Department of Mathematics and Statistics

Localised adaptive Particle Filters for large scale operational NWP model

Anne Sophie Rojahn

A thesis submitted for the degree of

Doctor of Philosophy

April 2024

Abstract

Data assimilation is the process of combining observations with a model's first guess to determine the best estimate of the current state of the atmosphere. In the case of numerical weather prediction the data assimilation step is carried out using various types of Kalman filter equations, which are developed based on the Gaussian assumption and are therefore unable to deal with non-Gaussianity. Particle filters, by definition, can deal with non-Gaussianity and are well known in statistics. They have a long tradition in the framework of ensemble data assimilation (EDA) as well as Markov-Chain Monte Carlo (MCMC) methods. A key challenge today is to employ such methods in a high-dimensional environment, since the naive application of the classical particle filter usually leads to filter divergence or filter collapse when applied within the very high dimensionality of many practical assimilation problems (known as the curse of dimensionality).

In the first part of this thesis we introduce the Localised Adaptive Particle Filter (LAPF), which follows closely the idea of the classical MCMC or bootstrap type particle filter, but overcomes the problems of filter collapse and divergence using localisation in the sense of the Local Ensemble Transformed Kalman Filter (LETKF) and adaptivity with an adaptive Gaussian resampling or rejuvenation scheme in ensemble space. We have implemented the particle filter in the data assimilation system for the global forecast model ICON at the German Meteorological Service (DWD). We carry out simulations over a period of one month with a global horizontal resolution of 52 km and 90 vertical layers. With four variables analysed per grid point, this results in $6.6 \cdot 10^6$ degrees of freedom. The LAPF can be run stably and shows a reasonable performance. We compare its results with the operational LETKF implementation of DWD for the ICON model.

Based on this work, we investigate the implementation of the Gaussian uncertainty of individual particles in the assimilation step of the localised adaptive particle filter. We obtain a local representation of the prior distribution as a mixture of basis functions. In the assimilation step, the filter calculates the individual weight coefficients and new particle locations. It can be thought of as a combination of the LAPF and a localised version of a Gaussian mixture filter, i.e., a Localised Mixture Coefficients Particle Filter (LMCPF).

Again, we have implemented the LMCPF within the global operational framework ICON of the DWD and evaluate the relationship between prior and posterior distributions and observations. Our simulations are carried out in the same standard pre-operational experimental setup as for the LAPF and additionally in a setup closer to the currently operational system. We are able to show that the mixture approach is able to deal with the discrepancy between the prior distribution and the observation location in ensemble space in a real-world framework and to pull the particles towards the observations. This shows that the use of Gaussian uncertainty can be an important tool to improve the analysis and forecast quality in a particle filter framework.

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Anne Sophie Rojahn

Publications

The content of chapter 3 has been published in

- Potthast, R., Walter, A., and Rhodin, A. A localized adaptive particle filter within an operational nwp framework. *Monthly Weather Review*, 147 (1):345–362, 2019. doi: 10.1175/MWR-D-18-0028.1

My contribution to this paper has been to implement the LAPF code in the operational data assimilation system of DWD. I have done all experiments and evaluations. Writing the paper has been done by Prof. Dr. Roland Potthast and me with same amount.

The content of chapter 4 has been accepted for publication in

- Rojahn, A., Schenk, N., van Leeuwen, P. J., and Potthast, R. Particle filtering and gaussian mixtures - on a localized mixture coefficients particle filter (lmcpf) for global nwp. *Journal of the Meteorological Society of Japan. Ser. II*, advpub:2023–015, 2023. doi: 10.2151/jmsj.2023-015.

In this work, I again did the code work, performed the experiments and the evaluations. I led the writing of the manuscript.

Further work related to this PhD thesis but not explicitly mentioned has been published in

- Schenk, N., Potthast, R., and Rojahn, A. On two localized particle filter methods for lorenz 1963 and 1996 models. *Frontiers in Applied Mathematics and Statistics*, 8, 2022. ISSN 2297-4687. doi: 10.3389/fams.2022.920186.

Acknowledgements

First of all I would like to thank my first supervisor Prof. Dr. Roland Potthast. During our first meeting, even before you really got to know me, you offered me the opportunity to write a PhD thesis. Over the past seven years, you have always had an open ear for me. Thank you for always taking time for me when there was a gap in your appointment calendar. This thesis wouldn't exist without you!
I would also like thank Prof. Dr. Peter Jan van Leeuwen for taking time during my stays in Reading. You explained the world of data assimilation in a way that a meteorologist like me could understand. Thank you very much for that!

Since I've been working at DWD my colleagues from FE12 and FE11 have always had time to answer my questions, thanks a lot for that Harald, Andreas, Christina, Robin, Elisabeth and Steffi. Robin and Christina, I thank you so much for your patience when I didn't have time to take care of the satellites.
Lilo thank you for all the talks about our work and all the other topics, thank you for being my friend!
Gernot thanks a lot for all the conversations about writing dissertations! Thanks for the Threema chats, they often made my day ;-).

Last but not least, I would like to thank my family. Mum and Dad, I literally wouldn't be here without you ;-P. Mum, you told me everything I know, you made me the person I am today. Thank you so much for always being there!
Patrick, I know, it's not easy with me, but I thank you for all your understanding, for taking the kids so that I had the time to write down this thesis. I love you so much!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Numerical Weather Prediction (NWP) models determine the evolution of physical processes in the atmosphere and at the surface to obtain weather forecasts by solving mathematical equations based on physical laws like the conservation of energy. As the atmosphere is a high dimensional, chaotic (Lorenz, 1963, 1995) and non-linear system, a huge amount of observational data is needed to describe its current state. Since this is not given, NWP is an initial value problem as the initial weather conditions at each point on earth are unknown. Therefore, NWP models run an ensemble of forecasts with slightly perturbed initial conditions (Kalnay (2002), Kalnay et al. (2006)) to estimate the uncertainty in the meteorological variables.

Data assimilation is the process of determining the initial states for NWP from observations of the Atmosphere and the Earth's surface. Therefore, initial conditions (i.e. observations of the atmosphere like temperature, pressure and wind) with a high accuracy are needed. To provide these high-quality measurements and observations a combination of different sources is taken into account, including data from direct observations (e.g. measurements provided by radiosondes, aircrafts and buoys), as well as indirect observations (e.g. satellite data and

radar data). Data assimilation then determines an integrated full state of the Earth's surface and the Atmosphere, taking into account the measurements and the calculated model states. For the global NWP model ICON (see section 1.3) a hybrid Ensemble Variational (EnVar) system is used. It consists of a Local Ensemble Transform Kalman Filter (LETKF - based on Hunt et al. (2007)), combined with a three-dimensional variational analysis (3d-Var).

A lot of research into all of these parts of NWP (the numerical model, data assimilation and the use of more observations) has improved the accuracy of forecasts enormously (see Fig. 1.1). 50 years ago the 1-day forecast has been as good as the 7-day forecast is nowadays. In Figure 1.1 the verification of the tendency correlation coefficient[1] of the geopotential height[2] at 500 hPa is shown, and thus the development of the accuracy of the NWP at DWD.

