10$ (b) $M = 20$



(c) $M = 30$ (d) $M = 50$

Figure 3.10: Estimated and experimental curves $\tilde{\alpha}^M$ and α^M respectively for the hybrid filter applied to bimodal distributions.

3.2 Gaussian mixture approach

Another approach to bridge between the group of Kalman filters and particle filters is the following: given a forecast ensemble x_i^{f}, $i = 1, \ldots, M$, and a bridging parameter $\alpha \in [0, 1]$, we define the prior distribution to be the following Gaussian mixture:

$$\pi_\alpha^{\mathrm{f}}(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^{M} \mathcal{N}(x; \bar{x}_i^{\mathrm{f}}, (2\alpha - \alpha^2) P_M^{\mathrm{f}}). \tag{3.2.1}$$

Here,

$$P_M^{\mathrm{f}} = \frac{1}{M-1} \sum_{i=1}^{M} (x_i^{\mathrm{f}} - \bar{x}^{\mathrm{f}})(x_i^{\mathrm{f}} - \bar{x}^{\mathrm{f}})^{\mathrm{T}} \tag{3.2.2}$$

denotes the empirical covariance matrix,

$$\bar{x}^{\mathrm{f}} = \frac{1}{M} \sum_{i=1}^{M} x_i^{\mathrm{f}} \tag{3.2.3}$$

is the empirical mean, and

$$\bar{x}_i^{\mathrm{f}} := x_i^{\mathrm{f}} - \alpha(x_i^{\mathrm{f}} - \bar{x}^{\mathrm{f}}) \tag{3.2.4}$$

are the centres of the components of the Gaussian mixture. By this definition, the mean and covariance matrix of π_α^{f} are equal to $\bar{x}^{\mathrm{f}}$ and P_M^{f} respectively, regardless of the value of the bridging parameter α. For $\alpha = 1$, the prior (3.2.1) reduces into a single Gaussian and for $\alpha = 0$, it leads to the empirical measure

$$\pi_0^{\mathrm{f}}(x) = \frac{1}{M} \sum_{i=1}^{M} \delta(x - x_i^{\mathrm{f}})\,. \tag{3.2.5}$$

This approximation goes back to Liu & West (2001).

If we assume that the likelihood is Gaussian with linear forward model H and measurement error covariance matrix R, then the posterior distribution is again a weighted Gaussian mixture, see for example Alspach & Sorenson (1972). In particular, the weights are given by

$$w_i \propto \exp\left(-\frac{1}{2}(H\bar{x}_i^{\mathrm{f}} - y_{\mathrm{obs}})^{\mathrm{T}}((2\alpha - \alpha^2)HP_M^{\mathrm{f}}H^{\mathrm{T}} + R)^{-1}(H\bar{x}_i^{\mathrm{f}} - y_{\mathrm{obs}})\right)\,. \tag{3.2.6}$$

The centres and covariance matrix of the posterior Gaussian mixture follow from applying a Kalman filter to each mixture component separately. Hence, we can calculate a Kalman gain matrix

$$K = \frac{(2\alpha - \alpha^2)P_M^{\mathrm{f}}H^{\mathrm{T}}}{\rho + (2\alpha - \alpha^2)HP_M^{\mathrm{f}}H^{\mathrm{T}}} \tag{3.2.7}$$

and from this,

$$\bar{x}_i^{\mathrm{h}} = \bar{x}_i^{\mathrm{f}} - K(H\bar{x}_i^{\mathrm{f}} - y_{\mathrm{obs}}), \tag{3.2.8}$$

$$P_M^{\mathrm{a}} = (2\alpha - \alpha^2)P_M^{\mathrm{f}} - KH(2\alpha - \alpha^2)P_M^{\mathrm{f}}. \tag{3.2.9}$$

Then, the posterior is given by the formula

$$\pi_\alpha^{\mathrm{a}}(x) = \sum_{i=1}^{M} w_i \, \mathcal{N}(x; \bar{x}_i^{\mathrm{h}}, P_M^{\mathrm{a}}). \tag{3.2.10}$$

Again note that $\alpha = 1$ leads to equal weights while $\alpha = 0$ leads to the classic importance weights associated with (3.2.5).

The analysis means (3.2.8) have a superscript 'h' here, because they are not yet representing the final centres, but rather an intermediate step. In fact, we can now apply an ETPF by solving the optimal transport problem (2.3.10) for the ensemble $\bar{x}_1^{\mathrm{h}}, ..., \bar{x}_M^{\mathrm{h}}$ with the weights given by (3.2.6). This produces equally weighted analysis centres $(\bar{x}_i^{\mathrm{a}})_{i=1}^M$ for each Gaussian mixture component. Afterwards, one has to sample from each component to obtain the analysis ensemble. In the following examples, we do that via

$$x_i^{\mathrm{a}} = \bar{x}_i^{\mathrm{a}} + (P_M^{\mathrm{a}})^{\frac{1}{2}} \left((2\alpha - \alpha^2)P_M^{\mathrm{f}}\right)^{-\frac{1}{2}} (x_i^{\mathrm{f}} - \bar{x}_i^{\mathrm{f}}) \tag{3.2.11}$$

but for higher dimensional problems, this will be too expensive and one can use an EnKF or ESRF.

As mentioned in the beginning of this chapter, adapting the Kalman filter to be applied to Gaussian mixtures, has already been done, in order to represent the involved probability distributions in a less restrictive way than assuming actual Gaussianity. See for example Bengtsson et al. (2003), Sun et al. (2009) and Dovera & Della Rossa (2011). Although these approaches base on the same concept of representing the posterior in the form (3.2.10), they do not include a bridging parameter that controls the proportion of the Kalman filtering and the importance sampling in the update step. There is also no ETPF used in the other approaches to produce equally weighted analysis centres.

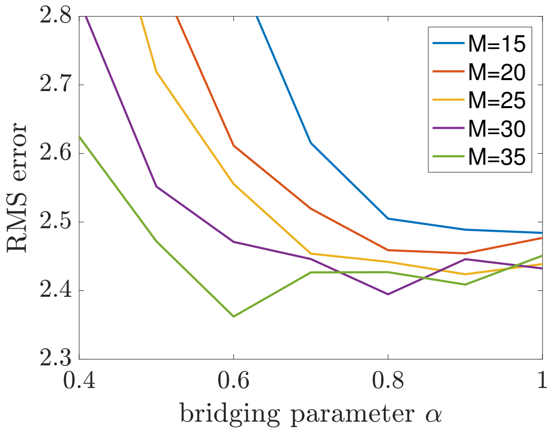


Figure 3.11: RMSE values for the Gaussian mixture hybrid filter applied to the Lorenz-63 model plotted against the bridging parameter α. Remark: now $\alpha = 0$ corresponds to ETPF and $\alpha = 1$ corresponds to EnKF.

3.2.1 Example: Lorenz-63

We apply this again to the Lorenz-63 model with the same setting as before in the beginning of this chapter. It turns out, that this approach does not work quite as well as the first hybrid approach (which was based on the likelihood splitting), at least for this example. The results can be seen in Figure 3.11. Again, it can be seen for all ensemble sizes but $M = 15$, that the RMSE becomes the smallest for an $\alpha \in (0, 1)$. The optimal α is decreasing with increasing ensemble size, since in this case, $\alpha = 0$ corresponds to a pure ETPF and $\alpha = 1$ to a Kalman filter. However, in general, the errors are bigger than with the hybrid, that was based on splitting the likelihood (compare to Figure 3.2).

3.2.2 Example: Lorenz-96

When applying this version of the hybrid to the Lorenz-96 model, there are two minor adjustments that need to be made. In formula (3.2.6), we have to calculate the weights for every grid point $j = 1, ..., 40$ separately since we want to apply

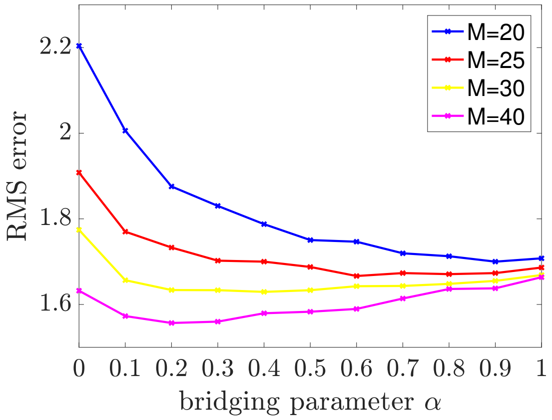


Figure 3.12: RMSE values for the Gaussian mixture hybrid filter applied to the Lorenz-96 model plotted against the bridging parameter α.

localisation. Therefore, we get

$$\begin{aligned} y_i &:= H\bar{z}_i^{\mathrm{f}} && \in \mathbb{R}^p \\ b &:= \mathbf{diag}(HP_M^{\mathrm{f}}H^{\mathrm{T}}) && \in \mathbb{R}^p \\ r &:= \mathbf{diag}(R) && \in \mathbb{R}^p \end{aligned} \tag{3.2.12}$$

$$w_i^j \propto \exp\left(-\frac{1}{2}\sum_{k=1}^{p}\rho_{kj}\frac{(y_i^k - y_{\mathrm{obs}}^k)^2}{b(k)+r(k)}\right), \tag{3.2.13}$$

with p being the number of observed grid points and ρ_{kj} the localisation factor for applying R-localisation depending on the distance of the k-th and j-th grid point. The second difference is, that in place of the P_M^{f} we use the localised version $\hat{P}_M^{\mathrm{f}} = C \circ P_M^{\mathrm{f}}$ as in formula (2.4.3). Thus, we are using both, R-localisation in the weights and also B-localisation. For both, we use a localisation radius of 4. Figure 3.12 shows the values of the RMSEs for the Gaussian mixture hybrid applied to the Lorenz-96 model in the same setting as before. For this example the Gaussian mixture approach works quite well, and yields almost the same results as the other hybrid (compare to Figure 3.4).

Chapter 4

Multi-scale models

The assimilation procedure applies an instantaneous change of the state of a system every time an observation becomes available by nudging the system closer to the observation. A multi-scale model usually fulfils some kind of balance relation between the variables that act on the different scales. As we have seen, the analysis ensemble is a linear transformation of the forecast ensemble and therefore the analysis ensemble members will no longer be balanced, unless the balance relation of the model is linear. When those unbalanced model states are fed into the model equations, high frequency oscillations are excited. The general problem of unbalanced initial conditions leading to non-physical predictions and introducing fast oscillations when considering systems with more than one time scale, and ways in which to mitigate this issue, has been investigated by several researchers, since it has been a problem in the field of numerical weather forecasting from the beginning. In his book 'The Emergence of Numerical Weather Prediction: Richardson's Dream' (Lynch (2006)), Lynch writes that the reason why Richardson's forecast of the change of surface pressure failed so drastically was that he started off with an unbalanced initial field. One of the other interesting points he presents is that, while the existence of the so-called *slow manifold*, introduced by Lynch in 1980, can be proven for some very simple equation systems, there have been many papers discussing the existence of a slow manifold in more complicated models. Lorenz published two papers, one called 'On the existence of a slow manifold' (Lorenz (1986)) and one 'On the nonexistence of a slow manifold' (Lorenz & Krishnamurthy (1987)) in 1986 and 1987 respectively. Other examples include Jacobs (1991), Lorenz (1992), Boyd (1994), Fowler & Kember (1996) and Camassa & Tin (1996). Lynch's conclusion is that, despite it being generally accepted that a slow manifold for realistic atmospheric models does not exist in the mathematical sense, the fast oscillations are of small amplitude and the atmosphere stays close to a balanced state. The issue of Data Assimilation techniques producing unbalanced fields has been studied for example in Bloom et al. (1996), Lorenc (2003*a*) and Ourmières et al. (2006).

In this chapter we will first introduce the concept of multi-scale models, balance and the issues of Data Assimilation methods in this context, by initially considering highly-oscillatory Hamiltonian models followed by spatially extended models. This discussion will be accompanied with a review of existing methods to prevent imbalance caused by Data Assimilation and afterwards, a new method will be introduced. At the end of this chapter, this method will be tested on the mentioned models in form of numerical simulations and the results will be presented.

4.1 Highly-oscillatory Hamiltonian models

Lynch states on page 137 in his book, that 'the solutions to the balance problem for this system [the stiff spring pendulum] have much wider applicability and are relevant to the problems arising in general forecasting models'. Let us consider a generalisation of that. The highly-oscillatory Hamiltonian system is a model with energy

$$H^\varepsilon(p,q) = \frac{1}{2}p^{\mathrm{T}}M^{-1}p + \frac{1}{2\varepsilon^2}g(q)^{\mathrm{T}}Kg(q) + V(q), \tag{4.1.1}$$

where $M \in \mathbb{R}^{N\times N}$ is a diagonal matrix of masses, $V : \mathbb{R}^N \to \mathbb{R}$ is a potential, $g : \mathbb{R}^N \to \mathbb{R}^L$, $L \geq 1$ is the balance relation, $K \in \mathbb{R}^{L\times L}$ is a diagonal matrix of force constants and $\varepsilon > 0$ is a stiffness parameter. For reasons of simplicity, we use $M = I$. The equations of motion are given by

$$\dot{q} = p \tag{4.1.2}$$

$$\dot{p} = -\varepsilon^{-2}G(q)^{\mathrm{T}}Kg(q) - \nabla_q V(q), \tag{4.1.3}$$

where $G(q) := Dg(q) \in \mathbb{R}^{L\times N}$ denotes the Jacobian matrix of g at q. This type of model is problematic concerning its numerical treatment for the limit $\varepsilon \to 0$.
If we want a solution $(p^\varepsilon(t), q^\varepsilon(t))$ to be of bounded energy, i.e.

$$H^\varepsilon(p^\varepsilon(t), q^\varepsilon(t)) = \mathcal{O}(\varepsilon^0), \tag{4.1.4}$$

the imbalance $g(q^\varepsilon(t))$ must be of order ε. In other words, solutions $q^\varepsilon(t)$ have to stay close to the so called slow manifold

$$\mathcal{M} = \{q \in \mathbb{R}^N : g(q) = 0\}. \tag{4.1.5}$$

Since $G(q)^\mathrm{T}$ denotes the derivative of $g(q)$, its columns are orthogonal to the curves defined in equation (4.1.5) at q and give the fast direction of the momentum at that point. Let $B(q)^\mathrm{T}$ be orthogonal to $G(q)^\mathrm{T}$ in the sense that $G(q)B(q)^\mathrm{T} = 0$ and therefore a matrix containing the tangential vectors of the curves in point q and the slow direction of the momentum. We can now decompose the momentum vector into the fast and the slow part via

$$\begin{aligned} p &= p_{\text{fast}} + p_{\text{slow}} && (4.1.6)\\ &= G(q)^\mathrm{T} p_x + B(q)^\mathrm{T} p_y, && (4.1.7)\end{aligned}$$

which is equivalent to

$$B(q)^\mathrm{T} p_y = p - G(q)^\mathrm{T} p_x, \tag{4.1.8}$$

with $p_x, p_y \in \mathbb{R}^{L\times 1}$. Since $G(q)B(q)^\mathrm{T} = 0$, we get $G(q)\left(p - G(q)^\mathrm{T} p_x\right) = 0$ and from that

$$p_x = \left(G(q)G(q)^\mathrm{T}\right)^{-1} G(q)p. \tag{4.1.9}$$

Hence, the fast part of the momentum vector p is

$$\begin{aligned} p_{\text{fast}} &= G(q)^\mathrm{T} p_x && (4.1.10)\\ &= G(q)^\mathrm{T} \left(G(q)G(q)^\mathrm{T}\right)^{-1} G(q)p. && (4.1.11)\end{aligned}$$

Note that

$$p_{\text{slow}} = p - p_{\text{fast}}. \tag{4.1.12}$$

From (4.1.10) follows

$$\begin{aligned} p_{\text{fast}}^\mathrm{T} p_{\text{fast}} &= p_x^\mathrm{T} G(q)G(q)^\mathrm{T} p_x && (4.1.13)\\ &= (G(q)p)^\mathrm{T} \left(G(q)G(q)^\mathrm{T}\right)^{-1} G(q)p && (4.1.14)\end{aligned}$$

and from that we can deduce that the Hamiltonian describing the energy of the fast oscillations looks as follows

$$H^{\varepsilon}_{\text{osc}}(q,p) = \frac{1}{2}(G(q)p)^{\mathrm{T}}(G(q)G(q)^{\mathrm{T}})^{-1}G(q)p + \frac{1}{2\varepsilon^2}g(q)^{\mathrm{T}}Kg(q). \tag{4.1.15}$$

Let us define the fast position variable $x := g(q)$ and the fast impulse is p_x as defined above. The oscillatory Hamiltonian with respect to these variables is the following

$$H^{\varepsilon}_{\text{osc}}(x,p_x) = \frac{1}{2}p_x^{\mathrm{T}}(G(q)G(q)^{\mathrm{T}})p_x^{\mathrm{T}} + \frac{1}{2\varepsilon^2}x^{\mathrm{T}}Kx. \tag{4.1.16}$$

and with

$$\dot{x} = \dot{g}(q) = G(q)p = G(q)G(q)^{\mathrm{T}}p_x = \nabla_{p_x}H^{\varepsilon}_{\text{osc}} \tag{4.1.17}$$

$$\dot{p_x} = \ddot{g}(g) = -G(q)\left[\nabla V(q) + G(q)^{\mathrm{T}}\frac{K}{\varepsilon^2}g(q)\right] + g_{qq}[p,p] \tag{4.1.18}$$

$$\approx -G(q)G(q)^{\mathrm{T}}\frac{1}{\varepsilon^2}Kx \tag{4.1.19}$$

we see that this system is harmonic in x and p_x with frequency

$$\omega^{\varepsilon} = \sqrt{G(q)G(q)^{\mathrm{T}}\frac{K}{\varepsilon^2}}. \tag{4.1.20}$$

We will consider two different scenarios that fit into that model family to investigate how Data Assimilation procedures destroy balance relations and to demonstrate how we are able to overcome that issue with a post-processing step after the assimilation.