Figure 1.1 shows the tendency correlation coefficient of the geopotential height at 500 hPa forecasts from the global models used at DWD from 1968 to 2023 for the North Atlantic region and Central Europe. The blue shading indicates the changes in the numerical models (from BKL to ICON) and the coloured lines show the forecast-times (from 24 h to 168 h). It can clearly be seen that a change in the model and the resolution leads to a large improvement in all forecast times (e.g. the large increase in 1991 when the GM became operational or in 2004 when the GME was introduced) but the improvement after a change in the data assimilation system is even higher. The latest shown model change has been the update of the resolution. Since 2015, the ICON model with a horizontal resolution of 13 km for the deterministic run is used , causing a large improvement which can be seen as an increase in the tendency correlation coefficient. The last

[1]The correlation coefficient ist given by: $CC = \frac{\overline{(F-R-\overline{(F-R)})(A-R-\overline{(A-R)})}}{\sqrt{\overline{F-R-\overline{(F-R)})^2} \cdot \overline{A-R-\overline{(A-R)})^2}}}$ with R is the reference forecasts. To calculate the tendency correlation coefficient, the long-range dependence has to be taken as reference forecast. (Damrath, 2002)

[2]The geopotential height is a measure of the height of a point in the atmosphere in relation to its potential energy. This parameter plays an important role in synoptic meteorology (analysis of weather patterns). Charts of geopotential height plotted at constant pressure levels can be used to identify weather systems such as cyclones, anticyclones, troughs and ridges. (European Centre for Medium-Range Weather Forecasts, 2019)

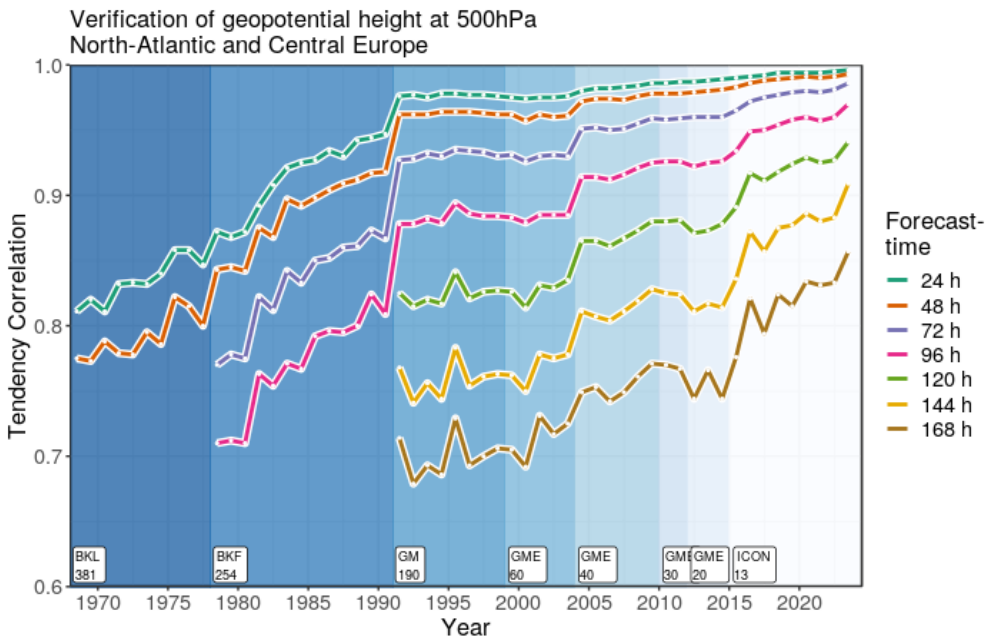

Figure 1.1: Tendency correlation coefficient of the geopotential height at 500 hPa forecasts of the global models used at DWD and corresponding analysis on a 1.5°x1.5° grid over Cenral Europe and the North Atlantic. The forecast period ranges from 24 to 168 hours in steps of 24 hours, as indicated by the colour of the lines. Operational models are indicated by the coloured background, model name and horizontal resolution (km) are indicated by labels. The tendency correlation is the correlation of anomalies, forecasts and analysis were corrected by the geopotential height at forecast initialisation. Yearly correlations were obtained by arithmetic averaging of daily correlations. (Source: Image created by DWD, FE1, Verification group, accessed 01.02.2023)

improvement in the horizontal resolution has been the update on 26 km for the ensemble runs for the ICON model which has been operational since January 2023. An even larger improvement than that of the introduction of ICON can be seen in 2016 were the new data assimilation system has been made operational. Since 2016 the DWD is using an Ensemble Variational (EnVar) Data Assimilation System, a combination of the LETKF for the 40 member ensemble and a 3d-Var for the high-resolution deterministic run. Before 2016, the DWD didn't run an ensemble itself, so this really high improvement is due to use the combination of the ensemble with the LETKF and a 3d-Var for the deterministic run.

This improvement is also visible in Figure 1.2 where the World Meteorological Organization (WMO) verification against observations for the root mean square error (RMSE) of the temperature at 850 hPa for a lead time of 24 h is shown.

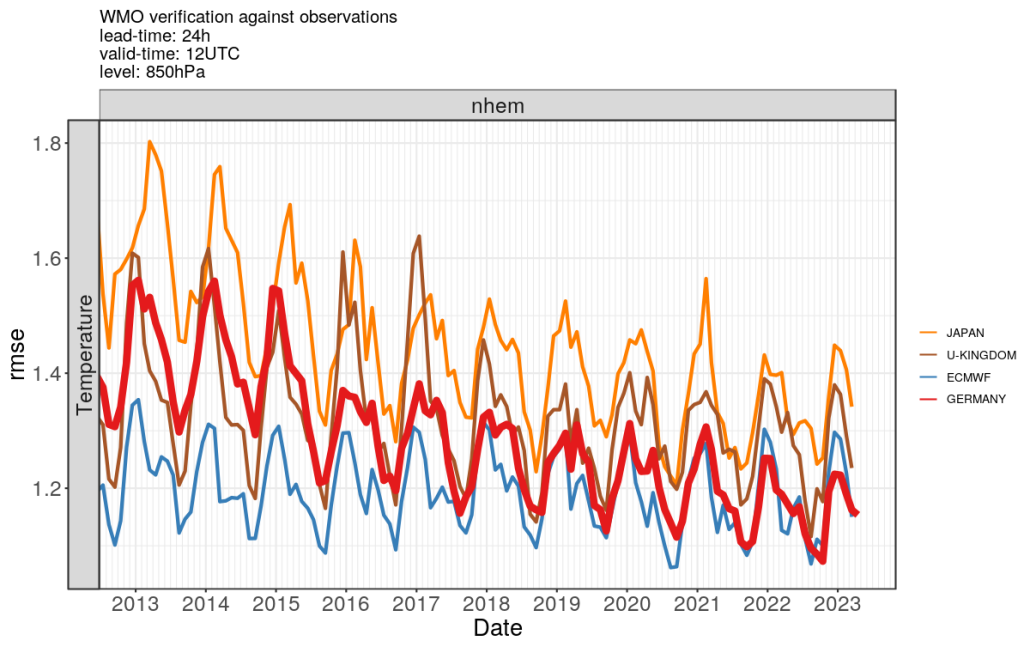

Figure 1.2: WMO verification against observations. Shown is the the root mean square error of the temperature at 850 hPa for 24 h lead time for four different NWP centres as timeline for the last ten years. The line colour indicate the country - Japan orange, Met Office (UK) brown, ECMWF blue and DWD (Germany) red. (Source: Image created by DWD, FE1, Verification group, accessed 01.02.2023)

Here, the colours are indicating four different operational NWP centres (orange for the Japan Meteorological Agency, brown for the United Kingdom Met Office, blue for the ECMWF and red for the DWD). It can clearly be seen that the RMSE of the DWD is getting better (lower) since 2013, in 2016 the same very large improvement can be seen as in Figure 1.1, when the Ensemble Data Assimilation (EDA) was made operational.

The RMSEs in the extratropics are higher in the winter months than in the summer because the baroclinicity of the atmosphere is higher and therefore the low pressure areas are more intense than in the summer. This means that the potential for errors is greater (e.g. due to double-penalty effects), which is further accentuated by quadratic error measures. Note that RMSE and predictability are not the same thing. There is no simple, universally valid measure of predictability, but the added value of a prediction compared to the climate mean or persistence is usually considered. In the extratropics, for example, the anomaly

correlation is used, which is normalised by the amplitude of the deviations and thus largely ignores amplitude-related double-penalty effects. The anomaly correlation is usually better in winter than in summer. (Zängl, 2024)
Since the first launch of meteorological satellite in the 1960s, they have provided the largest amount of observations worldwide and thus have a really big influence on the accuracy of the data assimilation (assimilated at ECMWF since the late 1990's) and thus on the forecasts. The assimilation of these data is very important (Eyre et al., 2022), especially in regions with less conventional data. Because of the non-Gaussianity of the error (Geer and Bauer, 2011) of this large contributor to the observations, the aim of this thesis is to develop a particle filter that has the potential to deal with non-Gaussianity.

1.2 Data Assimilation

Data assimilation is concerned with the use of observational data to determine the state of a dynamical system. In the framework of weather forecasting, data assimilation has a long history, ranging from early work by Bjerknes (1904) and Richardson (1922) to modern ensemble data assimilation systems (see Bauer et al. (2015)). In NWP, data assimilation links the model world to reality by using a wide range of observations to determine the evolution of the atmosphere and to provide initial conditions for weather forecasting.
The history of data assimilation methods which are used in an operational NWP framework began with optimal interpolation from the 1960's to the 1990's. Variational methods such as three-dimensional variational assimilation (3d-Var) have been used operationally since about 1990, with four-dimensional variational assimilation (4d-Var) since the late 1990's (operational at ECMWF since 1997). Variational methods have the strength to calculate a best estimate either for one time slice (3d-Var) or over some temporal window (4d-Var). The development of ensemble data assimilation started in the mid 1990s and has been in operational use since about 2010.

The Ensemble Kalman Filter was developed by Evensen (1994), see also Evensen and van Leeuwen (2000) and Evensen (2009). The idea has been applied to global numerical weather prediction by Houtekamer and Mitchell (1998, 2001, 2005) and Houtekamer et al. (2005). In the area of geophysical data assimilation, Burgers et al. (1998) developed the theoretical basis of the EnKF methods based on perturbations of the observations. Whitaker and Hamill (2002) proposed the alternative approach called ensemble square root filter (EnSRF). It does not use randomly perturbed observations, but formulates a deterministic calculation of the posterior ensemble.

Further variants of ensemble filters include for example the Singular Evolutive Extended Kalman Filter (SEEKF) of Pham et al. (1998), the Ensemble Adjustment Kalman Filter (EAKF) of Anderson (2001) and the Ensemble Transform Kalman Filter (ETKF) of Bishop et al. (2001). Localisation is a key ingredient of the Ensemble Kalman Filter of Houtekamer and Mitchell (1998) (denoted as data selection with a cut-off radius), the work of Brusdal et al. (2003), the Local Ensemble Kalman Filter (LEKF) of Ott et al. (2004) and the Local Ensemble Transform Kalman Filter (LETKF) of Hunt et al. (2007), where all locally available observations are assimilated in one step (locally in this context means that only observations close to the analysis grid-point are used). Various other forms of filters have been developed, see for example the GIGG Filter by Bishop (2016). For an overview of ensemble-based data assimilation methods we refer to Vetra-Carvalho et al. (2018).

Important current research topics on ensemble Kalman filters are covariance localisation and inflation, see van Leeuwen (2003a), Miyoshi et al. (2007), Miyoshi and Sato (2007), Campbell et al. (2010), Greybush et al. (2011), Janjić et al. (2011) and Periáñez et al. (2014). Flow-adaptive localisation has been described by Bishop and Hodyss (2007, 2009a,b) and Anderson (2007); multiscale localisation by Miyoshi and Kondo (2013); flow-adaptive inflation by Anderson (2007, 2009), Li et al. (2009) and Miyoshi (2011). The investigation of large ensembles

has been carried out for example by Miyoshi et al. (2014). However, these methods rely on the approximation of Gaussianity, which is a strong limitation when applied to highly non-linear dynamical systems such as global or high-resolution Numerical Weather Prediction (NWP).

Leaving the Gaussian regime, for which ensemble Kalman filters are best, particle filters take into account the full non-linearity of both the model dynamics and the observation operators in applications, leading to strongly non-Gaussian distributions on all temporal and spatial scales. Particle filters have a long history in stochastic modelling, where they have been used since the 1960s under the name of iterative Markov Chain Monte Carlo (MCMC) methods (Bain and Crisan (2009); Crisan and Rozovskii (2011)). The idea is to sample some probability distribution, where the number of samples reflects the local strength of the probability density, their weights are adapted using observations and then resampling is carried out. Several particle filter methods have been formulated and tested for small-dimensional problems, ranging from early work by Gordon et al. (1993) to the review by van Leeuwen (2009).

It is well-known that in a high-dimensional framework, standard particle filters suffer from so-called filter collapse or filter divergence under the curse of dimensionality. This means that usually only very few or only one of the ensemble members carry all the weight in the assimilation step. This immediately destroys the diversity of the ensemble and leads to useless behaviour when applied in an iterative way, compare van Leeuwen (2010); Snyder et al. (2008, 2015) or Bickel et al. (2008). Different ideas have been developed to overcome filter collapse, for example guiding the particles to the right places in the high-dimensional space as in van Leeuwen (2010). Using localisation for particle filters has become popular, first introduced by Bengtsson et al. (2003); van Leeuwen (2003b,a), see also for example Reich and Cotter (2015); Poterjoy and Anderson (2016) or Penny and Miyoshi (2016). Frei and Künsch (2013) have developed a hybrid Ensemble Kalman Particle Filter which Robert et al. (2017) have tested for the regional

NWP model COSMO. To avoid filter divergence, they combine the particle filter step with an ensemble Kalman filter step. Poterjoy (2016) works in a convective-scale framework in comparison to the global setup, he also uses localisation as a core component of his Localized Particle Filter. Poterjoy et al. (2019) have successfully implemented the Localized Particle Filter (LPF) of Poterjoy (2016) in the convective-scale ensemble system (NEWS-e), Poterjoy et al. (2017) implemented it in the Weather Research and Forecasting (WRF) model. Ades and van Leeuwen (2013, 2015) have adapted their Equivalent-Weights Particle Filter (EWPF) to high-dimensional systems using a simple relaxation technique and a proposal to ensure equal weights for the particles (see also Browne and van Leeuwen (2015) who used the EWPF in the full scale coupled ocean atmosphere circulation model HadCM3). Zhu et al. (2016) have further improved the EWPF to their Implicit Equal-Weights Particle Filter (IEWPF). Wang et al. (2020) have further developed the IEWPF to the Implicit Equal-Weights Variational Particle Smoother (varPS) - a combination of IEWPF and 4d-Var, that uses implicit sampling to avoid filter divergence and have tested it in the Lorenz96 model. Pinheiro et al. (2019) improves the IEWPF by using an efficient ensemble based synchronisation scheme as the proposal density function and have also tested it in the Lorenz96 model.

Instead of localisation Kawabata and Ueno (2020) have used an adaptive observation error estimator to avoid the filter collapse in a regional mesoscale model. Hu and van Leeuwen (2021) implemented the Particle Flow Filter (PFF) in the Lorenz96 model. The PFF overcomes the filter collapse by keeping the weights of all particles equal at all times, by transforming the particles from the prior to the posterior state space without the need to resample the weights.

Further, Gaussian Mixture models have been developed, based on the estimation of the model error probability distribution by a set of Gaussians, see Hoteit et al. (2008) or Stordal et al. (2011) for a hybrid method of a Gaussian Mixture and the particle filter. Frei and Künsch (2013) develop a hybrid method for an EnKF and

a particle filter. Kotsuki et al. (2022) have implemented the LPF of Penny and Miyoshi (2016) and its Gaussian Mixture extension (LPFGM) to the Simplified Parametrisations, Primitive Equation Dynamics (SPEEDY) model.

For recent reviews we refer to van Leeuwen (2009); Reich and Cotter (2015); Farchi and Bocquet (2018); Vetra-Carvalho et al. (2018); Carrassi et al. (2018) and van Leeuwen et al. (2019), see also Nakamura and Potthast (2015) and Evensen et al. (2022).

1.3 The ICON Model and Deterministic Forecast System

The ICOsahedral Nonhydrostatic (ICON) model is the operational global numerical weather prediction (NWP) model of the German Meteorological Service (DWD). It is a joint project of DWD and the Max-Planck-Institute for Meteorology (MPI-M) (Zängl et al. (2014), see also Reinert et al. (2018) for a complete documentation of the ICON model or the official website of the ICON model (the ICON partnership , MPI-M et al.(2024)). ICON is based on the prognostic variables suggested by Gassmann and Herzog (2008) (for further information see also Reinert et al. (2018)). But instead of using the three-dimensional Lamb transformation it uses the two-dimensional version to convert the non-linear momentum advection into a vector-invariant form (Zängl et al. (2014)).