4.1.1 Scenario A

The first one provides a simple class of toy problems that represent balanced flow regimes in atmosphere-ocean dynamics, see Reich & Hundertmark (2011). We look at solutions $q^{\varepsilon}(t)$, $p^{\varepsilon}(t)$ of (4.1.2)–(4.1.3) with bounded first derivative

$$\dot{q}^{\varepsilon}(t) = \mathcal{O}(\varepsilon^0) \quad \text{and} \quad \dot{p}^{\varepsilon}(t) = \mathcal{O}(\varepsilon^0). \tag{4.1.21}$$

Looking again at equation (4.1.3)

$$\dot{p} = -\varepsilon^{-2} G(q)^{\mathrm{T}} K g(q) - \nabla_q V(q),$$

we can deduce

$$\nabla_q V(q^\varepsilon) + \varepsilon^{-2} G(q^\varepsilon) K g(q^\varepsilon) = \mathcal{O}(\varepsilon^0) \tag{4.1.22}$$

and therefore $g(q^\varepsilon(t)) = \mathcal{O}(\varepsilon^2)$. Formula (4.1.15) describes the oscillatory Hamiltonian as the sum of its kinetic energy part and its potential energy part. The law of equipartition of energy, first postulated by John James Waterston in 1845, states, that both parts must be equal in a statistical sense. It follows

$$H^\varepsilon_{\mathrm{osc}}(q^\varepsilon(t), p^\varepsilon(t)) = \mathcal{O}(\varepsilon^2) \tag{4.1.23}$$

and consequently $G(q^\varepsilon(t))p^\varepsilon(t) = \mathcal{O}(\varepsilon)$. From now on $z^\varepsilon(t) = (q^\varepsilon(t), p^\varepsilon(t))$. The tangent manifold of the models we consider in this scenario is defined by

$$T\mathcal{M} = \{(q,p) \in \mathbb{R}^{2N} : \, g(q) = 0, \, G(q)p = 0\}, \tag{4.1.24}$$

and solutions with $\dot{z}^\varepsilon(t) = \mathcal{O}(\varepsilon^0)$ stay ε-close to it, see for example Rubin & Ungar (1957) and Bornemann & Schütte (1997).

To fulfill (4.1.23) at the initial time, we use initial conditions of the form

$$q^\varepsilon_0 := q^0_0 + \varepsilon^2 \tilde{q}_0 \tag{4.1.25}$$

$$p^\varepsilon_0 := p^0_0 + \varepsilon \tilde{p}_0 \tag{4.1.26}$$

with $(q^0_0, p^0_0) \in T\mathcal{M}$ and $\tilde{q}_0, \tilde{p}_0 \in \mathbb{R}^N$.

Looking at the limit for $\varepsilon \to 0$, the main results of Rubin & Ungar (1957) state that solutions to the initial value problem

$$\dot{q^\varepsilon} = p^\varepsilon \qquad q^\varepsilon(0) = q^\varepsilon_0 \tag{4.1.27}$$

$$\dot{p^\varepsilon} = -\varepsilon^{-2} G(q^\varepsilon)^{\mathrm{T}} K g(q^\varepsilon) - \nabla_q V(q^\varepsilon) \qquad p^\varepsilon(0) = p^\varepsilon_0 \tag{4.1.28}$$

converge to solutions (q,p) of the differential algebraic limit system

$$\dot{q} = p \qquad q(0) = q^0_0 \in \mathcal{M} \tag{4.1.29}$$

$$\dot{p} = -G(q)^{\mathrm{T}} K \lambda - \nabla_q V(q) \qquad p(0) = p^0_0 \in T_{q_0}\mathcal{M} \tag{4.1.30}$$

$$0 = g(q)\,. \tag{4.1.31}$$

Hereby $\lambda \in \mathbb{R}^L$ is the corresponding Lagrange multiplier, which is determined by

$$0 = \ddot{g}(q) = -G(q)\left[\nabla V(q) + G(q)^{\mathrm{T}} K \lambda\right] + g_{qq}(q)[p,p]. \tag{4.1.32}$$

In order for the constraint formulation (4.1.29)–(4.1.31) to be consistent with the balanced free dynamics, it is necessary that

$$g(q^{\varepsilon}) = \varepsilon^2 \lambda(q^{\varepsilon}, p^{\varepsilon}), \tag{4.1.33}$$

where $\lambda(q,p)$ is determined by (4.1.32). It is demonstrated by Reich (1995), that we can improve the zero-order balance relation $(q,p) \in T\mathcal{M}$ therefore by (4.1.33).

It should be noted that initial conditions of the form

$$q_0^{\varepsilon} := q_0^0 + \varepsilon \tilde{q}_0 \qquad q_0^0 \in \mathcal{M},\ \tilde{q}_0 \in \mathbb{R}^N \tag{4.1.34}$$

$$p_0^{\varepsilon} := \tilde{p}_0 \qquad \tilde{p}_0 \in \mathbb{R}^N\,. \tag{4.1.35}$$

lead to an oscillatory energy (4.1.15) which is of order $\mathcal{O}(\varepsilon^0)$. In this case, an additional force term appears in the limiting equations (4.1.29)–(4.1.31), as we will see later in this chapter. See also Rubin & Ungar (1957), Takens (1980), Bornemann & Schütte (1997) for more details.

We now describe the stiff elastic double pendulum as an example of (4.1.27)-(4.1.28) as we will use it later for numerical experiments. It corresponds to system (4.1.1)–(4.1.3) with $q, p \in \mathbb{R}^{N=4}$ and

$$g(q) = \begin{pmatrix} \|q^1\| - l_1 \\ \|q^1 - q^2\| - l_2 \end{pmatrix}, \tag{4.1.36}$$

with q^1 and q^2 denoting the two mass points of the systems and l_1 and l_2 the lengths of the two pendulums. Thus, for this example we have

$$G(q)^{\mathrm{T}} = \begin{pmatrix} \frac{q_x^1}{\|q^1\|} & \frac{q_x^1 - q_x^2}{\|q^1 - q^2\|} \\ \frac{q_y^1}{\|q^1\|} & \frac{q_y^1 - q_y^2}{\|q^1 - q^2\|} \\ 0 & -\frac{q_x^1 - q_x^2}{\|q^1 - q^2\|} \\ 0 & -\frac{q_y^1 - q_y^2}{\|q^1 - q^2\|} \end{pmatrix}. \tag{4.1.37}$$

It is a simple interacting multi-scale model which is useful for Data Assimilation experiments due to its chaotic behaviour. In Figure 4.1 we show how the positions of a solution with initial conditions of the form (4.1.25)–(4.1.26) develop in time compared to a solution that has more general initial conditions, and in Figure 4.2 we show the time evolution of the total energy (4.1.1) and the oscillatory energy (4.1.15) for both sets of initial conditions. Here, we used $\varepsilon = 10^{-3}$, $V(q) = 10(q_2 + q_4)$, $K = \operatorname{diag}(1, 0.04)$, $l_1 = l_2 = 1$ to create these figures. We will also use these values in later parts for numerical experiments.

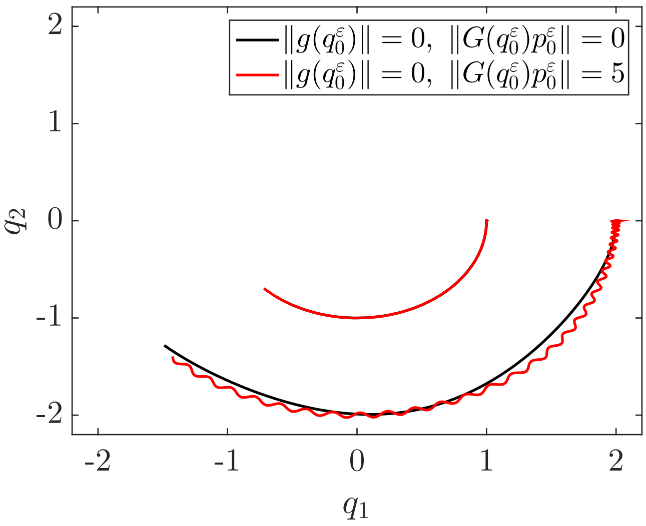


Figure 4.1: Evolution in time of a solution of the double pendulum model with initial conditions of the form (4.1.25)–(4.1.26) compared to a solution that has more general initial conditions, i.e. unbalanced initial velocity.

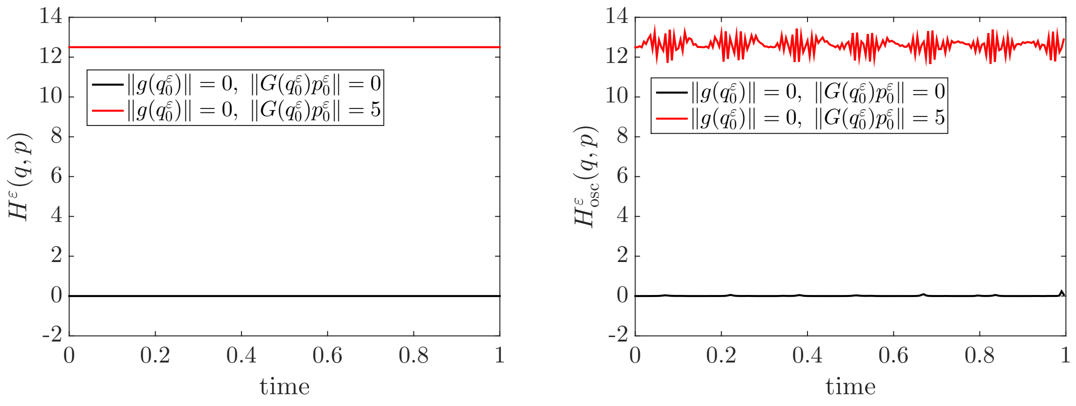
In Figure 4.3 we rerun the balanced solution for a series of decreasing values of ε and monitor values of $g(q)$. As one would expect $g(q)$ is of order ε^2 and $\tilde{g} := g(q) - \varepsilon^2\lambda(q,p)$ is slightly better preserved under the model dynamics than g, especially for small values of ε.

4.1.2 Scenario B

This second scenario couples the model (4.1.2)–(4.1.3) to a heat bath, which gives rise to Langevin dynamics of the form

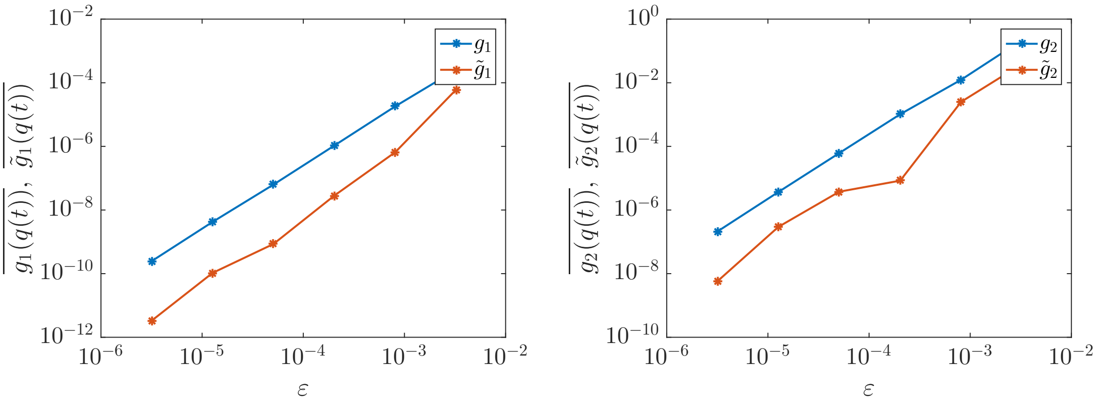
$$\dot{q} = p \tag{4.1.38}$$

$$\dot{p} = -\varepsilon^{-2} G(q)^{\mathrm{T}} K g(q) - \nabla_q V(q) - \gamma p + \sqrt{2\gamma \mathrm{k_B} T}\dot{W}, \tag{4.1.39}$$



(a) $H^\varepsilon(q,p)$ (b) $H^\varepsilon_{\mathrm{osc}}(q,p)$

Figure 4.2: Evolution in time of the total energy and the energy of the fast variable for a solution of the double pendulum model with initial conditions of the form (4.1.25)–(4.1.26), shown in black, compared to a solution that has more general initial conditions, shown in red.



(a) $g_1(q)$ (b) $g_2(q)$

Figure 4.3: Time-averaged values of $g_1(q(t)), g_2(q(t))$ and $\tilde g_1(q(t)), \tilde g_2(q(t))$ with $\tilde g(q) := g(q) - \varepsilon^2\lambda(q,p)$ for a series of balanced solutions of the double pendulum model corresponding to decreasing values of the stiffness parameter ε.

where $\gamma > 0$ is the friction constant, T the temperature, k_B the Boltzmann constant, and $W(t)$ standard N-dimensional Brownian motion. Model (4.1.38)–(4.1.39) has the invariant distribution

$$\pi^\varepsilon(q,p) \propto e^{-H^\epsilon(q,p)/k_B T} \tag{4.1.40}$$

and the energy is finite in the sense of

$$\mathbb{E}[H^\varepsilon(q,p)] = \int_{\mathbb{R}^{2N}} H^\varepsilon(z)\,\pi^\varepsilon(z)\,\mathrm{d}z = \mathcal{O}(\varepsilon^0). \tag{4.1.41}$$

This scenario arises from statistical mechanics and in particular from molecular dynamics simulations; see Reich (1995), Bornemann & Schütte (1997), Reich (2000).

We define the action variable

$$J^\varepsilon(q,p) := \frac{H^\varepsilon_{\text{osc}}(q,p)}{\omega^\varepsilon(q)} \tag{4.1.42}$$

which is nearly conserved in time, for harmonic oscillators with slowly varying frequency, i.e. $J^\varepsilon \approx$ const. For thermally embedded systems, this is not the case. Instead, it exists a stochastic differential equation for the evolution of J^ε, we refer the reader to Reich (2000).
Furthermore, we can define a correcting force term

$$F_{\text{corr}}(q) := -J^\varepsilon(t)\nabla_q\omega^\varepsilon(q), \tag{4.1.43}$$

which represents the influence of the harmonic oscillations normal to $T\mathcal{M}$ on the slow motion of the system. The existence of (4.1.43) makes estimating the model state via Data Assimilation techniques even harder as it requires estimation of $J^\varepsilon(t)$ to predict the slow motion along the constraint manifold $\mathcal{M}$.

The example from that family of models that we will use for our experiments is the *elliptic* elastic pendulum which is given by

$$\dot{q} = p \tag{4.1.44}$$

$$\dot{p} = -\varepsilon^{-2}G(q)^{\mathrm{T}}g(q) - \begin{pmatrix} g^0 \\ 0 \end{pmatrix} - \gamma p + \sqrt{2\gamma \mathrm{k_B} T}\dot{W} \tag{4.1.45}$$

$$g(q) = \sqrt{q^{\mathrm{T}}Aq} - 1, \tag{4.1.46}$$

with $q, p \in \mathbb{R}^2$. The matrix $A = \operatorname{diag}(1,\alpha)$ gives the geometry of an ellipse with $\alpha > 0$.

For the moment we assume no stochastic forcing i.e. $\gamma = 0$ and obtain the Hamiltonian

$$H(q,p) = \frac{p^{\mathrm{T}}p}{2} + \frac{1}{2\varepsilon^2} g(q)^2 + g^0 q_1. \tag{4.1.47}$$

The derivative of $g(q)$ is given as

$$G(q)^{\mathrm{T}} = \frac{Aq}{\sqrt{q^{\mathrm{T}}Aq}}. \tag{4.1.48}$$

We recall the oscillatory Hamiltonian (4.1.15) and its frequency (4.1.20) and deduce

$$\omega^{\varepsilon} = \sqrt{\frac{q^{\mathrm{T}}A^{\mathrm{T}}Aq}{q^{\mathrm{T}}Aq}} \tag{4.1.49}$$

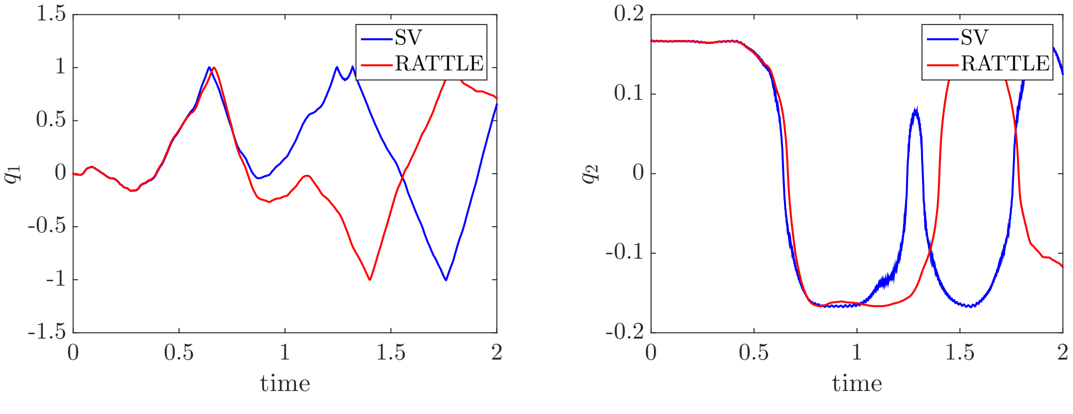
$$H^{\varepsilon}_{\mathrm{osc}} = (\omega^{\varepsilon})^2 \frac{{p_x}^2}{2} + \frac{1}{2\varepsilon^2} x^2. \tag{4.1.50}$$

Now we can determine the action variable as

$$J^{\varepsilon} := \frac{1}{2\omega^{\varepsilon}(q)} \left(\omega^{\varepsilon}(q)^2 p_x^2 + \frac{g(q)^2}{\varepsilon^2} \right). \tag{4.1.51}$$

This also enables us to find the correction force given by (4.1.43) for the limit system with non-tangential initial conditions. We also conclude that there is no additional correction term for the classical elastic pendulum (i.e. $\alpha = 1$), since $\omega^{\varepsilon} = \text{const}$ in this case. To consider the stochastic forcing i.e. $\gamma \neq 0$ introduced in (4.1.38)-(4.1.39) one can repeat all coordinate changes as above, but on the level of the equations of motion. Therefore one must split up the noise into x and y components. For a detailed analysis of this case we refer to Reich (2000).

In Figure 4.4 we show the evolution of q_1 and q_2 in the thermally embedded elliptic pendulum with $\varepsilon = 0.001$, $\gamma = 1$, $\mathrm{k_B}T = 16$, $\alpha = 36$ and $g^0 = 1$. The initial conditions are on the slow manifold but with an impuls, that is orthogonal to it. One run is done with the Störmer-Verlet method and the other one with the RATTLE method (Andersen (1983)), a numerical integration method which approximates solutions of the constrained (limit) problem (4.1.29)–(4.1.31). Hereby the numerical solution is as close to $T\mathcal{M}$, c.f. (4.1.24), as a given tolerance independent of the time step length. We used exactly the same realisations of the noise for both integrators. One can see that the two solutions behave very differently, which is caused by the RATTLE method setting the action variable (4.1.42), and therefore the correction force (4.1.43), to almost zero. In contrast the Störmer-Verlet method does not. This can be seen in Figure 4.5.



(a) q_1 (b) q_2

Figure 4.4: Evolution in time of the positions of the thermally embedded elliptic pendulum, once obtained by the RATTLE and once by the Störmer-Verlet method.

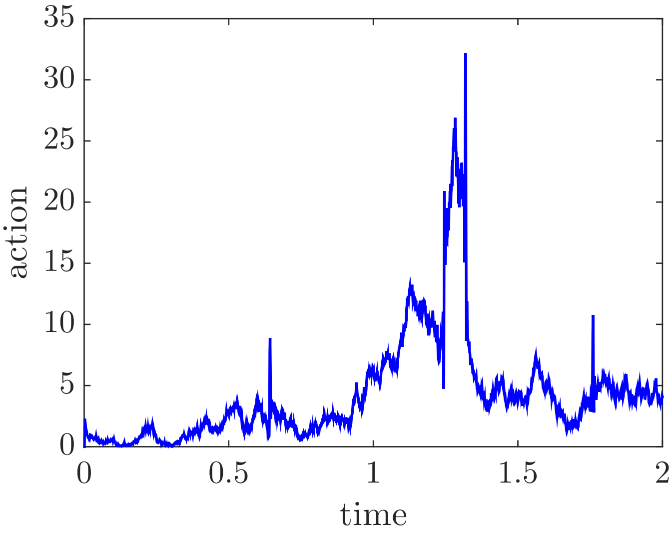


Figure 4.5: Evolution in time of the action variable of a solution obtained by the Störmer-Verlet method. In contrast to that, the action variable takes the value zero under the RATTLE approximation, because $x = g(q)$ and p_x are set to 0 in every time step.

4.2 Spatially extended models

In chapter 2.4 we have aforementioned that localisation is inevitable for spatio-temporal systems to deal with the undesired effects of the rather small sample sizes that we are restricted to, due to limited computational resources. An assimilation step including localisation looks like

$$z_j^{\mathrm{a}}(x_k) = \sum_{i=1}^{M} z_i^{\mathrm{f}}(x_k) d_{ij}(x_k),$$

as already stated in equation (2.4.6), while a standard Data Assimilation step is given by

$$z_j^{\mathrm{a}} = \sum_{i=1}^{M} z_i^{\mathrm{f}} d_{ij},$$

as in equation (1.3.6). This means that the global spatial update is replaced by seperate updates for each grid point, since the coefficients become dependent on the grid points. Consequently, localisation introduces further imbalances. See for example Houtekamer & Mitchell (2005), Cohn et al. (1998), Lorenc (2003*b*), Mitchell et al. (2002), Oke et al. (2007), Kepert (2009) and Greybush et al. (2011). In the following we will introduce two different spatially extended models, a variant of the aforementioned Lorenz-96 model and the well known shallow water equations.

4.2.1 Slow-fast Lorenz-96 model

We consider again the Lorenz-96 model

$$\dot{z}_l = (z_{l+1} - z_{l-2}) z_{l-1} - z_l + 8, \tag{4.2.1}$$

that we already used for experiments with the hybrid filter. Again, we choose $l = 1, ..., 40$ and we apply periodic boundary conditions, i.e. $z_0 = z_{40}$. In Bergemann & Reich (2010) this model is coupled with a discrete wave equation

$$\epsilon^2 \ddot{h}_l = -h_l + a^2 [h_{l+1} - 2h_l + h_{l-1}], \tag{4.2.2}$$

with a small parameter $\epsilon > 0$ that leads to the wave equation evolving on a faster time scale than the Lorenz equation, and a parameter $a > 0$ that determines the wave dispersion. As stated before, the energy

$$E_{\text{Lorenz}} = \frac{1}{2}\sum_{l=1}^{40} z_l^2 \tag{4.2.3}$$

is preserved in the advective part, $\dot{z}_l = (z_{l+1} - z_{l-2})z_{l-1}$ of the Lorenz model. The wave equation conserves the energy

$$E_{\text{wave}} = \frac{\varepsilon^2}{2}\sum_{l=1}^{40} \dot{h}_l^2 + \frac{1}{2}\sum_{l=1}^{40}[h_l^2 + a^2(h_{l+1} - h_l)^2]. \tag{4.2.4}$$

The equations are coupled together, leading to the following model:

$$\dot{z}_l = (1-\delta)(z_{l+1} - z_{l-2})z_{l-1} + \delta(z_{l+1}h_{l+1} - z_{l-2}h_{l-1}) - z_l + 8, \tag{4.2.5}$$
$$\epsilon^2\ddot{h}_l = -h_l + a^2[h_{l+1} - 2h_l + h_{l-1}] + z_l, \tag{4.2.6}$$

where $\delta \in [0,1]$ denotes the coupling strength between the two models. This is done by introducing an exchange energy term

$$E_{\text{coupling}} = -\delta\sum_{l=1}^{40} h_l z_l. \tag{4.2.7}$$

The pure wave-advection system

$$\dot{z}_l = (1-\delta)(z_{l+1} - z_{l-2})z_{l-1} + \delta(z_{l+1}h_{l+1} - z_{l-2}h_{l-1}), \tag{4.2.8}$$
$$\epsilon^2\ddot{h}_l = -h_l + a^2[h_{l+1} - 2h_l + h_{l-1}] + z_l, \tag{4.2.9}$$

conserves the total energy

$$E_{\text{total}} = (\delta - 1)E_{\text{Lorenz}} + \delta E_{\text{wave}} + E_{\text{coupling}}. \tag{4.2.10}$$

As it is desirable in physical applications, we want to find solutions to system (4.2.5) on the slow time scale. By the bounded derivative principle, see Kreiss (1980), this yields the conditions

$$\dot{z}_l \approx \mathcal{O}(\epsilon^0) \tag{4.2.11}$$
$$\ddot{h}_l \approx \mathcal{O}(\epsilon^0) \tag{4.2.12}$$

and therefore the slow manifold

$$\mathcal{M} = \{(z, h) \in (\mathbb{R}^{40}, \mathbb{R}^{40}) : -h_l + a^2[h_{l+1} - 2h_l + h_{l-1}] + z_l = 0\}. \tag{4.2.13}$$

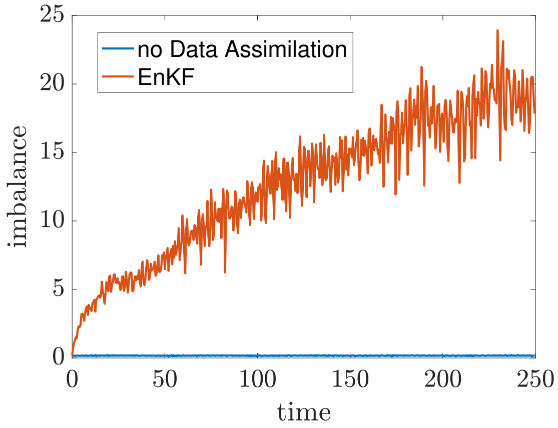


Figure 4.6: Evolution of the imbalance g of the slow-fast Lorenz-96 model with and without application of the EnKF.

If the initial state satisfies the balance relation (4.2.13), then the system (4.2.5) will stay in approximate balance over a fixed time interval, see for example Wirosoetisno (2004). For the choice of parameters $\epsilon = 0.0025$, $\delta = 0.1$, $a = 0.5$ we can see in Figure 4.6 on the left, that the imbalance $\|g\|$, with

$$g_l := -h_l + a^2[h_{l+1} - 2h_l + h_{l-1}] + z_l, \tag{4.2.14}$$

stays approximately constant. To illustrate the issue of regular Data Assimilation techniques regarding balance in multi-scale models, the imbalance $\|g\|$ is plotted against time in Figure 4.6 without and with the application of the LETKF with a localisation radius of $r_{\text{loc}} = 2$, observing every 4th grid point (i.e. $p = 10$ observation points) of z with a measurement error variance of $\rho = 1$.

4.2.2 Shallow water equations

We consider the following model

$$\partial_t h + \partial_x(hu) + \partial_y(hv) = 0 \tag{4.2.15}$$

$$\partial_t u + u\partial_x u + v\partial_y u - f^0 v + g^0 \partial_x h = 0 \tag{4.2.16}$$

$$\partial_t v + u\partial_x v + v\partial_y v + f^0 u + g^0 \partial_y h = 0. \tag{4.2.17}$$

These are the 2-dimensional shallow water equations. They describe the evolution of an incompressible, inviscid liquid or gas with a horizontal scale much larger than the vertical, in response to gravitational rotational accelerations. This model can be derived by depth-integrating the Navier Stokes equations and is widely used to describe flows in the earth's atmosphere and oceans. For a detailed derivation see for example Zeitlin (2007). The variable $h = h(x, y, t)$ describes the height of the liquid, $u = u(x, y, t)$ is the velocity in x-direction and $v = v(x, y, t)$ is the velocity in y-direction. The parameter g^0 is the acceleration due to gravity, while f^0 is the Coriolis coefficient. To conduct a scale analysis of this model, we first have to non-dimensionalise it. For this, we choose dimensionless reference values for the state variables, time and space as follows:

$$(h, u, v) = (\tilde{h}H_{\text{ref}}, \tilde{u}U_{\text{ref}}, \tilde{v}U_{\text{ref}}) \tag{4.2.18}$$

$$t = \tilde{t}T_{\text{ref}} \tag{4.2.19}$$

$$(x, y) = (\tilde{x}L_{\text{ref}}, \tilde{y}L_{\text{ref}}). \tag{4.2.20}$$

The derivatives with respect to the new dimensionless then beome

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial \tilde{t}}\frac{\partial \tilde{t}}{\partial t} = \frac{\partial}{\partial \tilde{t}}\frac{1}{T_{\text{ref}}} \tag{4.2.21}$$

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial \tilde{x}}\frac{\partial \tilde{x}}{\partial x} = \frac{\partial}{\partial \tilde{x}}\frac{1}{L_{\text{ref}}} \tag{4.2.22}$$

and model (4.2.15)-(4.2.17) turns into

$$\frac{L_{\mathrm{ref}}}{T_{\mathrm{ref}}U_{\mathrm{ref}}}\partial_{\tilde{t}}\tilde{h} + \partial_{\tilde{x}}(\tilde{h}\tilde{u}) + \partial_{\tilde{y}}(\tilde{h}\tilde{v}) = 0 \tag{4.2.23}$$

$$\frac{L_{\mathrm{ref}}}{T_{\mathrm{ref}}U_{\mathrm{ref}}}\partial_{\tilde{t}}\tilde{u} + \tilde{u}\partial_{\tilde{x}}\tilde{u} + \tilde{v}\partial_{\tilde{y}}\tilde{u} - \frac{f^0 L_{\mathrm{ref}}}{U_{\mathrm{ref}}}\tilde{v} + \frac{g^0 H_{\mathrm{ref}}}{U_{\mathrm{ref}}^2}\partial_{\tilde{x}}\tilde{h} = 0 \tag{4.2.24}$$

$$\frac{L_{\mathrm{ref}}}{T_{\mathrm{ref}}U_{\mathrm{ref}}}\partial_{\tilde{t}}\tilde{v} + \tilde{u}\partial_{\tilde{x}}\tilde{v} + \tilde{v}\partial_{\tilde{y}}\tilde{v} - \frac{f^0 L_{\mathrm{ref}}}{U_{\mathrm{ref}}}\tilde{u} + \frac{g^0 H_{\mathrm{ref}}}{U_{\mathrm{ref}}^2}\partial_{\tilde{y}}\tilde{h} = 0. \tag{4.2.25}$$

The dimensionless Strouhal, Rossby and Froude numbers are given as follows

$$\mathrm{Sr} := \frac{L_{\mathrm{ref}}}{T_{\mathrm{ref}}U_{\mathrm{ref}}} \tag{4.2.26}$$

$$\mathrm{Ro} := \frac{U_{\mathrm{ref}}}{f^0 L_{\mathrm{ref}}} \tag{4.2.27}$$

$$\mathrm{Fr} := \sqrt{\frac{U_{\mathrm{ref}}^2}{g^0 H_{\mathrm{ref}}}}. \tag{4.2.28}$$

On dropping the tilde symbols, the final dimensionless form of model (4.2.15)-(4.2.17) is

$$\mathrm{Sr}\partial_t h + \partial_x(hu) + \partial_y(hv) = 0 \tag{4.2.29}$$

$$\mathrm{Sr}\partial_t u + u\partial_x u + v\partial_y u - \frac{1}{\mathrm{Ro}}v + \frac{1}{\mathrm{Fr}^2}\partial_x h = 0 \tag{4.2.30}$$

$$\mathrm{Sr}\partial_t v + u\partial_x v + v\partial_y v + \frac{1}{\mathrm{Ro}}u + \frac{1}{\mathrm{Fr}^2}\partial_y h = 0. \tag{4.2.31}$$

In an experimental setting, one would choose a reference velocity U_{ref}, given a reference time T_{ref} and length L_{ref}, such that the Strouhal number becomes 1. Therefore, we ignore it from now on. Majda explains in his book (Mayda (2002)), that this model converges to the quasi-geostrophic model for $\mathrm{Fr}^2 \propto \mathrm{Ro} \propto \varepsilon \to 0$. If we let $\mathrm{Ro} \propto \mathrm{Fr}^2 \propto \varepsilon \to 0$ in (4.2.30) and (4.2.31), we obtain

$$v = \frac{H_{\mathrm{ref}}}{U_{\mathrm{ref}}L_{\mathrm{ref}}}\frac{g^0}{f^0}\partial_x h \tag{4.2.32}$$

$$u = -\frac{H_{\mathrm{ref}}}{U_{\mathrm{ref}}L_{\mathrm{ref}}}\frac{g^0}{f^0}\partial_y h, \tag{4.2.33}$$

which means that the coriolis term and the pressure gradient balance each other and consequently the flow is parallel to the isobars. This is called geostrophic balance, and solutions $(h^\varepsilon, u^\varepsilon, v^\varepsilon)$ of the full model (4.2.15)-(4.2.17) converge to solutions

(h^0, u^0, v^0) of the quasi geostrophic model. The flows in the earth's atmosphere are approximately in geostrophic balance, see for example Gill (1982), which makes this example very relevant from a practical viewpoint. As in the models before, if this balance relation is fulfilled in the initial state, it will stay approximately satifsfied, at least over a fixed time interval, see Klainerman & Majda (1981). Data Assimilation techniques however, can and will destroy it, leading to unphysical predictions.
The divergence of the velocity vector of a geostrophic flow is 0, because

$$\nabla \cdot \begin{pmatrix} u \\ v \end{pmatrix} = \partial_x u + \partial_y v \tag{4.2.34}$$

$$= \frac{g^0}{f^0}\partial_x\partial_y h - \frac{g^0}{f^0}\partial_y\partial_x h \tag{4.2.35}$$

$$= 0. \tag{4.2.36}$$

In the experiment section we will therefore use the balance function

$$g(h, u, v) = \nabla \cdot \begin{pmatrix} u \\ v \end{pmatrix}. \tag{4.2.37}$$

Let us first consider a simpler approximation of this model for the moment. The 1.5-dimensional shallow water equations

$$\partial_t h + \partial_x(hu) = 0 \tag{4.2.38}$$

$$\partial_t u + u\partial_x u - f^0 v + g^0 \partial_x h = 0 \tag{4.2.39}$$

$$\partial_t v + u\partial_x v + f^0 u = 0, \tag{4.2.40}$$

assume that the surface height and velocities do not vary in y direction, therefore all derivatives with respect to y are 0. Geostrophic balance reduces to the stationary state

$$u = 0 \tag{4.2.41}$$

$$v = \frac{g^0}{f^0}\partial_x h. \tag{4.2.42}$$

To simplify things even further, we consider a linearisation of model (4.2.38) - (4.2.40). Therefore, we choose a motionless basic state with constant surface height, i.e. (h^*, u^*, v^*) with

$$u^* = 0 \tag{4.2.43}$$

$$v^* = 0 \tag{4.2.44}$$

$$\partial_x h^* = 0 \tag{4.2.45}$$

(without the constant surface height, this state would not be motionless). Now, we apply the model to a small perturbation of this state:

$$\partial_t(h^* + \delta\tilde{h}) + \partial_x\left((h^* + \delta\tilde{h})(u^* + \delta\tilde{u})\right) = 0 \tag{4.2.46}$$

$$\partial_t(u^* + \delta\tilde{u}) + (u^* + \delta\tilde{u})\partial_x(u^* + \delta\tilde{u}) - f^0(v^* + \delta\tilde{v}) + g^0\partial_x(h^* + \delta\tilde{h}) = 0 \tag{4.2.47}$$

$$\partial_t(v^* + \delta\tilde{v}) + (u^* + \delta\tilde{u})\partial_x(v^* + \delta\tilde{v}) + f^0(u^* + \delta\tilde{u}) = 0. \tag{4.2.48}$$

Using the product rule in the derivatives, ignoring terms of order δ^2 (since we let $\delta \to 0$), using (4.2.43) - (4.2.45) and afterwards dropping δ and the tilde, we get

$$\partial_t h + h^*\partial_x u = 0 \tag{4.2.49}$$

$$\partial_t u - f^0 v + g^0\partial_x h = 0 \tag{4.2.50}$$

$$\partial_t v + f^0 u = 0. \tag{4.2.51}$$

Note that due to the motionlessness, $\partial_t h^*$ is also implied. We choose $x \in [-\pi, \pi]$ and periodic boundary conditions from now on.