The model grid is based on an unstructured triangular grid that is generated by successive refinement of a spherical icosahedron, which consists of 20 equilateral triangles with an edge length of about 7054 km. These triangles are subdivided (e.g., by bisection, trisection) into smaller triangles, leading to a model grid with the desired spatial resolution. The use of this icosahedral grid provides a nearly homogeneous coverage of the globe. After dividing the triangles the operational grid of the ICON model consists of 2.949.120 triangles on each horizontal level. Each triangle has an average area of 173 km^2. This corresponds to the global

horizontal resolution of 13 km for the deterministic run. For the operational setup a two-way nested area over Europe with 6.5 km horizontal resolution is included. Since November 23, 2022 the ensembles run with an operational horizontal resolution of 26 km with a 13 km resolution nest over Europe.

The scalar prognostic variables (e.g., temperature, humidity) are located in the centre of the triangles whereas the wind components are located at the edge midpoints of the triangles. The most important prognostic variables (e.g., wind, humidity, cloud water, cloud ice, temperature, snow, precipitation) are calculated for all grid-cells on 120 terrain-following vertical model levels which range from the surface up to a height of 75 km, leading to over 353 million grid points in the operational setup. Additional prognostic equations are solved over land on eight soil levels for soil temperature and soil water content. If snow is present, several snow variables are also determined. Once per day (at 00 UTC) the sea surface temperature (just over ice-free ocean) is analysed from observations and is kept constant during the forecasts. For the sea ice fraction of the ice-covered oceans we proceed in the same way. However, ice thickness and ice surface temperature are determined by a simple sea-ice model.

Beyond the adiabatic processes in the atmosphere (horizontal and vertical transport processes), diabatic processes (e.g., radiation, turbulence) play a major role in NWP. Describing these small scale processes is part of the physics parametrisation of ICON.

Within the operational workflow, we distinguish the data assimilation cycle and the forecast mode. During the data assimilation cycle a three-hour-forecast starting from the previous analyses (the first guess) is blended with all observations valid for a three-hour time window centred at the analysis date. It is an important detail to note that traditionally the atmospheric analysis calculates increments to four core variables (temperature, humidity, two wind components) at each grid point in its three-dimensional grid, with pressure adapted according to the hydrostatic equation. The model adapts further prognostic variables itself. Ad-

ditional two-dimensional fields which are also adapted have been neglected in our variable counts. To obtain an optimal initial state for the subsequent forecasts we calculate a variational analysis with a dynamic covariance matrix, known as Ensemble Variational Data Assimilation (EnVar). 70% of the covariance matrix is calculated from the ensemble runs based on a Local Ensemble Transform Kalman Filter (LETKF). Further 30% of the covariance matrix are given by its climatological part based on the NMC method (Parrish and Derber (1992)). The EnVar and the LETKF were made operational on January 20, 2016.

Based on the analyses for 00 and 12 UTC ICON provides a 180 hours forecast in just one hour wall clock time. Forecasts over 120 hours are based on the analyses of the 06 and 18 UTC run and the 30 hours forecasts are based on the 03, 09, 15 and 21 UTC analyses.

1.3.1 The ICON Ensemble Prediction System

The ICON Ensemble Prediction System (ICON-EPS) has run pre-operationally since January 2016 providing background error correlations for the operational global EnVar system of DWD, with operational forecasts including all ensemble products since January 17, 2018.

The ICON-EPS initial conditions are provided by the LETKF, which is part of the hybrid data assimilation suite LETKF+EnVar for the ICON model. The ICON-EPS is run up to 180 hours at 00 and 12 UTC, up to 120 hours at 06 and 18 UTC and up to 30 hours at the three hourly intervals in between with the main purpose to generate forecasts and ensemble boundary conditions in the short range of up to 30 hours lead time every three hours.

1.4 Objectives

In the operational NWP centres Kalman filters in various ways are used to calculate the data assimilation step (e.g. the DWD uses a LETKF, the ECMWF

uses a Reduced Rank Kalman Filter (RRKF) within their 4d-Var framework). The Kalman filter equations are developed based on the Gaussian assumption and therefore aren't designed for dealing with non-Gaussianity. Classical particle filters are able to deal with non-Gaussianity, but in high-dimensional setups they suffer from filter divergence under the curse of dimensionality (van Leeuwen, 2010), and so they aren't able to run stable over a long period of time and thus can't be used for weather prediction. Therefore, the aim of this thesis is to develop particle filters that do not suffer from degeneration and therefore run stable over a longer period of time. For this aim, two different localised particle filters are formulated and tested. They will be compared with the operational LETKF implementation of the DWD in the ICON model.

The first particle filter is the Localised Adaptive Particle Filter (LAPF). It is the first particle filter implemented in a global high dimensional and large-scale operational NWP model. We will show that, with appropriate adaptation, as e.g. localisation, ensemble transformation (based on LETKF), the particle filter can be a stable and useful approach in a large-scale operational framework.

The LAPF is per construction not able to pull the particle towards the observations, as it just chooses the nearest particles to the observations and than draw around them with a determined spread. So that the LAPF isn't able to handle biases or even correct them.

The second particle filter shall be able to pull the chosen particles towards the observations to receive better results. The LMCPF overcomes this problem by using Gaussian Mixtures. The Mixtures approach is able to deal with the discrepancy between prior distribution and observation location and thus is able to pull the particles towards the observations. By giving each particle a certain bias, the LMCPF is able to pull them closer to the observations. The smaller the number of particles selected, the more important is the pulling towards the observations steps.

1.5 Thesis structure

This thesis is divided into 6 chapters. Chapter 2 gives a brief introduction into Bayes' Theorem and the LETKF (Hunt et al. (2007)) implementation at DWD. In Chapter 3 we describe the Localised Adaptive Particle Filter implemented in the operational NWP model ICON of DWD. This work has been published in Potthast et al. (2019). Chapter 4 deals with Particle Filtering with Model Error - a Localised Mixture Coefficients Particle Filter (LMCPF), where we develop a particle filter that overcomes the deficiencies of the LAPF. The results presented in this chapter have been accepted for publication in Rojahn et al. (2023). In Chapter 5, we discuss an interpretation of the LMCPF as Implicit Equal-Weights Particle Filter (IEWPF by Zhu et al. (2016)). In the last Chapter of this Thesis, Chapter 6, we give a general conclusion and discuss some suggestions how to extend the work at particle filters.

Chapter 2

Background Methodology

This chapter gives an overview of the data assimilation methods which are used as the basis for the particle filters developed in this thesis.

As mentioned in the introduction of this thesis, we need a data assimilation step to determine initial conditions for the weather forecast given some observations using a NWP model. In our case this means, that we have determined the prior state of the atmosphere, given by a prior distribution, by taking the analysis of the data assimilation step before. To receive the analysis - the posterior distribution - we have to use Bayes' Theorem and update the prior distribution. Therefore, Bayes' Theorem is explained in section 2.1.

The Localised Ensemble Transform Kalman Filter (LETKF) of Hunt et al. (2007) is used as the basis and reference for both particle filters developed in this thesis. Its implementation in the data assimilation system of the DWD is explained in section 2.2.

2.1 Bayes' Theorem in Data Assimilation

In this section a brief overview of Bayes' Theorem is given.

The basic idea of a Bayesian assimilation step is to calculate a posterior distribution based on a prior distribution, model states and some given observations.

The Bayes' formula is given by

$$p(x|y) = \frac{p(x) \cdot p(y|x)}{p(y)}. \tag{2.1.1}$$

The prior distribution $p(x)$

The prior distribution $p(x)$ is the a priori knowledge of the model states x.
In the case of operational numerical weather prediction (NWP) the prior distribution $p(x)$ is obtained from a model forecast. This means that the starting point of the data assimilation process is the so called 'analysis' distribution of the data assimilation cycle one time step before.