To get some insight in the imbalance that can be caused by localisation, we conduct a simple experiment. We want to apply the EnKF with a very small localisation radius and a very small observation error to an initially balanced ensemble of states of the 1.5D shallow water model. To create an initial ensemble, we draw M functions $h_1(x,0), ..., h_M(x,0)$ from $\mathcal{GP}\left(h^{\text{ref}}(x,0), K_{hh}(x,x')\right)$, a Gaussian process with the Gaussian Kernel function

$$K_{hh} : \left\{ \begin{array}{lcl} [-\pi,\pi]^2 & \to & \mathbb{R} \\ (x,x') & \mapsto & \sigma^2 \exp\left(-\frac{2\sin(|x-x'|)^2}{l^2}\right) \end{array} \right. . \tag{4.2.52}$$

We set

$$u_j(x,0) := 0, \quad j = 1, ..., M \tag{4.2.53}$$

$$v_j(x,0) := \frac{g^0}{f^0} h_j'(x,0), \quad j = 1, ..., M. \tag{4.2.54}$$

Feeding $h_j(x,0), u_j(x,0), v_j(x,0),\ j = 1, ..., M$ in the model (4.2.49) leads to

$$\partial_t h = 0 \tag{4.2.55}$$

$$\partial_t u = 0 \tag{4.2.56}$$

$$\partial_t v = 0, \tag{4.2.57}$$

i.e. the ensemble stays unaltered during the forecast of q time steps. The initial ensemble 'covariance matrix'

$$P_M^0 = \begin{pmatrix} \hat{K}_{hh}(x,x') & 0 & \hat{K}_{hv}(x,x') \\ 0 & 0 & 0 \\ \hat{K}_{vh}(x,x') & 0 & \hat{K}_{vv}(x,x') \end{pmatrix}, \tag{4.2.58}$$

is assumed to be obtained in a way that $P_M^0 \to P^0$ for $M \to \infty$. The exact covariance matrix $P^0 \in \{\mathcal{L}^2([-\pi,\pi]^2,\mathbb{R})\}^{3\times 3}$ is given by

$$P^0 = \begin{pmatrix} K_{hh}(x,x') & 0 & K_{hv}(x,x') \\ 0 & 0 & 0 \\ K_{vh}(x,x') & 0 & K_{vv}(x,x') \end{pmatrix}, \tag{4.2.59}$$

with

$$K_{hv}(x,x') := \mathrm{Cov}(h(x), v(x')) = \frac{g^0}{f^0}\mathrm{Cov}(h(x), h'(x')) \tag{4.2.60}$$

$$K_{vh}(x,x') := \mathrm{Cov}(v(x), h(x')) = \frac{g^0}{f^0}\mathrm{Cov}(h'(x), h(x')) \tag{4.2.61}$$

$$K_{vv}(x,x') := \mathrm{Cov}(v(x), v(x') = \left(\frac{g^0}{f^0}\right)^2 \mathrm{Cov}(h'(x), h'(x')). \tag{4.2.62}$$

The matrices P_M^0 and P^0 are understood as matrices of functions. The third column entry in the first row for example, is the covariance function between h and v. Obviously, P_M^0 stays invariant during the forecast as well.

We only observe h at one discrete point x_k, which leads to the observation operator

$$H(z) = H\begin{pmatrix} h \\ u \\ v \end{pmatrix} := h(x_k) \tag{4.2.63}$$

and therefore $y_{\text{obs}} = h^{\text{ref}}(x_k, t_q) + \Xi$ with $\Xi \sim \mathcal{N}(0, r)$. The assimilation step with

B-localisation as described in chapter 2.4 then leads to

$$z_j^a = z_j^f - \begin{pmatrix} \tilde{\rho}_k(x)\hat{K}_{hh}(x,x_k) \\ 0 \\ \tilde{\rho}_k(x)\hat{K}_{vh}(x,x_k) \end{pmatrix} \underbrace{\frac{h_j^f(x_k) - h^{\mathrm{ref}}(x_k,t_q) - \Xi}{\sigma^2 + r}}_{:=c}. \tag{4.2.64}$$

Here, $\tilde{\rho}_k(x) := \rho\left(\frac{|x-x_k|}{r_{\mathrm{loc}}}\right)$ is defined via the Gaspari-Cohn function. This means that we are applying the same localisation function to h, u and v. We define

$$f_k(x) := K_{hh}(x,x_k) = \lim_{M\to\infty} \hat{K}_{hh}(x,x_k) \tag{4.2.65}$$

and

$$\begin{aligned} \lim_{M\to\infty} \hat{K}_{vh}(x,x_k) &= K_{vh}(x,x_k) && (4.2.66)\\ &= \mathrm{Cov}(v(x),h(x_k)) && (4.2.67)\\ &= \mathbb{E}\left[\frac{g^0}{f^0}\partial_x h(x)h(x_k)\right] - \mathbb{E}\left[\frac{g^0}{f^0}\partial_x h(x)\right]\mathbb{E}[h(x_k)] && (4.2.68)\\ &= \frac{g^0}{f^0}\partial_x\left(\mathbb{E}[h(x)h(x_k)] - \mathbb{E}[h(x)]\mathbb{E}[h(x_k)]\right) && (4.2.69)\\ &= \frac{g^0}{f^0}\partial_x K_{hh}(x,x_k) && (4.2.70)\\ &= \frac{g^0}{f^0} f_k'(x). && (4.2.71) \end{aligned}$$

We can finally deduce

$$\begin{aligned} h_j^a(x) &= h_j^f(x) - c\tilde{\rho}_k(x)f_k(x) && (4.2.72)\\ v_j^a(x) &= v_j^f(x) - \frac{g^0}{f^0}c\tilde{\rho}_k(x)f_k'(x). && (4.2.73) \end{aligned}$$

When we now check for geostrophic balance in the analysis ensemble,

$$\begin{aligned}
g(z_j) &:= f^0 v_j^a(x) - g^0 \frac{d}{dx} h_j^a(x) && (4.2.74)\\
&= f^0 \left(v_j^f(x) - \frac{g^0}{f^0} c \tilde{\rho}_k(x) \frac{d}{dx} f_k(x) \right) && (4.2.75)\\
&\qquad - g^0 \left(\frac{d}{dx} h_j^f(x) - c \left(\tilde{\rho}_k(x) \frac{d}{dx} f_k(x) + f_k(x) \frac{d}{dx} \tilde{\rho}_k(x) \right) \right) && (4.2.76)\\
&= \underbrace{f^0 v_j^f(x) - g^0 \frac{d}{dx} h_j^f(x)}_{=0} \underbrace{- g^0 c \tilde{\rho}_k(x) \frac{d}{dx} f_k(x) + g^0 c \tilde{\rho}_k(x) \frac{d}{dx} f_k(x)}_{=0} && (4.2.77)\\
&\qquad + g^0 c f_k(x) \frac{d}{dx} \tilde{\rho}_k(x), && (4.2.78)
\end{aligned}$$

we find that it is not fulfilled in case $\frac{d}{dx}\tilde{\rho}_k(x) \neq 0$, which is the case whenever $\tilde{\rho}_k$ is not constant, i.e., whenever any sort of localisation is applied. Therefore

$$\delta_k(x) := g^0 c f_k(x) \frac{d}{dx} \tilde{\rho}_k(x) \tag{4.2.79}$$

is the resulting imbalance.

To illustrate all this, we choose $h^{\text{ref}}(x,0) = 1 + 0.2\sin(x)$ and set the parameters of the kernel function to $\sigma = 0.05,\ l = 1$ for creating the initial ensemble. We discretise with $N = 1257$ grid points and we observe the 200-th grid point of h, which is approximately $h(-2.147)$ with an observation error variance of $r = 0.00005$. Now we apply the EnKF once with a localisation radius of $r_{\text{loc}} = 50$ and once $r_{\text{loc}} = 300$.

In Figure 4.7 we can see the kernel function $f_{200}(x)$ (see formula (4.2.65)), which is also the limiting function of $P^f_{1:N,200}$ for $N \to \infty$. The localisation function $\tilde{\rho}_k$ and the imbalance δ_k depend on the chosen localisation radius, which is why we will denote the radius as an additional subscript. In Figures 4.8 and 4.9, the localisation functions $\tilde{\rho}_{200,50}(x)$ and $\tilde{\rho}_{200,300}(x)$ and the resulting imbalances $\delta_{200,50}(x)$ and $\delta_{200,300}(x)$ are depicted respectively. It is clear from these figures that the smaller the chosen localisation radius, the larger will be the imbalance.

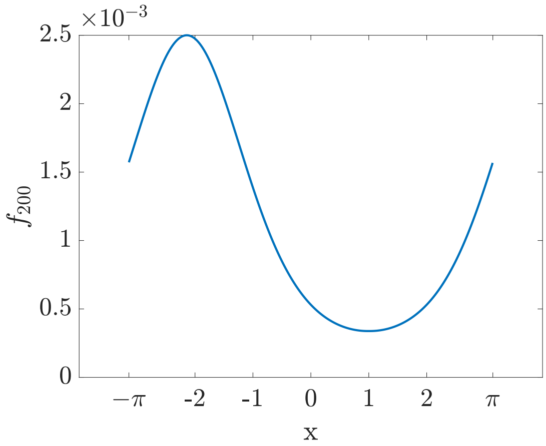
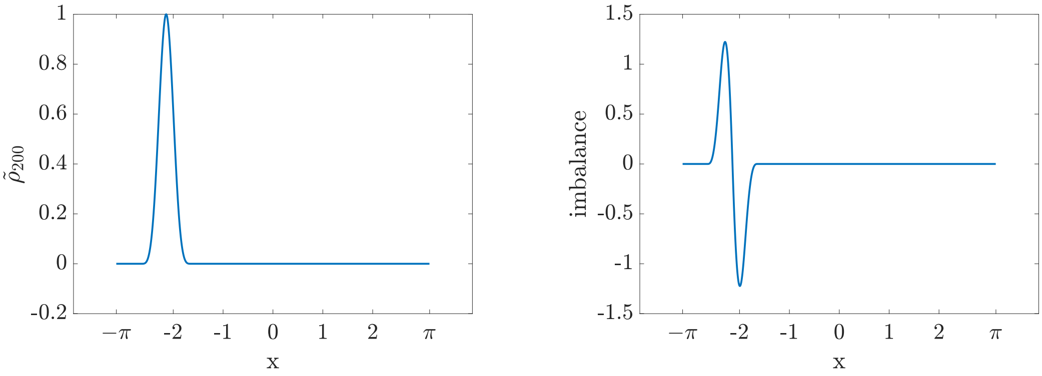


Figure 4.7: Kernel function $f_{200}(x)$.



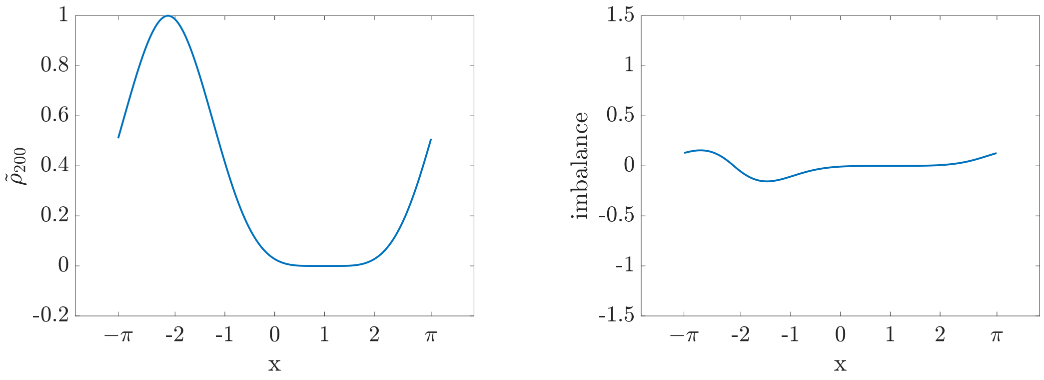
(a) Localisation function $\tilde{\rho}_{200,50}(x)$

(b) Imbalance $\delta_{200,50}(x)$

Figure 4.8: Localisation function and resulting imbalance in the 1.5D shallow water equation when applying an EnKF with B-localisation and localisation radius $r_{\text{loc}} = 50$.

4.3 Existing methods

Due to the sequential alternation between forecasting and assimilating of observations, one faces an initialisation problem every time after the analysis ensemble is obtained and before feeding it into the model. Therefore, normal mode initialisation as done in Machenhauer (1977) and Baer & Tribbia (1977) is a possible remedy. In Lynch & Huang (1992) it is suggested to apply a digital filter after every assimilation step to filter out spurious fast oscillations and this is applied to a weather prediction



(a) Localisation function $\tilde{\rho}_{200,300}(x)$

(b) Imbalance $\delta_{200,300}(x)$

Figure 4.9: Localisation function and resulting imbalance in the 1.5D shallow water equation when applying an EnKF with B-localisation and localisation radius $r_{\text{loc}} = 300$.

model.

This is probably the most intuitive approach. Let us assume we have an analysis ensemble $\{z_i^{\text{a}}(t_*)\}_{i=1}^{M}$. For each ensemble member we create a time sequence

$$z_i^{\text{a}}(t_{*-N}), z_i^{\text{a}}(t_{*-(N-1)}), ..., z_i^{\text{a}}(t_*), z_i^{\text{a}}(t_{*+1}), ..., z_i^{\text{a}}(t_{*+N}) \tag{4.3.1}$$

by using the model. Then we want to calculate the Fourier transform $Z_i^{\text{a}}(\omega)$ of $z_i^{\text{a}}(t)$.

In a continuous setting, the Fourier transform of an integrable function $f \in \mathcal{L}^1(\mathbb{R})$ is defined by

$$(\mathcal{F}f)(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt. \tag{4.3.2}$$

This transforms the time dependent signal f into its frequency representation. Each frequency is mapped to a coefficient, which describes the proportion of this frequency in the original signal. The reverse transform is defined by

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (\mathcal{F}f)(\omega)e^{i\omega t}d\omega. \tag{4.3.3}$$

An important property of the Fourier transform is the result of the convolution theorem. Let $*$ denote the convolution of two functions f and g:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-t)g(t)dt. \tag{4.3.4}$$

The convolution theorem states that the Fourier transform of the convolution of f and g is the same as the product of the Fourier transforms of f and g:

$$\mathcal{F}\left(f * g\right) = \mathcal{F}(f)\mathcal{F}(g). \tag{4.3.5}$$

A discrete approximation of the Fourier transform of $z_i^{\mathrm{a}}(t)$ is given by

$$\hat{Z}_i^{\mathrm{a}}(\omega) = \sum_{n=-N}^{N} z_i^{\mathrm{a}}(t_{*+n})e^{-i\omega n}. \tag{4.3.6}$$

Since we want to filter out high frequencies, we multiply $\hat{Z}_i^{\mathrm{a}}(\omega)$ with a cutoff function

$$H(\omega) = \begin{cases} 1, & |\omega| \leq |\omega_c|; \\ 0, & |\omega| > |\omega_c|. \end{cases} \tag{4.3.7}$$

This yields an approximation of $\hat{Z}_i^{\mathrm{a}}(\omega)$, in which the coefficients that belong to frequencies higher than ω_c, have been set to zero. Finally, the inverse Fourier transform has to be applied to $H(\omega)\hat{Z}_i^{\mathrm{a}}(\omega)$, to transform the filtered signal back into time-domain. Due to the convolution theorem, these three steps can be summarized to convoluting $z_i^{\mathrm{a}}(t)$ with the inverse Fourier transform $h(t) = \frac{\sin(\omega_c t)}{\pi t}$ of $H(\omega)$.

Strategies that incorporate the observational data in the model evolution in a gradual and smooth way instead of using all the information about the observation at one single point in time have been suggested for example in the from of *incremental analysis update* (IAU) by Bloom et al. (1996) and as the so-called mollified Data Assimilation by Bergemann & Reich (2010). In the latter, Bergemann and Reich used their continuous formulation of the EnKF to develop this approach. The unified differential equation (describing the model f and the EnKF step) is

$$\frac{dz_i}{dt} = f(z_i) - \frac{1}{2}\sum_{j=1}^{J} \delta(t-t_j)PH^{\mathrm{T}}R^{-1}(Hz_i + H\bar{z} - 2y^{\mathrm{obs}}(t_j)), \tag{4.3.8}$$

where $\delta(t)$ denotes the Dirac delta function. This is not actually a function, but a distribution. However, it can can loosely be thought of as a map, that is zero everywhere except for the origin where it is inifite:

$$\delta(t) = \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases} \tag{4.3.9}$$

This map is approximated by

$$\delta_\varepsilon(t) = \frac{1}{\varepsilon}\psi\left(\frac{t}{\varepsilon}\right). \tag{4.3.10}$$

Here $\varepsilon > 0$ is a small parameter and $\psi(t) \geq 0$ is a symmetric, continuous function with

$$\int_{-\infty}^{\infty} \psi(s)ds = 1, \tag{4.3.11}$$

for example the standard hat function

$$\psi(s) = \max\{0, 1 - |s|\}. \tag{4.3.12}$$

Replacing $\delta(s)$ in (4.3.8), gives the formulation of the mollified EnKF.

I partly worked together with Gottfried Hastermann, who developed a time-stepping method that is able to seamlessly bridge between a full oscillatory model and its stiff limiting system. This is achieved by using two integrators and bridging between them. One integrator has to project the solution onto the slow manifold, denoted by ψ^{slow} whereas the other, denoted by ψ^{fast}, has to be energy preserving. The full integrator is then composed via

$$z_{n+1} = \psi^\alpha(z_n) \tag{4.3.13}$$

$$\psi^\alpha := \alpha\psi^{\text{fast}} + (1-\alpha)\psi^{\text{slow}}. \tag{4.3.14}$$

For the two Hamiltonian models introduced in chapter 4.1, Hastermann used the classical symplectic Störmer Verlet method as ψ^{fast} and invented a completely new integrator for ψ^{slow}. Integrator (4.3.13) reduces to the limit model for $\alpha = 0$ and the full model for $\alpha = 1$. The idea is then to use ψ^{α_i} for k steps with $\alpha_1 = 0 < \alpha_2 < ... < \alpha_k = 1$ and to use ψ^1 for the remaining steps in the forecasting part of the Data Assimilation routine. For details, we refer the reader to Reinhardt et al.

(2017).

In Cotter (2013) a highly-oscillatory Hamiltonian form of the shallow water equations is investigated and a method for stabilizing the 4D-VAR technique is suggested.

Concerning imbalances introduced by localisation, Kepert (2009) modified the way of applying localisation into a more balance respecting manner. Gottwald (2014) incorporated information about the balance in the observations.

For variational Data Assimilation, it appears natural to include a penalty term describing the degree of imbalance in the minimisation problem. This has been done for example in Thépaut & Courtier (1991), Polavarapu et al. (2000), Gauthier & Thépaut (2001), Neef et al. (2006), Watkinson et al. (2007), Cotter (2013). This ansatz is related to what is going to be described now.

4.4 Post-processing approach

We now describe a post-processing method that allows the enforcement of balance relations on the analysis ensemble $\{z_i^{\mathrm{a}}(x_k, t_*)\}_{i=1}^{M},\ k = 1, ..., n$. From now on we drop the time-dependence for notational convenience and introduce the notation $\hat{z}_i$ for the unbalanced analysis ensemble, i.e. before applying the post-processing.

After an assimilation step is carried out, we have an ensemble, that was created using a linear transform of the form

$$\hat{z}_j(x_k) := \sum_{i}^{M} z_i^{\mathrm{f}}(x_k) d_{ij}(x_k) \qquad \text{or} \qquad \hat{z}_j := \sum_{i}^{M} z_i^{\mathrm{f}} d_{ij} \tag{4.4.1}$$

for $j = 1, \dots, M,\ k = 1, \dots, n$. We can apply the same transformation to the forecasted values of g as well, i.e.

$$\hat{g}_j(x_k) := \sum_{i} g(z_i^{\mathrm{f}}(x_k)) d_{ij}(x_k) \qquad \text{or} \qquad \hat{g}_j := \sum_{i} g(z_i^{\mathrm{f}}) d_{ij}, \tag{4.4.2}$$

$j = 1, \dots, M,\ k = 1, \dots, n$. The method works the same, regardless of whether the model has a space dimension or not. Therefore, we drop the emphasis on the grid points x_k from here on. It is important to note that $\hat{g}_j \neq g(\hat{z}_j)$ in general and that $g(\hat{z}_j)$ is usually a lot larger than $\hat{g}_j$.

The general idea now, is to solve M minimisation problems of the form

$$\mathcal{J}_j(z) = \frac{1}{2}\left(\hat{z}_j - z\right)^{\mathrm{T}} A \left(\hat{z}_j - z\right) + \frac{1}{2}\left(\hat{g}_j - g(z)\right)^{\mathrm{T}} B \left(\hat{g}_j - g(z)\right)^{\mathrm{T}}, \tag{4.4.3}$$

with $j = 1, ..., M$ and where A is a matrix weighting the uncertainty in $\hat{z}_j$ and B the uncertainty about $\hat{g}_j$. However, this agrees with the cost functional of 3D-Var (2.1.2) for $(P^b)^{-1} = A,\ R^{-1} = B$ and $y_{\text{obs}} = \hat{g}$. In case the slow manifold was linear, the observation operator would be $Hz = g(z)$. In the case that is most interesting for us, however, it is non-linear and we need to linearise $g(z) \approx g(\bar{z}) + Dg(\bar{z})(z - \bar{z})$. We can then interpret the Jacobian $Dg(\bar{z})$ as linear observation operator H^{T}. Since we know, that the solution of that minimisation problem will coincide with the result of applying a Kalman filter and we have M different observations $\hat{g}_j$, which we can use like perturbed observations, an EnKF implementation like

$$z_j^{\mathrm{a}} = \hat{z}_j - P_{zz}(Dg)\left(R + (Dg)^{\mathrm{T}} P_{zz}(Dg)\right)^{-1}\left(g(\hat{z}_j) - \hat{g}_j\right) \tag{4.4.4}$$

seems sensible. Here, we replaced the background covariance matrix P^b by the ensemble approximation

$$P_{zz} = \frac{1}{M-1}\sum_i \hat{z}_i(\hat{z}_i - \bar{\hat{z}})^{\mathrm{T}}. \tag{4.4.5}$$

However, instead of using $y_{\text{obs}}^j = \hat{g}_j$, we introduce a damping parameter $\gamma \in [0, 1]$ and use $y_{\text{obs}}^j = \gamma\hat{g}_j$ as observations. Here, as $\gamma \to 0$ the observations tend to 0 and for $\gamma \to 1$ they tend to $\hat{g}_j$, allowing us to control the amount of imbalance we permit in the analysis ensemble. Another possibility on how to implement this post-processing step is to use a Kalman-Bucy filter (as described in chapter 2.2.2), leading to the following two implementations:

$$\frac{\mathrm{d}}{\mathrm{d}s} z_j = -P_{zg} R^{-1}\left(g(z_j) - \gamma\hat{g}_j\right), \qquad j = 1, \ldots, M \text{ or} \tag{4.4.6}$$

$$\frac{\mathrm{d}}{\mathrm{d}s} z_j = -P_{zz}(Dg) R^{-1}\left(g(z_j) - \gamma\hat{g}_j\right), \qquad j = 1, \ldots, M \tag{4.4.7}$$

with $z_j(0) = \hat{z}_j$ and

$$P_{zg} = \frac{1}{M-1} \sum_i \hat{z}_i (g(\hat{z}_i) - \bar{g})^{\mathrm{T}}, \qquad \bar{g} = \frac{1}{M} \sum_i g(\hat{z}_i). \tag{4.4.8}$$

In formulation (4.4.6), the empirical covariance between z and g is used directly, which makes the observation operator $H^{\mathrm{T}} = Dg(\bar{z})$ dispensable. Therefore, we will mostly use formulation (4.4.7) in our experiments. Note also that implementation (4.4.4) can be made derivative free in the same way.

It is not immediately clear, which values we should choose for R but one could use

$$R := P_{gg} = \frac{1}{M-1} \sum_i \hat{g}_i (\hat{g}_i - \bar{\hat{g}})^{\mathrm{T}}. \tag{4.4.9}$$

In case of a spatially extended model we choose a localized version of P_{zz} in (4.4.7), i.e. we use $P_{zz}^{\mathrm{loc}} := C \circ P_{zz}$ instead of P_{zz}, where C is a localisation matrix as introduced in chapter 2.4. It is important to note, that $(C \circ P_{zz})H^{\mathrm{T}} \neq C \circ (P_{zz} H^{\mathrm{T}})$. If we want to use formulation (4.4.6), we therefore have to apply the operation g to all lines of P_{zz}^{loc} separately, to obtain P_{zg}^{loc}.

In all the experiments presented in section 4.5, we use a simple forward Euler scheme to numerically solve these ODEs. We start at $z_j(0) := \hat{z}_j$ from formula 4.4.1 and instead of integrating them artificially in time from $s = 0$ until $s = 1$ (as it is done in the classical Kalman-Bucy filter), we stop whenever

$$|g(z_j) - \gamma \hat{g}_j| \leq \mathrm{tol}, \tag{4.4.10}$$

with tol> 0 being a sufficiently small tolerance.

Note that $\gamma = 0$ implies that we nudge $g(\hat{z}_j)$ towards zero while $\gamma = 1$ nudges towards $\hat{g}_j$. When using formulation (4.4.4) the only additional computational effort comes from applying one more EnKF each time step than in a standard sequential Data Assimilation algorithm. Therefore, the total complexity stays $\mathcal{O}(M^3)$. When using the Kalman-Bucy variant, the cost depends on the tolerance value and on the stiffness of the underlying problem as this determines the time step and the numerical integration scheme for solving the ODE (4.4.6) or (4.4.7).

4.4.1 Modified method for scenario B

An alternative version, that can be applied in the special case of the thermally embedded highly oscillatory Hamiltonian, that we considered in chapter 4.1.2 is the following. For this model, the distribution of the imbalance is explicitly known, namely $g(q) \sim \mathcal{N}(0, \mathrm{k_B} T \varepsilon^2)$. Therefore we choose our 'measurement error' covariance $R = \mathrm{k_B} T \varepsilon^2$. When $g(q)$ is the observation operator, the EnKF with perturbed observations creates the assimilated ensemble via

$$q_i^{\mathrm{a}} = \hat{q}_i - K(g(\hat{q}_i) + \xi_i - y_{\mathrm{obs}}) \tag{4.4.11}$$

where ξ_i are identically distributed Gaussians with mean 0 and covariance equal to the measurement error covariance $\mathrm{k_B} T \varepsilon^2$, and $y_{\mathrm{obs}} = 0$ is our pseudo-observation. The Kalman gain matrix is

$$K = P_{qq}(Dg)\left(\mathrm{k_B} T \varepsilon^2 + (Dg)^{\mathrm{T}} P_{qq}(Dg)\right)^{-1}. \tag{4.4.12}$$

This can also again be formulated in the Kalman-Bucy framework, which leads to

$$\frac{\mathrm{d}}{\mathrm{d}s} q_j = -P_{zg} \frac{1}{\mathrm{k_B} T \varepsilon^2} \left(g(q_j) + \xi_i - y_{\mathrm{obs}}\right), \qquad j = 1, \ldots, M \text{ or} \tag{4.4.13}$$

$$\frac{\mathrm{d}}{\mathrm{d}s} q_j = -P_{qq}(Dg) \frac{1}{\mathrm{k_B} T \varepsilon^2} \left(g(q_j) + \xi_i - y_{\mathrm{obs}}\right), \qquad j = 1, \ldots, M. \tag{4.4.14}$$

4.5 Experiments

4.5.1 Hamiltonian Scenario A

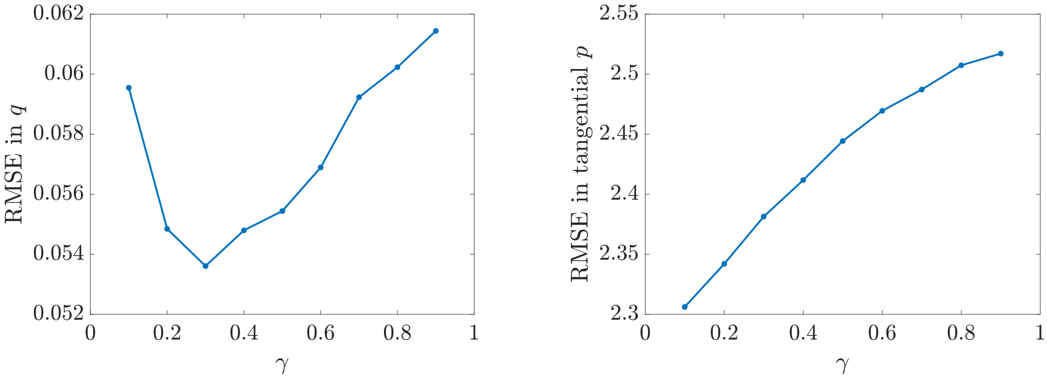
For all experiments we first create a reference trajectory denoted as z^{ref} which serves as our truth. We use this to create our observations by applying the observation operator H after each Δt_{obs} time units and perturbing the resulting quantities by adding noise with mean zero and covariance matrix $R = \rho I$.

For the double pendulum experiment we use $\varepsilon = 0.001$ and a stiffness matrix $K = \mathrm{diag}(1, 0.04)$. We set the gravitational constant to $g_0 = 10$, the ensemble size to $M = 20$ and we observe the coordinates of both pendulums every 0.02 time units

ε	K	g_0	δt	M	obs	Δt_{obs}	ρ	ini. var.	infl	total time
0.001	$\text{diag}(1, 0.04)$	10	0.001	20	q	0.02	0.05	0.1	1.05	200

Table 4.1: Parameters for the numerical experiments with the double pendulum.

with an observation error variance of $\rho = 0.05$. The initial uncertainty is Gaussian with a variance of 0.1 but all ensemble members are initially balanced. We apply multiplicative ensemble inflation with parameter infl $= 1.05$ after every assimilation step to counterbalance the underestimation of the covariance, as explained in chapter 2.4. We let the experiment run for 200 time units in total. The parameters of our chosen setting are summarised in Table 4.1.



(a) RMSE in q

(b) RMSE in p in the tangential direction

Figure 4.10: Time-averaged values of the RMSE in q and p in the tangential direction for the Kalman-Bucy post-processing method applied to the double pendulum, plotted against different values of γ.

Figure 4.10 displays the results of the Kalman-Bucy post-processing method (implementation version (4.4.7)) for different values of γ in terms of the RMSE in the positions and the momenta in the tangential direction:

$$\text{RMSE}_q := \left\| \begin{pmatrix} \overline{(q_x^1)}_M \\ \overline{(q_y^1)}_M \\ \overline{(q_x^2)}_M \\ \overline{(q_y^2)}_M \end{pmatrix} - \begin{pmatrix} (q_x^1)^{\text{ref}} \\ (q_y^1)^{\text{ref}} \\ (q_x^2)^{\text{ref}} \\ (q_y^2)^{\text{ref}} \end{pmatrix} \right\| \tag{4.5.1}$$

$$\text{RMSE}_p^{\text{tang}} := \left\| \begin{pmatrix} \overline{(p_x^{\text{tang},1})}_M \\ \overline{(p_y^{\text{tang},1})}_M \\ \overline{(p_x^{\text{tang},2})}_M \\ \overline{(p_y^{\text{tang},2})}_M \end{pmatrix} - \begin{pmatrix} (p_x^{\text{tang},1})^{\text{ref}} \\ (p_y^{\text{tang},1})^{\text{ref}} \\ (p_x^{\text{tang},2})^{\text{ref}} \\ (p_y^{\text{tang},2})^{\text{ref}} \end{pmatrix} \right\|. \tag{4.5.2}$$

As before $q^j = \begin{pmatrix} q_x^j \\ q_y^j \end{pmatrix}$, $j = 1, 2$ denotes the positions of the two mass points and $p^{\text{tang},j} = \begin{pmatrix} p_x^{\text{tang},j} \\ p_y^{\text{tang},j} \end{pmatrix}$, $j = 1, 2$ denotes the tangential momenta for both mass points. We recall, that the formula to obtain the tangential momenta was given by (4.1.12). The superscript ref means the reference solution and by $\overline{(q_x^1)}_M$, we denote the ensemble average of q_x^1. It can be seen that the RMSE in the positions is smallest for $\gamma = 0.3$, whereas the error in the tangential momenta increases with γ.

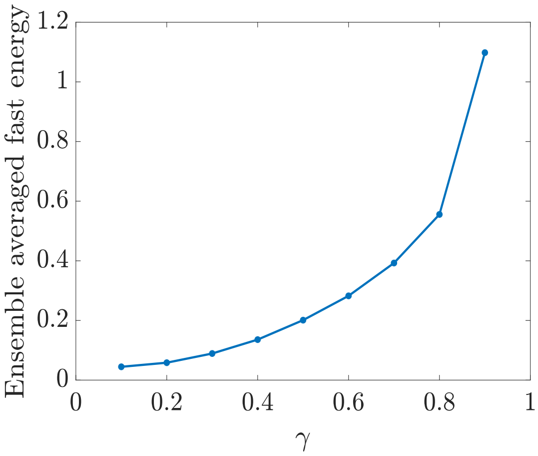
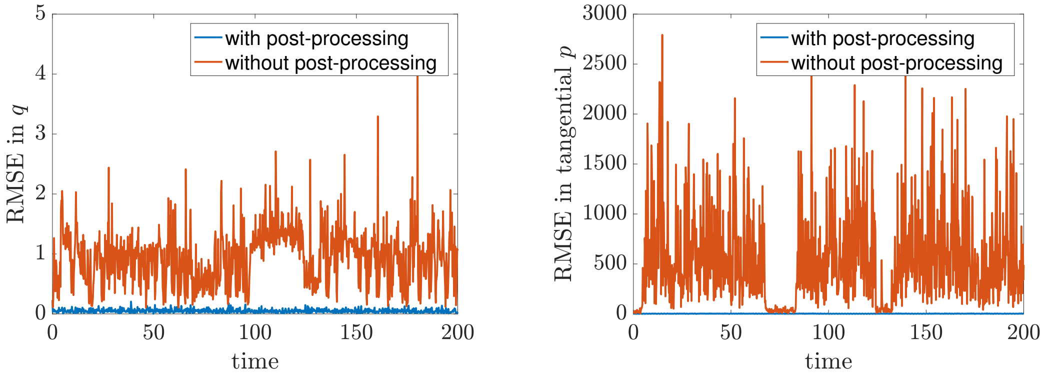


Figure 4.11: Time-averaged values of the ensemble average of the fast energy for the Kalman-Bucy post-processing method applied to the double pendulum, plotted against different values of γ.

The ensemble averaged fast energy of the double pendulum is depicted in Figure 4.11. As expected, this value is clearly increasing with γ.

In Figure 4.12, on the left, we see the RMSE in q plotted against time once with the post-processing method for $\gamma = 0.3$ (the choice of γ that yields the optimal result in terms of the RMSE in q) and once for the normal Data Assimilation procedure, without any post-processing. On the right we can see the same comparison for the RMSE in p in tangential direction, where $\gamma = 0.1$ is chosen for the post-processing. It is clearly visible that the balancing procedure has a significant positive effect on the prediction errors.



(a) RMSE in q

(b) RMSE in p in the tangential direction

Figure 4.12: RMSE values of the double pendulum plotted against time, once using the Kalman-Bucy post-processing and once without. On the left $\gamma = 0.3$ and on the right $\gamma = 0.1$.

So far, we used an EnKF to obtain (4.4.1) and (4.4.2), but it only seems natural to try and replace the EnKF by the ETPF or the hybrid formulation (3.1.6). Therefore we have to use particle rejuvenation instead of multiplicative inflation (see chapter 2.4) and also a larger ensemble size to avoid a degeneracy of weights. It turned out, that the whole algorithm, when using full or partly ETPF, becomes very sensitive to the choice of the rejuvenation parameter τ. A setting that worked well, is to choose the rejuvenation parameters $\tau_q = 0.1$, $\tau_p = 0.5$ and $\tau_g = 0.06$. We use $M = 100$ ensemble members, set $\gamma = 0.7$ and let the hybrid filter run for $\alpha = 0, 0.1, ..., 1$, once with Kalman-Bucy post-processing and once without. The time-averaged RMSE values in q are presented in Figure 4.13. One can see that using the ETPF instead of the EnKF already eliminates the problem with the imbalances. Nevertheless, the RMSE in q for the ETPF ($\alpha = 1$) is 0.072 with and 0.076 without post-processing. Using the EnKF together with post-processing with just $M = 20$ instead of $M = 100$ ensemble members yielded a smaller error of 0.058 (when using multiplicative inflation instead of rejuvenation).