The data distribution $p(y|x)$

The data distribution $p(y|x)$ gives the probability of the observations y given the model states x. It is also known as the conditional likelihood.
For example, in the case of NWP: x are the model states of the temperature (i.e. the determined temperatures) and y are the true, observed observations.

The distribution of observations $p(y)$

$p(y)$ is the probability distribution of the observations y.

The posterior distribution $p(x|y)$

The posterior distribution $p(x|y)$ gives the probability of the model states x given the observations y. It is also known as the conditional probability and it is proportional to the product of the prior $p(x)$ and the data distribution $p(y|x)$.

Bayes' Theorem can be derived from the definition of conditional probability (for further details see for example Nakamura and Potthast (2015), Section 4.2). The conditional probability of x under the condition y is given by

$$p(x, y) = p(x|y) \cdot p(y). \tag{2.1.2}$$

Taking Equation 2.1.2 twice and swapping the roles of x and y leads to

$$p(x, y) = p(x|y) \cdot p(y) = p(y|x) \cdot p(x), \tag{2.1.3}$$

which leads directly to the Bayes' formula:

$$p(x|y) = \frac{p(x) \cdot p(y|x)}{p(y)}. \tag{2.1.4}$$

In our case of numerical weather prediction, Bayes' theorem is applied as follows. Let us recall the stochastic notation where $x_k \in \mathbb{R}^n$ denote model states of dimension n and $y_k \in \mathbb{R}^m$ is the vector of m observations at time t_k. Bayesian data assimilation starts with some prior distribution. The analysis step employs observations to derive an analysis distribution. This analysis distribution is propagated to the next time step, where it acts as a prior distribution for the subsequent assimilation step. We consider the analysis distribution $p^{(a)}(x)$ for time t_0 as the initial state. Then the analysis distribution is propagated in time by a short range ensemble forecast, from t_{k-1} to t_k to obtain the first guess distribution $p_k^{(b)}(x)$. Afterwards, Bayes' formula $p(x|y_k) = \frac{p_k^{(b)}(x) \cdot p(y_k|x)}{p(y_k)},\ x \in \mathbb{R}^n,\ y_k \in \mathbb{R}^m$ is employed to calculate the new analysis distribution

$$p_k^{(a)}(x) := p(x|y_k) = cp(y_k|x)p_k^{(b)}(x), \quad x \in \mathbb{R}^n, \tag{2.1.5}$$

where c is a normalisation constant so that

$$\int_{\mathbb{R}^n} p_k^{(a)}(x)\, dx = 1. \tag{2.1.6}$$

2.2 The operational Localised Ensemble Transform Kalman Filter (LETKF) implementation of DWD

This section explains the operational LETKF implementation of DWD and all its inherent parts, e.g. observation handling and localisation. The LETKF serves as the reference for all experiments with particle filters.

2.2.1 Observations

Observation operators are evaluated at three hourly intervals in the cycled data assimilation code. Observation types currently include profiles of e.g. temperature and humidity measured by weather balloons (TEMP and PILOT), ships (SHIP), classical land-based weather stations (SYNOP), buoyes and drifting buoyes (DRIBU), wind profiler, aircraft (AIREP), atmospheric motion vectors (SATOB), radio occulations (GPSRO), scatterometers (SCATT) and satellite radiances (RAD). Observational quality control (observation minus first guess check) and bias correction (for radiances and individual aircrafts) is performed in the deterministic data assimilation system. Bias corrected observations and quality control flags are then further passed to the ensemble data assimilation system.

Obs.type	processed	dismissed	passive	rejected	active
SYNOP	108480	418	88447	588	19027
AIREP	267469	214216	294	4226	48733
SATOB	598548	554036	1200	2690	40622
DRIBU	1663	515	45	20	1083
TEMP	113696	0	83443	2067	28186
PILOT	31968	19672	8498	492	3306
SCATT	413105	313788	17911	18262	63144
RAD	58654581	42741985	15464187	298628	149781
GPSRO	20178	0	1224	383	18571
total	60209688	43844630	15665249	327356	372453

Table 2.1: Shown are statistics of the use of different observation types. Example for 31.05.2016 0 UTC - taken from the first LETKF reference experiment

Table 2.1 shows some statistics on the use of the different observation types. The statistics are determined for the reference experiment of 2016 with which the LAPF and LMCPF are compared. It is shown how many measurements per observation type are processed and how many are actively assimilated at the end of the quality control. The following observations conditions are shown:

processed	all observation going into the data assimilation system
dismissed	bad observations which can't be used, e.g. observations with missing or corrupted values
passive	good observations which passed the quality control but aren't assimilated actively (e.g. because of thinning), but they are further used, e.g. for bias correction or cloud detection
rejected	observations which didn't pass the quality control, e.g. satellite data which are cloudy
active	good observations which are actively assimilated

In the case of this example of our reference experiment an amount of 91568 observations are actively assimilated. As the ICON model and data assimilation system as well as the number of considered observation types have changed during the last 7 years, the current operational ICON model uses a much larger number of observations (see Table 2.2).

In Figure 2.1 an example for the coverage of the used radiance observed by satellite instruments for one assimilation step (31.05.2016 0 UTC) is shown.

The figure shows the active assimilated observations (brightness temperatures

Obs.type	processed	dismissed	passive	rejected	active
total	92637101	70131543	20336098	1387224	782236

Table 2.2: Shown are statistics of the number of all observations for the currently operational ICON model for 24.04.2023 0 UTC.

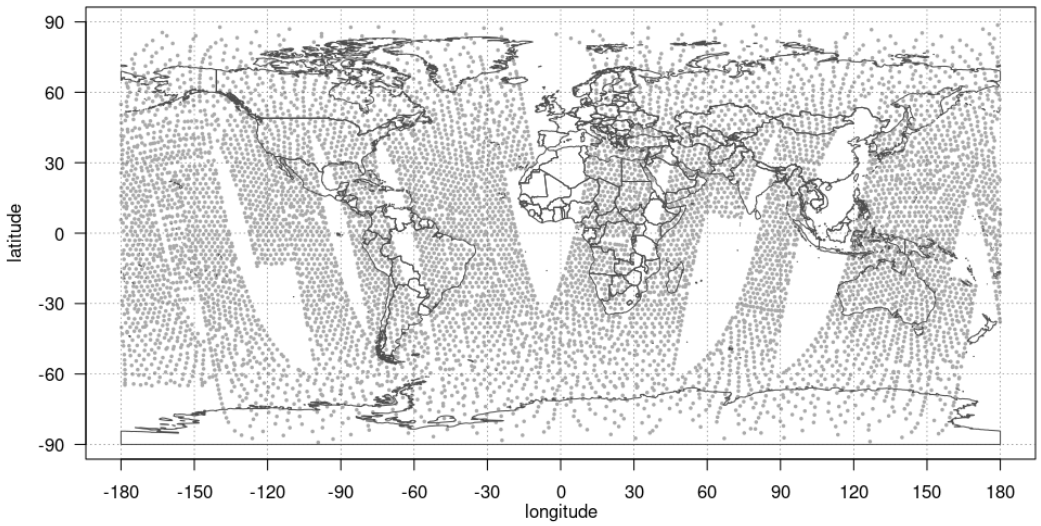

Figure 2.1: Shown is the coverage of observations measured by two satellite instruments (AMSU-A and ATMS) on totally 5 satellites (MetOp-A and -B, NOAA-15 and -19 and NPP) for one assimilation step (31.05.2016 0UTC) at approx. 50 hPa. Each grey dot shows an observation. Data taken from the LETKF reference experiment.

determined from radiances) of two microwave satellite instruments (AMSU-A and ATMS) flying on a total of five satellites (MetOp-A and -B, NOAA-15 and -19 and NPP). It can clearly be seen that even with these two instruments an almost complete coverage of the Earth can be achieved. In total, four satellite instruments, measuring radiances on five satellites, have been assimilated for the 2016 reference experiment. The second experimental period of the LMCPF has been determined for January 2022 and therefore assimilates a much larger number of satellite radiances (15 instruments on 22 satellites) than the older reference experiment. The current operational setup of the ICON model assimilates radiances measured by 12 instruments on 23 satellites.