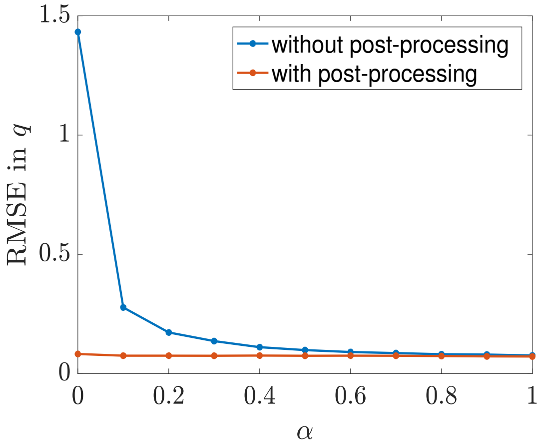


Figure 4.13: Time-averaged values of the RMSE in q with and without Kalman-Bucy post-processing applied to the double pendulum with the hybrid filter, plotted against different values of the bridging parameter α.

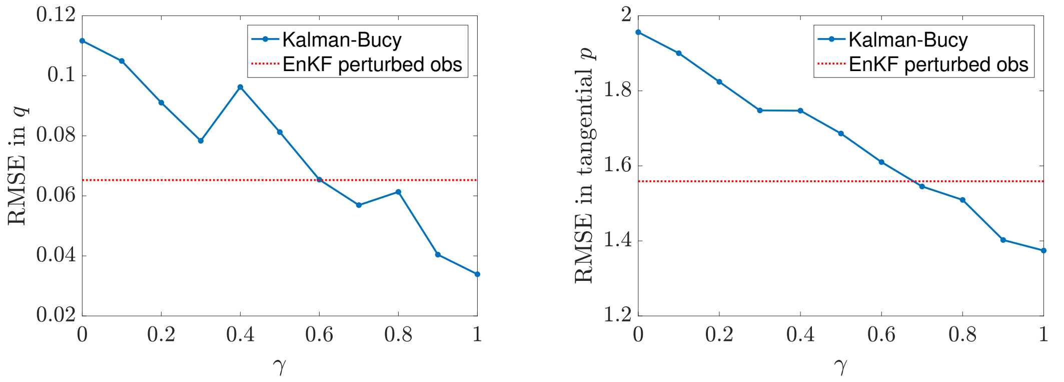
4.5.2 Hamiltonian Scenario B

ε	A	g^0	ζ	$\mathrm{k_B}T$	δt	M	obs	Δt_{obs}	ρ	ini. var.	time
0.001	$\mathrm{diag}(1, 36)$	0	1	16	5^{-6}	20	p	0.01	0.1	0.1	200

Table 4.2: Parameters for the numerical experiments with the thermally embedded system.

For our experiments with the thermally embedded elliptic elastic pendulum, we use model (4.1.44)-(4.1.46) with $A = \mathrm{diag}(1, 36)$. We set $\varepsilon = 0.001$, $g^0 = 0$, $\gamma = 1$ and $\mathrm{k_B}T = 16$. To avoid confusion with the tuning parameter in the post processing step, which we also denote by γ, we will change the notation of the damping parameter γ to ζ in this example. This time, we observe the momenta of the pendulum every 0.01 time units with an observation error variance of $\rho = 0.1$. Initial conditions are at the slow manifold but with an impulse in the normal direction and a Gaussian initial uncertainty with variance 0.1. We let the experiment run for 200 time units. The setting we chose for this model can be found in Table 4.2.

Figure 4.14 shows that choosing γ too small in the post-processing method (implementation version 4.4.7) yields non-optimal results. The best result is attained for $\gamma = 1$, i.e. when $\hat{g}_j$ is used as observation of the balance. The influence of the parameter γ on the action variable (4.1.42) of the ensemble is displayed in Figure



(a) RMSE in q (b) RMSE in p in the tangential direction

Figure 4.14: Time-averaged values of the RMSE for the Kalman-Bucy post-processing applied to the thermally embedded system, plotted against different values of γ. The red dotted line shows the same for the modified post-processing method.

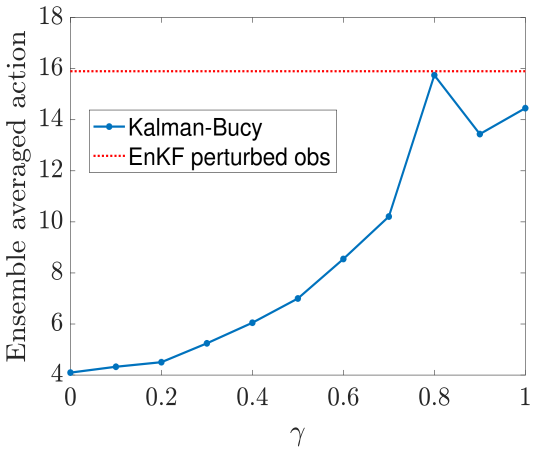
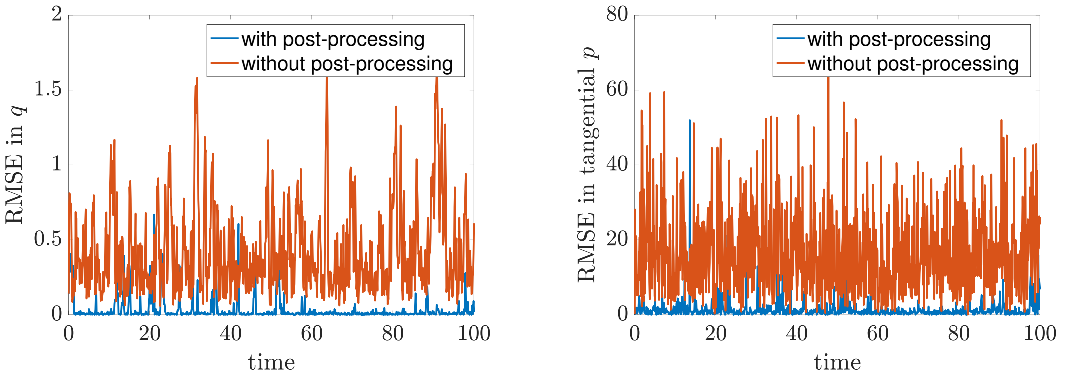


Figure 4.15: Dependence of the action variable J averaged over the ensemble and over time of the thermally embedded system on the balance parameter γ of the Kalman-Bucy post-processing method. The red dotted line shows the same for the modified post-processing method.

4.15. When choosing γ too small, J will be too small and therefore the correction force (4.1.43) will be too small as well. That leads to correspondingly erroneous predictions of the positions and the tangential velocity.

The results of using the EnKF with perturbed pseudo-observations as post-processing, the modified method, that was introduced in chapter 4.4.1, are depicted as a red dotted line in Figures 4.14 and 4.15. We obtained a RMSE in q of 0.065 and in p in the tangential direction of 1.56, which is equal to what was yielded by the standard post-processing for medium sized values of γ. The time and ensemble averaged action variable has a value of 15.9, which is close to the averaged action value we obtained with the standard post-processing for high values of γ and almost coincides with our chosen value for $\mathrm{k_B}T$.

In Figure 4.16, the RMSE values of the EnKF with perturbed pseudo observations post-processing are plotted against time in comparison to the RMSE values of a run without any post-processing. As in scenario A, both RMSEs are improved drastically by applying the post-processing strategy.



(a) RMSE in q

(b) RMSE in p in the tangential direction

Figure 4.16: RMSE values of the thermally embedded system, plotted against time, once using the post-processing and once without.

For this scenario and setting it was not possible to use an ETPF or hybrid formulation at all, as the weights collapsed in every simulation after some time.

4.5.3 Slow-fast Lorenz-96

In the setting for the Lorenz-96, as before, we use $\varepsilon = 0.0025$, wave dispersion $a = 0.5$ and coupling strength $\delta = 0.1$. We apply the LETKF with an ensemble size of $M = 50$, a localisation radius of $r_{\mathrm{loc}} = 2$ (R-localisation), observing every 4th grid point of z (resulting in $p = 10$ observed grid points) with an observation error variance of $\rho = 1$ and an observation interval of $\Delta t_{\text{obs}} = 0.05$. The variance of the initial ensemble is set to 1. The parameters are summarised in Table 4.3.

ε	a	δ	M	obs	p	Δt_{obs}	ρ	ini. var.	r_{loc}	total time
0.0025	0.5	0.1	50	z	10	0.05	1	1	2	500

Table 4.3: Parameters for the numerical experiments with the slow-fast Lorenz-96.

In the post-processing method, we use a localised version of P_{zz}, namely $P_{zz}^{\text{loc}} = C \circ P_{zz}$ where C is the localisation matrix as in chapter 2.4. We used a localisation radius of 3 for the post-processing in this experiment.

In Figure 4.17 we can see that the imbalance is drastically reduced by the Kalman-Bucy method and controlled by the tuning parameter γ, in the sense of it increasing monotonically with increasing γ. We didn't include the value of the imbalance for $\gamma = 1$, as it was 0.8 and hence much larger than the other depicted values.

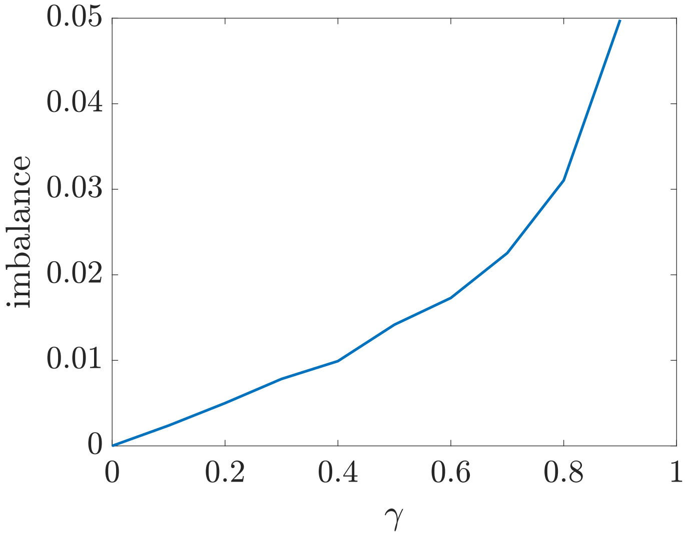


Figure 4.17: Imbalance of the slow-fast Lorenz-96 model using the Kalman-Bucy post-processing plotted against different values of γ.

The balance has a positive effect on the RMSE in z as can be seen in Figure 4.18. It didn't have an effect on the RMSE in h, though.

The more grid points we observed and the larger the localisation radius we used, the smaller was the effect of the balancing until the point, where balancing was

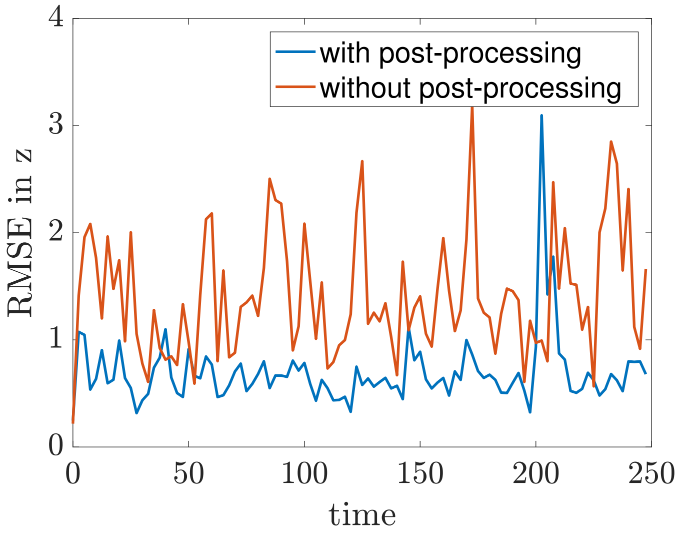


Figure 4.18: RMSE in z of the slow-fast Lorenz-96 model plotted against time, once using the Kalman-Bucy post-processing and once without.

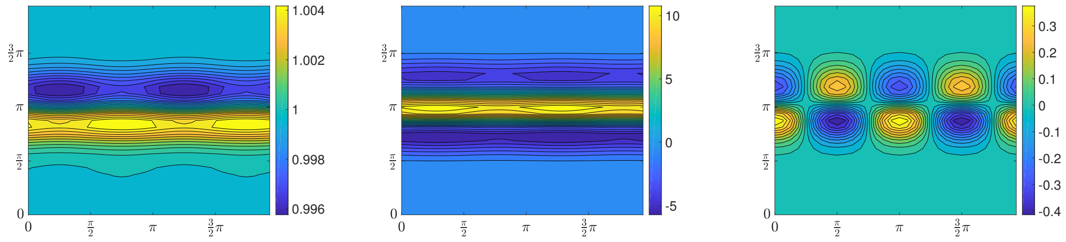
completely unnecessary.

It was possible to apply the ETPF (and a hybrid formulation) to the slow-fast Lorenz-96 model as well, yielding comparable results as the LETKF, but every attempt to combine the ETPF with the post-processing approach failed for this model, in the sense that the resulting RMS errors were much higher than using the LETKF. One could see, that the variance of the $\hat{g}_j,\ j = 1, ..., M$ when using the ETPF (or a hybrid formulation) was converging to 0 very quickly with time, while it stayed more or less constant in time (after a short calibration phase) when using the LETKF. Surprisingly this happened even when we applied rejuvenation to the ensemble of $\hat{g}_j,\ j = 1, ..., M$. Applying the post-processing step with almost equal $\hat{g}_j,\ j = 1, ..., M$ is comparable to using the EnKF without perturbed observations, which is not possible as we aforementioned in chapter 2.2.1.

4.5.4 Shallow water equations

For our experiment with the shallow water model, we consider the polar jet stream and therefore look at a shear flow. We examine it on a square with length $L = 2\pi$ discretised with $N = 32$ grid points in each direction (i.e. 1204 grid points in total).

As in Staniforth et al. (2006), we choose a scaler $\tilde{L} = \frac{60}{dx}$ km, such that the physical length and width of the domain we are looking at are $L \times \tilde{L} = 3840$ km and the grid width is $\frac{3840}{N-1} \approx 124$ km. We consider the SWE on an f-plane placed at 45° latitude and at a reference height of $H^0 = 9665$ m. The Coriolis parameter is given by $f^0 = 2 \cdot \omega \cdot \sin(\varphi)$, with ω being the angular frequency $\frac{2\pi}{T}$ ($T = 1$ day in our case) and φ is the latitude. Therefore we get $f^0 \approx 8.9 \frac{\text{rad}}{\text{day}}$ or $1.0284 \times 10^{-4} \frac{\text{rad}}{s}$. The acceleration due to gravity is set to $g^0 = 9.81 \frac{\text{m}}{s^2}$. We assume periodic boundary conditions and use the time-staggered semi-Lagrangian discretisation from Staniforth et al. (2006) with a time step of $\delta_t = 0.0139$ which represents time steps of 20 minutes. The initial state is depicted in Figure 4.19.



(a) H (b) u (c) v

Figure 4.19: Initial state of the reference solution for the simulations with the shallow water equations.

At time $t = 0$, the height of the liquid is 1, with a small valley and elevation, parallel to the x-axis, respectively to the left and to the right of $y = \pi$. The depth and height of the valley and the elevation are sinusoidal depending on x. In detail, at first $H(x, y)$ is given as

$$H(x, y) = \frac{256L^2}{Ng^0} \left(\left[\frac{1}{\pi}(y - \pi)\exp\left(-2(y - \pi)^2\right)(1 + \kappa\sin(2x)) + 1 \right]^{-1} - 1 \right) + 1, \tag{4.5.3}$$

with $\kappa = 0.05$.

The velocities u and v are chosen via

$$u = -\frac{g^0}{f^0}\partial_y H \tag{4.5.4}$$

$$v = \frac{g^0}{f^0}\partial_x H \tag{4.5.5}$$

to get geostrophically balanced initially conditions. Afterwards $H(x, y)$ is again altered, such that the following balance relation is satisfied

$$\beta^2[1-\beta^2\Delta]^{-1}(g^0\Delta H - f^0\zeta + 2J) = 0. \tag{4.5.6}$$

Here, Δ is the Laplacian and

$$\zeta = \partial_x v - \partial_y u \tag{4.5.7}$$

$$J = \partial_x v\ \partial_y u - \partial_x u\ \partial_y v \tag{4.5.8}$$

$$\beta = \frac{0.75\ \delta_t\sqrt{g^0}}{\sqrt{1+(\delta_t)^2\frac{(f^0)^2}{4}}}. \tag{4.5.9}$$

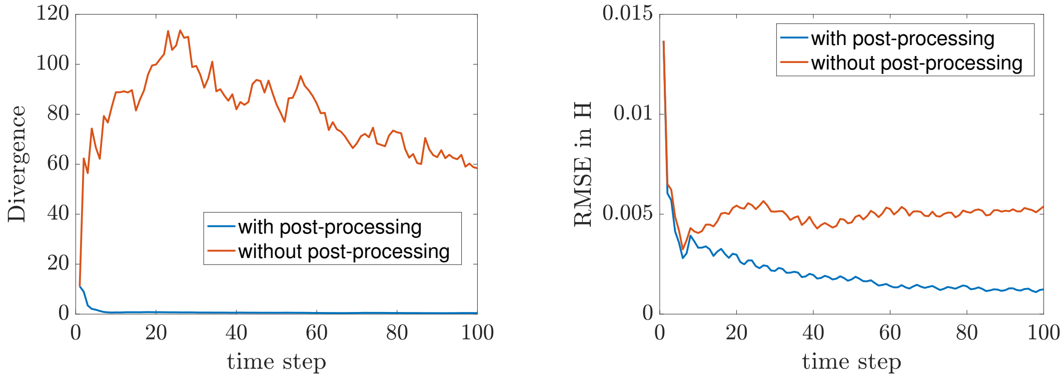
This balance condition is connected to the numerical integration and as well explained in Staniforth et al. (2006). To create an initial ensemble, we use formula (4.5.3) again, but with M different κ_i, drawn from $\mathcal{N}(0.05, 10^{-3})$ and an additive perturbation in x and y with variance 10^{-2}. Afterwards, u_i and v_i, $i = 1, ..., M$ are again given via formula (4.5.4) and finally the H_i are again adapted via formula (4.5.6).