To compare with the radiance data coverage in Figure 2.1 the global coverage of

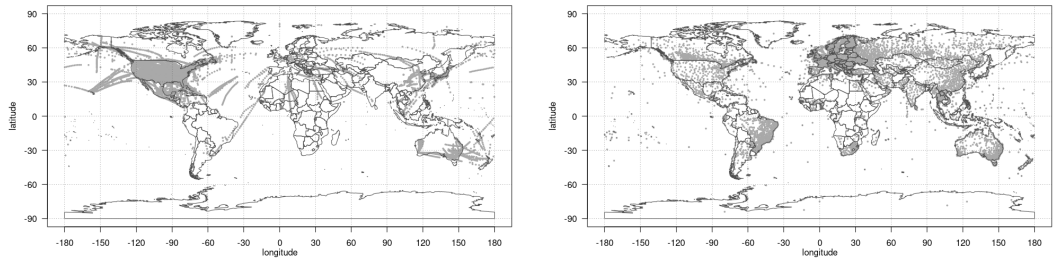

Figure 2.2: Shown is the coverage of observations of two different observation types - left: upper air temperatures measured by AIREPs, right: relative humidity measured by SYNOPs. Shown are the statistics for one assimilation step (31.05.2016 0UTC). Each grey dot shows an observation. Data taken from the LETKF reference experiment.

AIREPs and SYNOPs for the reference experiment of 2016 are shown in Figure 2.2. It can clearly be seen that AIREPs as well as SYNOPs cover some regions very well (e.g. North America resp. Europe) but especially over Africa there is nearly no observation. This again shows the importance of satellite data for the global coverage of observational data.

2.2.2 LETKF

The formulation of the LETKF at DWD is based on the proposal of Hunt et al. (2007). The Ensemble Kalman Filter equations are solved in ensemble space (39 dimensions in case of 40 members). In principle, the Kalman gain matrix uses the background error covariance matrix $\mathbf{P}^b$ in order to determine the analysis increment of the ensemble mean and the symmetric square root of the analysis ensemble covariance matrix $\mathbf{P}^a$. Practically, a weight matrix $\mathbf{W}$ is derived, which is used for the construction of the analysis ensemble as a linear combination of the forecast ensemble members. Since our particle filter implementations directly imitates the LETKF transform, we need to look into more detail here. The operational LETKF system implements equations (20) and (21) of Hunt et al. (2007), i. e.,

$$\overline{\mathbf{w}}^a = \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y}^0 - \overline{\mathbf{y}}^b) \tag{2.2.1}$$

for calculating the mean of the analysis ensemble and $\tilde{\mathbf{P}}^a$ given by

$$\tilde{\mathbf{P}}^a = [(L-1)\mathbf{I} + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b]^{-1}, \tag{2.2.2}$$

where we use the letter L for the number of ensemble members and the notation $\overline{\mathbf{w}}^a$ for the linear coefficients of the analysis mean. $\tilde{\mathbf{P}}^a$ denotes the $L \times L$ analysis covariance in the space of ensemble coefficients. $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{y}^0$ are the observations, $\overline{\mathbf{y}}^b$ is the mean of the model equivalents $H\mathbf{x}^b$ of the observations, H the observation operator and $\mathbf{Y}^b$ is the matrix of ensembles minus mean in observation space. Equation (2) of Hunt et al. (2007) in model space lead to (22) and (23) of Hunt et al. (2007),

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{X}^b \overline{w}^a, \tag{2.2.3}$$

$$\mathbf{P}^a = \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{X}^b)^T, \tag{2.2.4}$$

where $\overline{\mathbf{x}}^a$ is the analysis mean and $\mathbf{X}^b$ is the matrix of ensemble minus its mean. The analysis ensemble is calculated as in (24) of Hunt et al. (2007). We obtain

$$\mathbf{X}^a = \mathbf{X}^b \mathbf{W} \tag{2.2.5}$$

using

$$\mathbf{W} = [(L-1)\tilde{\mathbf{P}}^a]^{1/2} \tag{2.2.6}$$

with the symmetric square root denoted by the 1/2 power of the symmetric matrix $\tilde{\mathbf{P}}^a$. We note that a derivation of this algorithm with its links to classical inverse problems theory can also be found in Nakamura and Potthast (2015), Chapter 5.

2.2.3 Localisation on R

Localisation is performed by calculating independent analyses (weight matrices $\mathbf{W}$) at each analysis grid-point using only the observations in the vicinity of

that location. The observations are weighted smoothly in dependence on their distance to that point according to a localisation function chosen as the 5-th order polynomial described in Gaspari and Cohn (1999), which is similar to a Gaussian but has compact support. We use a horizontal localisation length scale of 300 km and a vertical length scale varying from 0.3 (given in ln(p)) at the surface and 0.8 at the model top (75 km). The length scales are defined following Daley (1993) using the second derivative of the localisation function c at its origin: $l = \sqrt{-1/(\nabla^2 c(0))}$. For a Gaussian this coincides with the standard deviation of the distribution. Formally, the inverse of the observation error covariance matrix $\mathbf{R}$ (in (2.2.2), respectively (2.2.6)) is weighted by a point wise multiplication with the function defined by Gaspari & Cohn, such that observations which are located in a larger distance from the current analysis grid point receive less weight when calculating the analysis. Therefore, this procedure is often denoted as localisation on $\mathbf{R}$.

2.2.4 Multiplicative Inflation and RTPP

The analysis ensemble spread is adjusted by multiplicative inflation with a factor ranging from 0.9 to 1.5, based on an online estimate of spread and ensemble mean RMSE in observation space, following Houtekamer et al. (2005). The inflation factor is estimated locally based on the statistics on observation minus first guess differences as described in Section 3.1.4 and the $\mathbf{W}$ matrices are adjusted respectively. In addition, a relaxation to prior perturbation (RTPP) is applied following Whitaker and Hamill (2012), with a rate of 0.75. The latter preserves a reasonable situation-dependent spread-skill relationship in the analysis cycle.

2.2.5 Assimilation Grid and Interpolation

Tapering the observations with a smooth function and taking the symmetric square root in the LETKF algorithm ensures that the weight matrices only change on scales of order of (or larger than) the localisation length scale. For this reason it is sufficient to derive the weight matrices $\mathbf{W}$ on a coarser analysis grid

$\mathcal{G}$, as for example described in Yang et al. (2009), with spacing of order of this prescribed length scale. Afterwards the $\mathbf{W}$ are interpolated to the model grid, and the final analyses are derived by taking linear combinations of the forecast ensemble members according to the interpolated weight matrices.

2.2.6 Additive Covariance Inflation

In order to account for model errors additive random perturbations consistent with 25% of the amplitude of the climatological B-matrix used in the deterministic EnVar assimilation system are added to the analysis ensemble members. In addition, the sea surface temperature (SST) is perturbed by random perturbations of 1 K which are a linear combination of perturbations with spatial correlation length scales of 100 km and 1000 km and a time scale of one day.

2.2.7 Incremental Analysis Update

The analysis increments applied by the cycled data assimilation system as well as the stochastic perturbations introduce imbalances and spin up effects which are diminished by using an incremental analysis update (IAU) scheme, see Bloom et al. (1996). The combined analysis increments from the LETKF, inflation schemes, and additive perturbations are added to the model trajectory dispensing them over a time interval of three hours, symmetrically adjusted around the analysis date.

2.2.8 Quality of the baseline system

To validate the quality of the NWP system of DWD two plots of the World Meteorological Organization (WMO) Ensemble Prediction System (EPS) verification against analysis are shown.

Figure 2.3 shows the skill of the EPS verification against analysis for the 72 h lead time forecast of temperature at 850 hPa for the Northern Hemisphere. The colours indicate four different forecast centres - orange for Japan Meteorological Agency (JMA), brown for Met Office of the United Kingdom (UKMO), blue for European Centre for Medium-Range Weather Forecasts (ECMWF) and red for

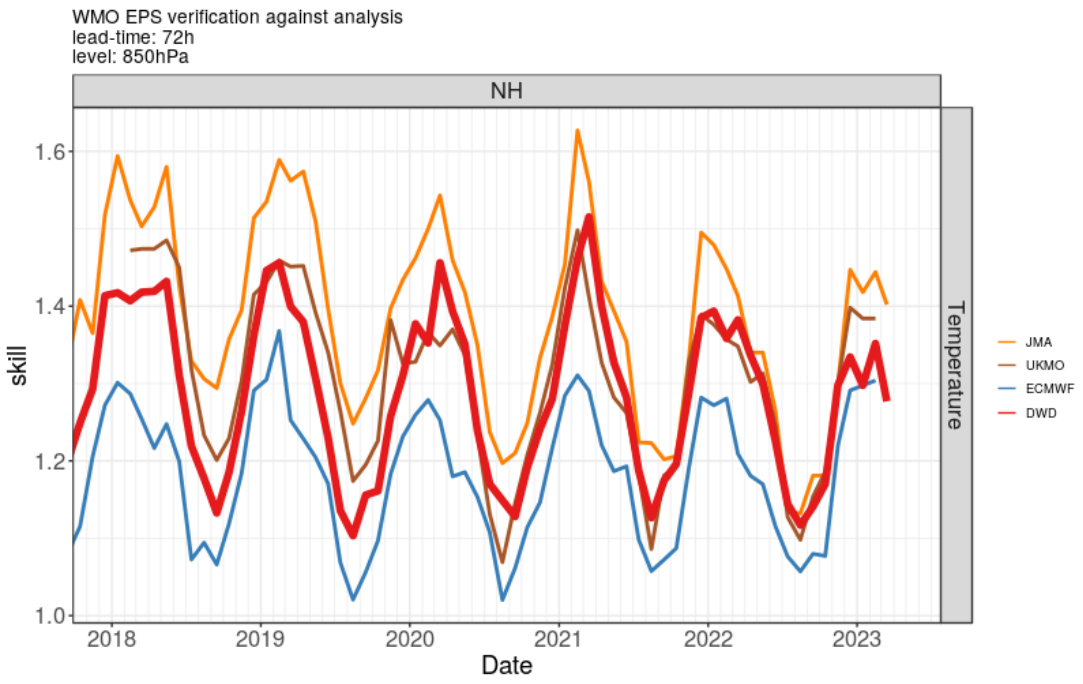

Figure 2.3: WMO Ensemble Prediction Systems (EPS) verification against analysis for four different countries. The line coloures indicate the country - JMA orange, UKMO brown, ECMWF blue and DWD red. Shown is the skill of verification against analysis as time line from 2018 to 2023 for the Northern Hemisphere for the temperature at 850 hPa for a lead time of 72 h. (Source: Image created by DWD, FE1, Verification group, accessed 24.04.2023)

German Meteorological Service (DWD). It can clearly be seen that the skill of the DWD forecasts is on the same level as the skill of the UKMO forecasts and for the end of 2022 and the beginning of 2023 it is even on the same level as the forecasts of the ECMWF.

Figure 2.4 shows again the EPS verification against analysis but for the continuous rank probability score (crps) for the u-component of the wind speed for the 48 h lead time forecasts. Again, it can be seen that the DWD system has been improved during the last five years and is now nearly on the same level as the ECMWF forecasts and even better than the systems of Japan and the Met Office.

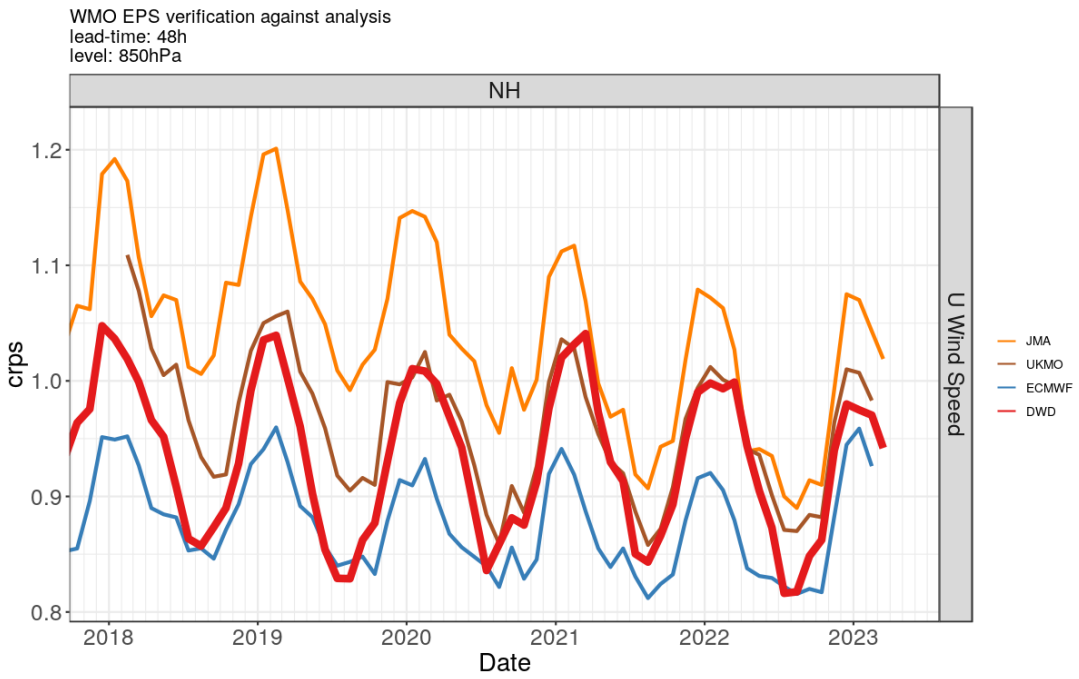

Figure 2.4: Shown are the same statistics as in Figure 2.3 but for the continuous rank probability score (CRPS) instead of the skill and for the u component of the wind speed for a lead time of 48 h at 850 hPa. (Source: Image created by DWD, FE1, Verification group, accessed 24.04.2023)

Chapter 3

The Localised Adaptive Particle Filter

The contents of this chapter have been published in Potthast et al. (2019). Since the reader should be able to read this as a standalone chapter, some parts of Chapter 2 are recapitulated.

Here, we introduce the Localised Adaptive Particle Filter (LAPF) with adaptive Gaussian resampling and spread control. The main objective of this work is to investigate the feasibility and performance of a stable particle filter for global numerical weather prediction with the ICON model in the setup explained in chapters 1.3 and 1.3.1.

The concept of the LAPF is chosen to be as parallel as possible to the Localised Ensemble Transform Kalman Filter described in section 2.2, which serves as a reference for comparison and whose core algorithm is replaced by the LAPF algorithm. Here, we basically replace the analysis step (section 2.2) by the steps described in sections 3.1.1 to 3.1.5, keeping the observation handling, quality control, and part of the inflation and localisation facilities unchanged.

Firstly, section 3.1.1 presents the classical particle filter basis of the method, where as an important first step we describe the ensemble transform version of this particle filter. Afterwards, we describe the projection onto ensemble

space in section 3.1.2. Then, we go into details of classical resampling in section 3.1.3, describe the indicator for spread control in section 3.1.4 and the adaptive Gaussian resampling or rejuvenation in section 3.1.5. We note that the adaptive Gaussian rejuvenation is carried out on top of classical resampling i.e., this is an additional tool for spread control added to the classical particle filter.

3.1 The Localised Adaptive Particle Filter (LAPF) with Gaussian Resampling

3.1.1 The classical Particle Filter

The classical Particle Filter uses an ensemble $x^{(l)}$ of states which represents the prior probability distribution $p_k^{(b)}$ at time t_k in the form of δ-distributions. Alternatively, particles are considered as draws from this prior distribution and, in the case of L particles, they each carry a weight of $\frac{1}{L}$. To carry out the analysis step at time t_k, weights are calculated by

$$w_{k,\ell} := cp(y_k|x^{(l)}), \quad \ell = 1, ..., L, \tag{3.1.1}$$

where c is a normalisation constant and for the particles $x^{(l)}$ corresponding to Equation (2.1.5) in Chapter 2.1 (page 17). We note that sometimes we use the normalisation to L for easier discussion, i.e. $\sum_{\ell=1,...,L} w_{k,\ell} = L$. Then for the prior each particle carries the weight 1.