We apply the LETKF with localisation radii of $r_{\text{loc}} = 1$ and $r_{\text{loc}} = 3$. We use $M = 20$ ensemble members and we observe the layer depth H on 4, 8 and 16 grid points in each direction (i.e. 16, 64 and 256 grid points in total, respectively) with an observation error variance of $\rho = 10^{-8}$ and $\rho = 10^{-4}$. The extreme setting (very small localisation radius and very small observation error) is chosen on purpose to elucidate the effect of Data Assimilation on the balance of the model and to point out the effectiveness of the post-processing method. We let the model run for a total time of $T = 69.4$ days and an observation time step of $\Delta t_{\text{obs}} = 0.694$, which corresponds to 16 hours and 40 minutes. We also conducted experiments with an observation interval of $\Delta t_{\text{obs}} = 0.42$, but there was not enough imbalance built up during the forecasting step to make any balancing strategy necessary and effective. In the post-processing method, we use P_{zz}^{loc} as we did in the experiments with the slow-fast Lorenz-96 model. Here, we use a localisation radius of three grid points. For R in formula (4.4.6), we use the identity matrix in this specific case because simulations with R as in (4.4.9) gave slightly worse results. The setting is summarised in Table 4.4.

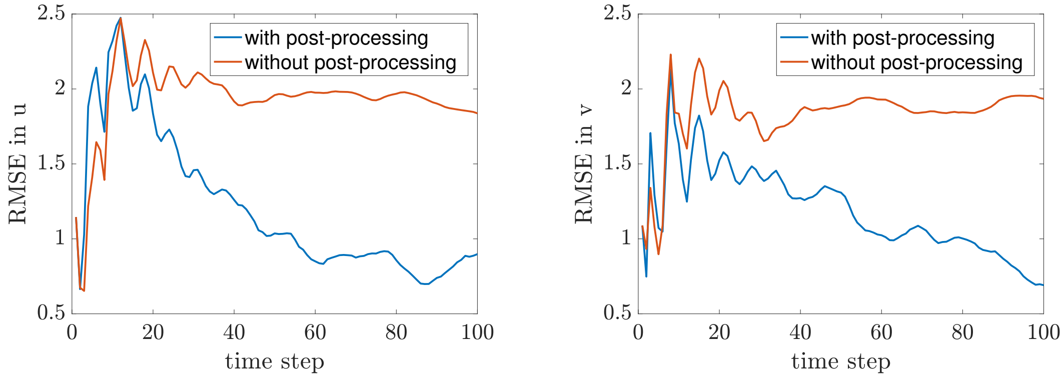
We conducted the experiments five times with different random seeds and averaged over the results.

L	N	f^0	g^0	M	obs	p	Δt_{obs}	ρ	r_{loc}	total time
3840 km	32	$1.0284 \times 10^{-4} \frac{\text{rad}}{s}$	$9.81 \frac{m}{s^2}$	20	H	$4,\ 8,\ 16$	0.694 days	$10^{-8},\ 10^{-4}$	$1,\ 3$	69.4 days

Table 4.4: Parameters for the numerical experiments with the shallow water model.



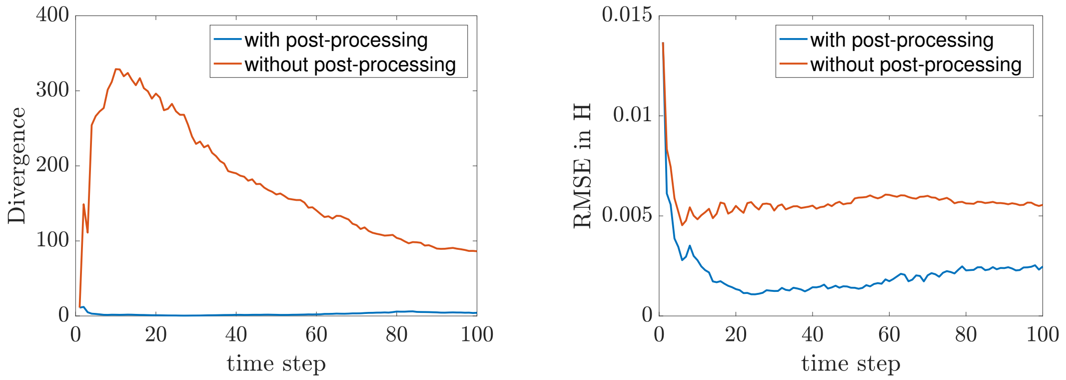
(a) Divergence (b) RMSE in H



(c) RMSE in u (d) RMSE in v

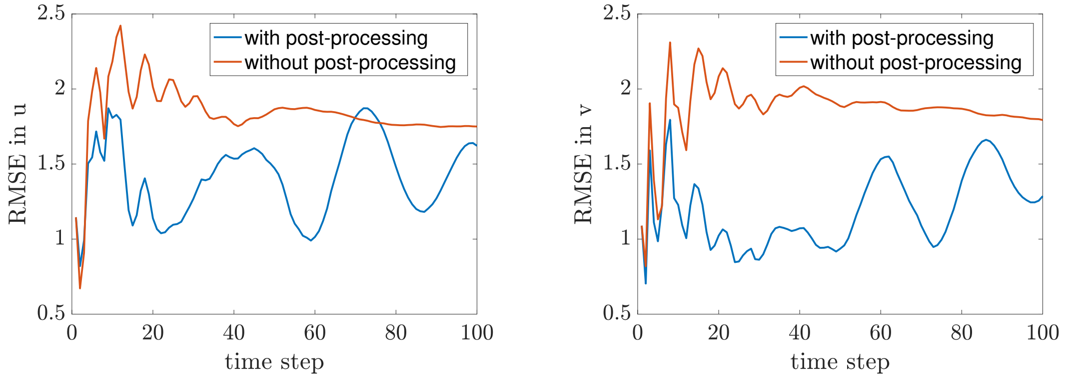
Figure 4.20: Values of the divergence and RMSE for the LETKF applied to the shallow water model, plotted against time, once with and once without Kalman-Bucy post-processing, for $r_{\text{loc}} = 1,\ p = 4,\ \rho = 10^{-8}$.

In Figures 4.20 - 4.22, we can see the divergence and the RMSE in all three state variables respectively plotted against time and compared to a run without any post-processing. This is done for $p = 4,\ 8,\ 16$ observed grid points in each direction and for the most extreme setting, in the sense, that the localisation radius is chosen as small as possible, namely $r_{\text{loc}} = 1$ and the observation error is very small, namely $\rho = 10^{-8}$. One can clearly see that the divergence is much smaller, when the post-processing method is applied and for $p = 4$ and $p = 8$, this has a significant positive



(a) Divergence

(b) RMSE in H



(c) RMSE in u

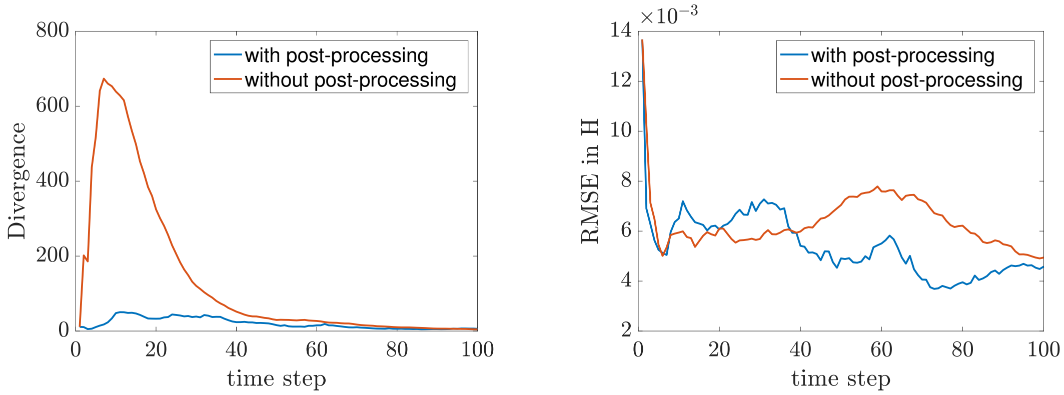
(d) RMSE in v

Figure 4.21: Values of the divergence and RMSE for the LETKF applied to the shallow water model, plotted against time, once with and once without Kalman-Bucy post-processing, for $r_{\text{loc}} = 1,\ p = 8,\ \rho = 10^{-8}$.

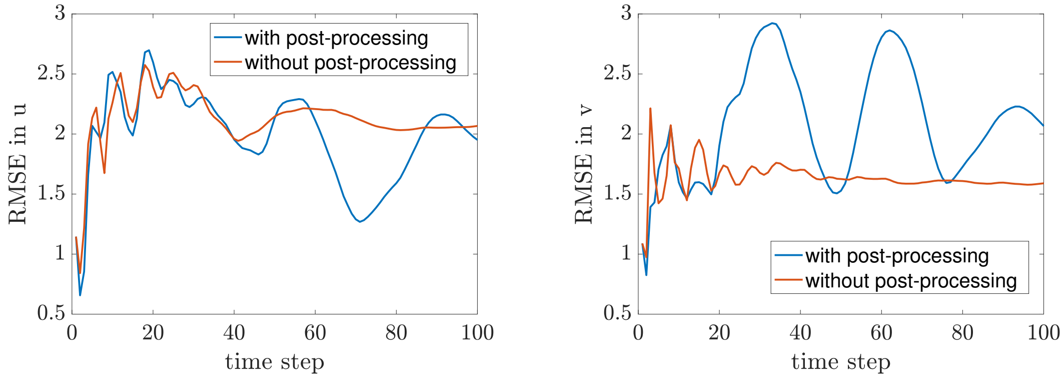
effect on the RMSE in all three state variables. For $p = 16$ on the other hand, the balancing strategy fails to improve the RMSE values. The benefit and effect of the balancing procedure becomes less apparent, the more grid points we observe. This behaviour coincides with our experience from the slow-fast Lorenz-96 model.

Figure 4.23, depicts the divergence and RMSE values for $p = 4,\ r_{\text{loc}} = 1$ and a larger measurement error of $\rho = 10^{-4}$. The post-processing reduces the error in the fluid depth clearly but only slightly the errors in the velocities.

In Figure 4.24, the results for a larger localisation radius of $r_{\text{loc}} = 3$ and $p = 4,\ \rho = 10^{-8}$ are presented. Also for this setting, an improvement of the RMSE values was



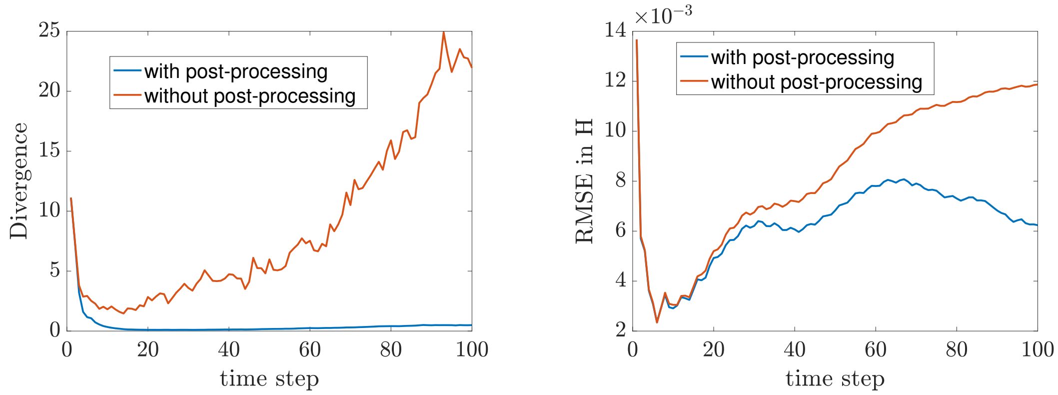
(a) Divergence (b) RMSE in H



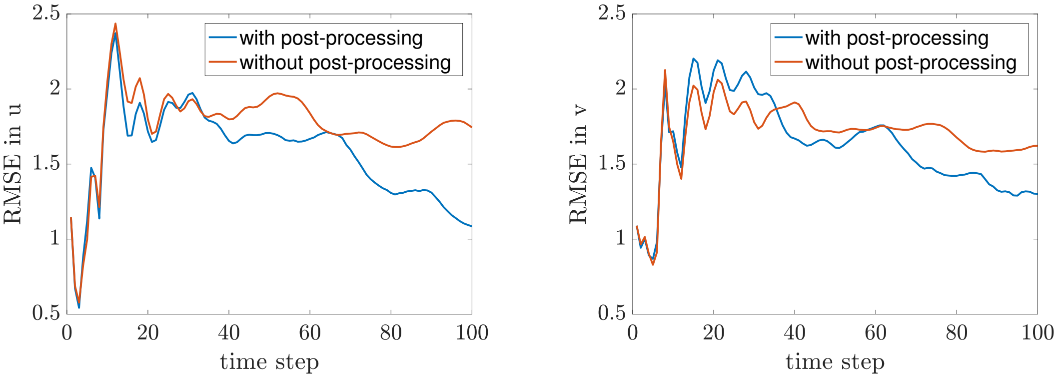
(c) RMSE in u (d) RMSE in v

Figure 4.22: Values of the divergence and RMSE for the LETKF applied to the shallow water model, plotted against time, once with and once without Kalman-Bucy post-processing, for $r_{\text{loc}} = 1,\ p = 16,\ \rho = 10^{-8}$.

achieved by applying the post-processing.

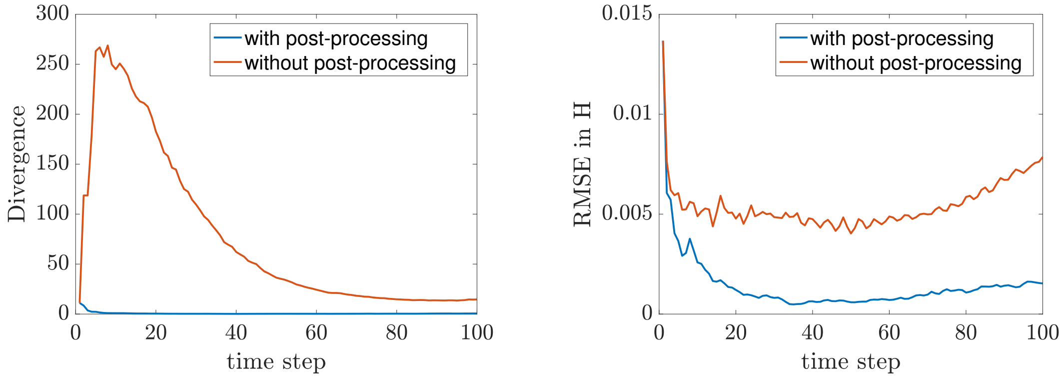


(a) Divergence (b) RMSE in H



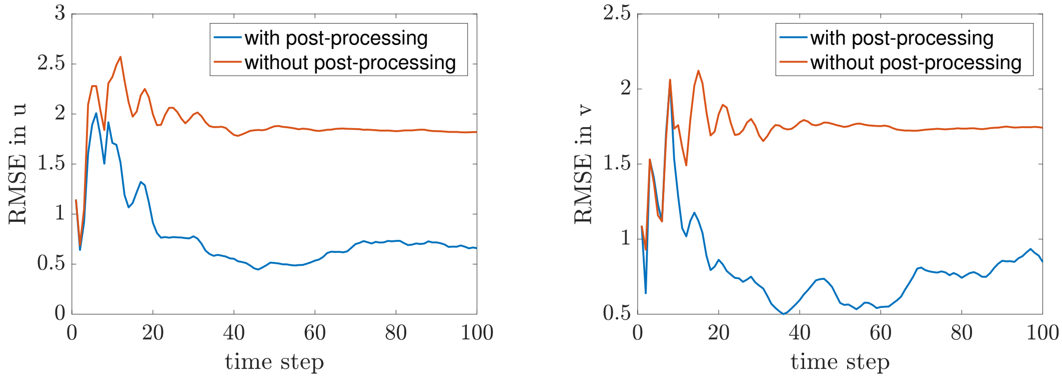
(c) RMSE in u (d) RMSE in v

Figure 4.23: Values of the divergence and RMSE for the LETKF applied to the shallow water model, plotted against time, once with and once without Kalman-Bucy post-processing, for $r_{\text{loc}} = 1,\ p = 4,\ \rho = 10^{-4}$.



(a) Divergence

(b) RMSE in H



(c) RMSE in u

(d) RMSE in v

Figure 4.24: Values of the divergence and RMSE for the LETKF applied to the shallow water model, plotted against time, once with and once without Kalman-Bucy post-processing, for $r_{\text{loc}} = 3,\ p = 4,\ \rho = 10^{-8}$.

Chapter 5

Conclusion

In this thesis we considered novel approaches to resolve issues arising in classical and significant algorithms in the field of Data Assimilation. In the first part we focused on hybrid filters, that bridge between ensemble Kalman filters (EnKF) and particle filters. This was achieved by splitting the likelihood function

$$\pi(y^{\mathrm{obs}}|x) \propto \exp\left(-\frac{1-\alpha}{2}(Hx - y^{\mathrm{obs}})^{\mathrm{T}}R^{-1}(Hx - y^{\mathrm{obs}})\right) \times \\ \exp\left(-\frac{\alpha}{2}(Hx - y^{\mathrm{obs}})^{\mathrm{T}}R^{-1}(Hx - y^{\mathrm{obs}})\right)$$

for $\alpha \in [0, 1]$ and then treating one part with the ESRF and the other one with the ETPF. For $\alpha \to 0$, the hybrid filter converges towards the ESRF and for $\alpha \to 1$ towards the ETPF. This parameter could either be a fixed value or chosen adaptively. For the adaptive choice we used the effective sample size of the ensemble and the error-vs-spread criterion. We tested the hybrid approach on the Lorenz-63 and the Lorenz-96 model and found out, that using the ETPF first, and afterwards the ESRF, yields better results. Moreover, even for small ensemble sizes, a bridging parameter $0 < \alpha < 1$ instead of $\alpha = 0$ or $\alpha = 1$, had a positive impact on the predictions and that the best choice of α increases with increasing ensemble size.

Another possibility how to choose α adaptively is by looking at the KL-divergence and the effective sample size. We showed, that there is a quadratic relation between these two explanatory variables and the optimal α, using regression analysis and a simple bimodal distribution with varying distance between the two modes and there for varying degree of Gaussianity.

We finally showed another way of combining an ETPF and a Kalman filter by approximating the prior distribution by a Gaussian mixture, the posterior by a weighted Gaussian mixture, applying the Kalman filter to each of the components of the sum separately and finally using an ETPF to produce equally weighted analysis centres for each Gaussian mixture component. This approach was again tested on the Lorenz-63 and the Lorenz-96 model. Although this method worked, the likelihood splitting was the better approach.

The second part of this thesis concerned the application of Data Assimilation to multi-scale models and related issues. Models that act on more than one scale often fulfil certain balance relations $g(z) = 0$ in the sense of staying close to a so-called slow manifold, given initial conditions on the slow manifold. Since LETFs are linear transformations, they destroy this balance in case g is non-linear. When dealing with spatially extended models, the inevitable use of localisation even violates linear balance relations. We also looked at the impact of localisation on the balance for a linear 1.5-dimensional shallow water model analytically. Some existing approaches on how to solve the issue were discussed and our own new strategy was presented. The idea was to include a post-processing step after each assimilation step in the sequential algorithm. With the analysis ensemble $\hat{z}_j$ and the assimilated imbalance values $\hat{g}_j$ at hand, we can solve the minimisation problem

$$\mathcal{J}_j(z) = \frac{1}{2}\left(\hat{z}_j - z\right)^{\mathrm{T}} A \left(\hat{z}_j - z\right) + \frac{1}{2}\left(\hat{g}_j - g(z)\right)^{\mathrm{T}} B \left(\hat{g}_j - g(z)\right)^{\mathrm{T}},$$

for $j = 1, ..., M$ and A and B being matrices that weight the uncertainty of $\hat{z}_j$ and $\hat{g}_j$ respectively. We formulated a way to solve this without even knowing the gradient of g by taking advantage of the similarity of this optimisation problem to the standard filtering problem and using the continuous Kalman-Bucy filter to solve it.

We applied this strategy to two highly-oscillatory Hamiltonian systems, namely the stiff spring double pendulum, which is chaotic and can be considered as a toy model for atmosphere and ocean dynamics, and a thermally embedded elliptic stiff spring pendulum, that has similarity to molecular dynamics simulations. We also used two spatially extended models for testing our rebalancing approach: a slow-fast extension of the Lorenz-96 model and the rotating 2-dimensional shallow water equations. The method worked remarkably well with all four systems. Imbalances were reduced and in some cases completely removed. In most of the simulations, it led to a considerable improvement of the errors, in some cases it was even essential to make the Data Assimilation algorithm work.

Bibliography

Acevedo, W., de Wiljes, J. & Reich, S. (2017), 'Second-order accurate ensemble transform particle filters', *SIAM Journal on Scientific Computing* **39**.

Alspach, D. & Sorenson, H. (1972), 'Nonlinear bayesian estimation using gaussian sum approximations', *IEEE Transactions on Automatic Control* **17**(4), 439–448.

Andersen, H. C. (1983), 'Rattle - a velocity version of the shake algorithm for molecular-dynamics calculations', *Journal of Computational Physics* **52**(1), 24–34.

Anderson, J. L. & Anderson, S. L. (1999), 'A monte carlo implementation of the nonlinear filter problem to produce ensemble assimilations and forecasts', *Monthly Weather Review* **127**, 2741–2758.

Baer, F. & Tribbia, J. (1977), 'On complete filtering of gravity modes through nonlinear initialization', *Monthly Weather Review* **105**.

Bengtsson, T., Snyder, C. & Nychka, D. (2003), 'Toward a nonlinear ensemble filter for highdimensional systems', *Journal of Geophysical Research* **108**, 8775.

Bergemann, K. & Reich, S. (2010), 'A mollified ensemble Kalman filter', *Quarterly Journal of the Royal Meteorological Society* **136**, 1636–1643.

Bergemann, K. & Reich, S. (2012), 'An ensemble Kalman-Bucy filter for continuous data assimilation', *Meteorolog. Zeitschrift* **21**, 213–219.

Bickel, P., Li, B. & Bengtsson, T. (2008), 'Sharp failure rates for the bootstrap particle filter in high dimensions', *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. Inst. Math. Stat. Collect.* **3**, 318–329.

Bishop, C., Ehterton, B. & Sharanya, M. (2001), 'Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects', *Monthly Weather Review* **129(3)**.

Bjerknes, V. (1904), 'Das problem der wettervorhersage: betrachtet vom standpunkte der mechanik und der physik', *Meteorologische Zeitrschrift* **21**.

Bloom, S., Takacs, L. L., Da Silva, A. & Ledvina, D. (1996), 'Data assimilation using incremental analysis updates', *Monthly Weather Review* **124**, 1256–1271.

Bornemann, F. A. & Schütte, C. (1997), 'Homogenization of Hamiltonian systems with a strong constraining potential', *Physica D* **102**(1-2), 57–77.

Boyd, J. P. (1994), 'The slow manifold of a five-mode model', *Journal of the Atmospheric Sciences* **51**.

Burgers, G., van Leeuwen, P. J. & Evensen, G. (1998), 'Analysis scheme in the ensemble kalman filter', *Monthly Weather Review* **126(6)**, 1719–1724.

Camassa, R. & Tin, S.-K. (1996), 'The global geometry of the slow manifold in the lorenz-krishnamurthy model', *Journal of the Atmospheric Sciences* **53**.

Cheng, Y. & Reich, S. (2015), 'Assimilating data into scientific models: An optimal coupling perspective', *Nonlinear Data Assimilation. Frontiers in Applied Dynamical Systems: Reviews and Tutorials* **2**.

Chutsagulprom, N., Reich, S. & Reinhardt, M. (2016), 'A hybrid ensemble transform filter for high-dimensional dynamical systems', *SIAM/ASA J. Uncertainty Quantification* **4-1**, 592–608.

Cohn, S., Da Silva, A., Guo, J., Sienkiewicz, M. & Lamich, D. (1998), 'Assessing the effects of data selection with the dao physicalspace statistical analysis system', *Monthly Weather Review* **126**, 2913–2926.

Cotter, C. (2013), 'Data assimilation on the exponentially accurate slow manifold', *Philosophical Transactions of the Royal Society A* **371**(1991).

Del Moral, P. (1996), 'Non linear filtering: Interacting particle solution', *Markov Processes and Related Fields* **2(4)**, 555–580.

Doucet, A. (1998), 'On sequential simulation-based methods for bayesian filtering. technical report', *Department of Engineering, University of Cambridge* .

Dovera, L. & Della Rossa, E. (2011), 'Multimodal ensemble kalman filtering using gaussian mixture models', *Computational Geosciences* **15**, 307–323.

Evensen, G. (1994), 'Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistic', *Journal of Geophysical Research: Oceans* **99(C5)**, 10143–10162.

Fowler, A. C. & Kember, G. (1996), 'The lorenz-krishnamurthy slow manifold', *Journal of the Atmospheric Sciences* **53**.

Frei, M. & Künsch, H. R. (2013), 'Bridging the ensemble kalman and particle filters', *Biometrika* **100(4)**, 781–800.

Gaspari, G. & Cohn, S. (1999), 'Construction of correlation functions in two and three dimensions', *Quarterly Journal of the Royal Meteorological Society* **125(554)**, 723–757.

Gauthier, P. & Thépaut, J. (2001), 'Impact of the digital filter as a weak constraint in the pre operational 4D-VAR assimilation system of météo-france', *Monthly Weather Review* **117**, 1225–1254.

Gill, A. (1982), 'Atmosphere-ocean dynamics', *Academic Press, New York* .

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

Gottwald, G. (2014), 'Controlling balance in an ensemble kalman filter', *Nonlinear Processes Geophysics* **21**, 417–426.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K. & Hunt, B. R. (2011), 'Balance and ensemble kalman filter localization techniques', *Monthly Weather Review* **139**, 511–522.

Hamill, T., Whitaker, J. & Snyder, C. (2001), 'Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter', *Monthly Weather Review* **129(11)**, 2776–2790.

Houtekamer, P. L. & Mitchell, H. L. (1998), 'Data assimilation using an ensemble kalman filter technique', *Monthly Weather Review* **126(3)**, 796–811.

Houtekamer, P. L. & Mitchell, H. L. (2001), 'A sequential ensemble kalman filter for atmospheric data assimilation', *Monthly Weather Review* **129**, 123–136.

Houtekamer, P. L. & Mitchell, H. L. (2005), 'Ensemble kalman filtering', *Quarterly Journal of the Royal Meteorological Society* **131**, 3269–32889.

Hunt, B., Kostelich, E. & Szunyogh, I. (2007), 'Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter', *Physica D: Nonlinear Phenomena* **230(1)**, 112–126.

Jacobs, S. J. (1991), 'Existence of a slow manifold in a model system of equations', *Journal of the Atmospheric Sciences* **48**.

Jazwinski, A. H. (1970), 'Stochastic processes and filtering theory', *Academic Press, New York* .

Kahn, H. (1950), 'Modification of the monte carlo method', *Proceedings, Seminar on Scientific Computation, November 1949* pp. 20–27.

Kahn, H. (1956), 'Controlling balance in an ensemble kalman filter', *Symposium on Monte Carlo methods, University of Florida, 1949* pp. 146–190.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Journal of Basic Engineering* **82**, 35–45.

Kalman, R. E. & Bucy, R. S. (1961), 'New results in linear filtering and prediction theory', *Journal of Basic Engineering* **83**, 95–108.

Kepert, J. (2009), 'Covariance localisation and balance in an ensemble kalman filter', *Quarterly Journal of the Royal Meteorological Society* **135**, 1157–1176.

Klainerman, S. & Majda, A. (1981), 'Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids', *Communications on Pure and Applied Mathematics* **34(4)**, 481–524.

Kreiss, H.-O. (1980), 'Problems with different time scales for partial differential equations', *Communications on Pure and Applied Mathematics* **33(3)**, 399–439.

Kullback, S. & Leibler, R. A. (1951), 'On information and sufficiency', *Annals of Mathematical Statistics* **22(1)**, 79–86.

Liu, J. & West, M. (2001), *Combined Parameter and State Estimation in Simulation-Based Filtering*, Springer New York.

Lorenc, A. C. (2003*a*), 'Modelling of error covariances by 4d-var data assimilation', *Quarterly Journal of the Royal Meteorological Society* **129**.

Lorenc, A. C. (2003*b*), 'The potential of the ensemble kalman filter for nwp: a comparison with 4dvar', *Quarterly Journal of the Royal Meteorological Society* **129**.

Lorenz, E. (1963), 'Deterministic nonperiodic flow', *Journal of the atmospheric sciences* **20(2)**, 130–141.

Lorenz, E. (1986), 'On the existence of a slow manifold', *Journal of the Atmospheric Sciences* **43**.

Lorenz, E. (1992), 'The slow manifold - what is it?', *Journal of the Atmospheric Sciences* **49**.

Lorenz, E. (1996), 'Predictability: A problem partly solved', *Proc. Seminar on predictability* **1**.

Lorenz, E. & Krishnamurthy, V. (1987), 'On the nonexistence of a slow manifold', *Journal of the Atmospheric Sciences* **44**.

Lynch, P. (2006), *The Emergence of Numerical Weather Prediction: Richardson's Dream*, Cambridge, University Press.

Lynch, P. & Huang, X.-Y. (1992), 'Initialization of the HIRLAM model using a digital filter', *Monthly Weather Review* **120**, 1019–1034.

Machenhauer, B. (1977), 'On the dynamics of gravity oscillations in a shallow water model, with applications to normal mode initialization', *Beiträge zur Physik der Atmosphäre* **50**.

Mandel, J. & Beezley, J. (2009), 'An ensemble kalman-particle predictor-corrector filter for non-gaussian data assimilation', *Computational Science - ICCS 2009* **5545**, 470–478.

Mayda, A. (2002), 'Introduction to p.d.e.'s and waves for the atmosphere and ocean', *New York University, Courant Institute of Mathematical Sciences and American Mathematical Society* .

Mitchell, H. L., Houtekamer, P. L. & Pellerin, G. (2002), 'Ensemble size, balance, and model-error representation in an ensemble kalman filter', *Monthly Weather Review* **130**, 2791–2808.

Neef, L. J., Polavarapu, S. M. & Shepherd, T. G. (2006), 'Four-dimensional data assimilation and balanced dynamics', *Journal of the Atmospheric Sciences* **63**.

Nerger, L. (2018), Ensemble data assimilation, algorithms - softwarte - applications. Department of Meteorology, University of Reading, UK.
URL: *https://epic.awi.de/id/eprint/48984/1/Nerger_EnDA_exp.pdf*

Oke, P., Sakov, P. & Corney, S. (2007), 'Impact of localization in the enkf and enoi: Experiments with a small model', *Ocean Dynamics* **57**, 32–45.

Ott, E., Hunt, B., Szunyogh, I., Zimin, A., Kostelich, E., Corrazza, M., Kalnay, E. & Yorke, J. (2004), 'A local ensemble kalman filter for atmospheric data assimilation', *Tellus A* **56**, 415–428.

Ourmières, Y., Brankart, J. M., Berline, L., Brassuer, P. & Verron, J. (2006), 'Incremental analysis update implementation into a sequential ocean data assimilation system', *Journal of Atmospheric and Oceanic Technology* **23**.

Papadakis, N., Mémin, E., Cuzol, A. & Gengembre, N. (2010), 'Data assimilation with the weighted ensemble kalman filter', *Tellus A* **62**, 673–697.

Pele, O. & Werman, M. (2009), 'Fast and robust earth mover's distances', *Computer vision, 2009 IEEE 12th international conference* pp. 460–467.

Peny, S. G. & Miyoshi, T. (2016), 'A local particle filter for high dimensional geophysical systems', *Nonlinear Processes in Geophysics* **23(6)**, 391–405.

Pham, D. T. (2001), 'Stochastic methods for sequential data assimilation in strongly non-linear systems', *Monthly Weather Review* **129(5)**, 1194–1207.

Polavarapu, S., Tanguay, M. & Fillion, L. (2000), 'Four-dimensional variational data assimilation with digital filter initialisation', *Monthly Weather Review* **128**.

Poterjoy, J. (2016), 'A localized particle filter for high-dimensional nonlinear systems', *Monthly Weather Review* **144(1)**.

Rebeschini, P. & van Handel, R. (2015), 'Can local particle filters beat the curse of dimensionality?', *Annals of Applied Probability* **25**, 318–329.

Reich, S. (1995), 'Smoothed dynamics of highly oscillatory Hamiltonian systems', *Physica D* **89**, 28—42.

Reich, S. (2000), 'Smoothed Langevin dynamics of highly oscillatory systems', *Physica D* **138**, 210–224.

Reich, S. (2013), 'A non-parametric ensemble transform method for bayesian inference', *SIAM/ASA J. Scientific Computing* **35(4)**, A2013–A2014.

Reich, S. & Cotter, C. (2015), 'Probabilistic forecasting and Bayesian data assimilation: a tutorial', *Cambridge University Press* .

Reich, S. & Hundertmark, T. (2011), 'On the use of constraints in molecular and geophysical fluid dynamics', *European Physical Journal Special Topics* **200**, 259—270.

Reinhardt, M., Hastermann, G., Klein, R. & Reich, S. (2017), 'Balanced data assimilation for highly-oscillatory mechanical systems', *arXiv:1708.03570* .

Robert, S. & Künsch, H. R. (2017), 'Localizing de ensemble kalman particle filter', *Tellus A: Dynamic Meteorology and Oceanography* **69(1)**.

Rubin, H. & Ungar, P. (1957), 'Motion under a Strong Constraining Force', *Communications on Pure and Applied Mathematics* **10**(1), 65–87.

Snyder, C., T., B., Bickel, P. & Anderson, J. (2008), 'Obstacles to high- dimensional particle filtering', *Monthly Weather Review* **136**.

Snyder, C., T., B. & Morzfeld, M. (2015), 'Performance bounds for particle filters using the optimal proposal', *Monthly Weather Review* **143**.

Staniforth, A., Wood, N. & Reich, S. (2006), 'A time-staggered semi-lagrangian discretisation the the rotating shallow-water equations', *Quarterly Journal of the Royal Meteorological Society* **132(621C)**, 3107–3116.

Stephenson, D. & Doblas-Reyes, F. (2000), 'Statistical methods for interpreting monte carlo ensemble forecasts', *Tellus A* **52(3)**, 300–322.

Sun, A. Y., Morris, A. P. & Mohanty, S. (2009), 'Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques', *Water Resources Research* **45**, W07424.

Swerling, P. A. (1958), 'A proposed stagewise differential correction procedure for satellite tracking and prediction', *Santa Monica, Calif.: Rand* .

Szunyogh, I., Kostelich, E., Gyarmat, G., Patil, D. J., Hunt, B., Kalnay, E., Ott, E. & Yorke, J. (2005), 'Assessing a local ensemble kalman filter: Perfect model experiments with the national centers for environmental prediction global model', *Tellus A* **57**, 528–545.

Takens, F. (1980), *Motion under the influence of a strong constraining force*, Springer Berlin Heidelberg.

Thépaut, J. N. & Courtier, P. (1991), 'Four-dimensional variational data assimilation using the adjoint of a multilevel primitive equation model', *Quarterly Journal of the Royal Meteorological Society* **129**.

Tippett, M., Anderson, J., Bishop, C., Hamill, T. & Whitaker, J. (2003), 'Ensemble square root filters*', *Monthly Weather Review* **131(7)**, 1485–1490.

Tödter, J. & Ahrens, B. (2015), 'A second-order exact ensemble square root filter for nonlinear data assimilation', *Monthly Weather Review* **143(6)**, 1347–1367.

van Leeuwen, P. J. (2003), 'Nonlinear ensemble data assimilation for the ocean', *Recent Developments in data assimilation for atmosphere and ocean* .

van Leeuwen, P. J. (2017), 'Particle filters for nonlinear data assimilation in high-dimensional systems', *Annales de la faulté des sciences de Toulouse Mathématiques* **26(4)**.

Watkinson, L. R., Lawless, A. S., Nichlos, N. K. & Roulstone, I. (2007), 'Weak constraints in four-dimensional variational data assimilation', *Meteorologische Zeitschrift* **16**.

Wirosoetisno, D. (2004), 'Exponentially accurate balance dynamics', *Adv. Diff. Eq.* **9**, 177–196.

Zeitlin, V. (2007), *Introduction: Fundamentals of rotating shallow water model in the geophysical fluid dynamics perspective*, Elsevier, chapter 2.