3.1.2 Projection onto ensemble space.

In the following, we drop the explicit declaration of the time index k and write y^o for the observations y_k at time t_k.

As seen from (2.2.1), the LETKF is based on the projection of the observation onto ensemble space of observation equivalents. For brevity we use $\mathbf{Y}$ for $\mathbf{Y}^b$. We note that the orthogonal projection of the observation difference $\mathbf{y}^o - \overline{\mathbf{y}}^b$ onto

the ensemble space $\{\mathbf{Y}w:\ w\in\mathbb{R}^L\}$ with respect to the scalar product weighted by $\mathbf{R}^{-1}$ in $\mathbb{R}^m$ is given by

$$P(\mathbf{y}^o-\overline{\mathbf{y}}^b)=\mathbf{Y}(\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{R}^{-1}(\mathbf{y}^o-\overline{\mathbf{y}}^b), \tag{3.1.2}$$

compare for example Lemma 3.2.3 of Nakamura and Potthast (2015) using the adjoint $\mathbf{Y}^*=\mathbf{Y}^T\mathbf{R}^{-1}$ based on the weighted scalar product in $\mathbb{R}^m$ and the Euclidean scalar product in the space $\mathbb{R}^L$ of ensemble coefficients[1]. The corresponding particle filter weights (3.1.1) at time t_k based on this projection are given by its ensemble transform projection

$$w_{k,\ell}\ :=\ ce^{-\frac{1}{2}[P(\mathbf{y}^o-H\mathbf{x}^{(\ell)})]^T\mathbf{R}^{-1}[P(\mathbf{y}^o-H\mathbf{x}^{(\ell)})]},\ \ \ell=1,...,L, \tag{3.1.3}$$

with normalisation constant c. Abbreviating $A:=\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}$ and $C:=A^{-1}\mathbf{Y}^T\mathbf{R}^{-1}(\mathbf{y}^o-\overline{\mathbf{y}}^b)$, we first note

$$\mathbf{y}^o-H\mathbf{x}^{(\ell)}=\mathbf{y}^o-(\overline{\mathbf{y}}^b+\mathbf{Y}e_\ell)=(\mathbf{y}^o-\overline{\mathbf{y}}^b)-\mathbf{Y}e_\ell,\ \ \ell=1,...,L \tag{3.1.4}$$

and

$$\begin{aligned}P(\mathbf{y}^o-H\mathbf{x}^{(\ell)}) &= \mathbf{Y}A^{-1}\mathbf{Y}^T\mathbf{R}^{-1}((\mathbf{y}^o-\overline{\mathbf{y}}^b)-\mathbf{Y}e_\ell)\\ &= \mathbf{Y}(C-e_\ell),\ \ \ell=1,...,L.\end{aligned} \tag{3.1.5}$$

where e_ℓ is the standard unit vector which is one in its ℓ-th component and zero otherwise. Now, the exponent of (3.1.3) is transformed into

$$P(\mathbf{y}^o-H\mathbf{x}^{(\ell)})]^T\mathbf{R}^{-1}P(\mathbf{y}^o-H\mathbf{x}^{(\ell)})=[C-e_\ell]^TA[C-e_\ell],\ \ \ell=1,...,L, \tag{3.1.6}$$

[1]This is readily obtained from $\langle z,\mathbf{Y}w\rangle_{\mathbf{R}^{-1}}=\langle z,\mathbf{R}^{-1}\mathbf{Y}w\rangle=\langle\mathbf{Y}^T\mathbf{R}^{-1}z,w\rangle=\langle\mathbf{Y}^*z,w\rangle$, where $\langle\cdot,\cdot\rangle$ denotes an inner product and where $z\in\mathbb{R}^m$ and $w\in\mathbb{R}^L$, see Nakamura and Potthast (2015) for a detailed introduction.

leading to

$$w_{k,\ell} = ce^{-\frac{1}{2}[C-e_\ell]^T A[C-e_\ell]}, \quad \ell = 1, ..., L. \tag{3.1.7}$$

The classical weight (3.1.1) is known to lead to filter divergence in high-dimensional spaces. Here, the ensemble transform and the projection P onto ensemble space lead to a significant reduction of the dimensionality. The observation vector $y \in \mathbb{R}^m$ is mapped onto the vector $A^{-1}\mathbf{Y}^T\mathbf{R}^{-1}(\mathbf{y}^o - \overline{\mathbf{y}}^b)$ in the space $\mathbb{R}^L$ with ensemble size L. The weights (3.1.3) now penalize the distance of the ensemble members e_ℓ, $\ell = 1, ..., L$ in $\mathbb{R}^L$ to the projection C of the observations onto ensemble space. The histograms in Figure 3.1 (see page 37) show the result of this projection step, which in combination with adaptive Gaussian resampling and localisation leads to a feasible behaviour of the particle filter weights.

To evaluate the relationship between the classical particle filter weights and the ensemble space particle filter weights we note

$$\begin{aligned} w_{k,\ell}^{classical} &= ce^{-\frac{1}{2}[(\mathbf{y}^o - H\mathbf{x}^{(\ell)})]^T\mathbf{R}^{-1}[(\mathbf{y}^o - H\mathbf{x}^{(\ell)})]} \\ &= ce^{-\frac{1}{2}[\Big(P+(I-P)\Big)(\mathbf{y}^o - H\mathbf{x}^{(\ell)})]^T\mathbf{R}^{-1}[\Big(P+(I-P)\Big)(\mathbf{y}^o - H\mathbf{x}^{(\ell)})]} \\ &= ce^{-\frac{1}{2}[P(\mathbf{y}^o - H\mathbf{x}^{(\ell)})]^T\mathbf{R}^{-1}[P(\mathbf{y}^o - H\mathbf{x}^{(\ell)})]} \cdot \underbrace{e^{-\frac{1}{2}[(I-P)(\mathbf{y}^o - H\mathbf{x}^{(\ell)})]^T\mathbf{R}^{-1}[(I-P)(\mathbf{y}^o - H\mathbf{x}^{(\ell)})]}}_{=\tilde{c}}, \end{aligned} \tag{3.1.8}$$

where we use the orthogonality of the projection P with respect to the scalar product with weight $\mathbf{R}^{-1}$ such that the mixed terms of P with $I - P$ vanish. The second exponential factor in the last line of (3.1.8) is equal to a constant $\tilde{c}$ for all $\ell = 1, ..., L$ since we have

$$\begin{aligned} (I-P)(\mathbf{y}^o - Hx^{(\ell)}) &= (I-P)(\mathbf{y}^o - \overline{\mathbf{y}}^b + \mathbf{Y}e_\ell) \\ &= (I-P)(\mathbf{y}^o - \overline{\mathbf{y}}^b) - \underbrace{(I-P)\mathbf{Y}e_\ell}_{=0}. \end{aligned} \tag{3.1.9}$$

If the ensemble space spans only a small part of the full state space, the constant $\tilde{c}$ can be very small, so the ensemble transformation can be performed in observation space, since the contribution perpendicular to P is common to all

ensemble members and gives a contribution independent of ℓ. Thus, the ensemble transformation effectively removes a very small but uniform factor from the ensemble weights.

After the determination of the weights (3.1.7) the classical resampling (Section 3.1.3) is carried out. Then, the spread control (Section 3.1.4) will be prepared. Subsequently, the adaptive Gaussian resampling step (Section 3.1.5) will be executed.

3.1.3 Classical Resampling

The LAPF carries out a classical resampling step based on (3.1.7), suggested already in Gordon et al. (1993). For resampling, accumulated weights w_{ac_ℓ}, $\ell = 1, ..., L$, are defined by

$$w_{ac_0} = 0, \;\; w_{ac_\ell} = w_{ac_{\ell-1}} + w_{k,\ell}, \qquad \ell = 1, ..., L, \tag{3.1.10}$$

where we now employ normalisation to the total weight of L. Then, similar to Bain and Crisan (2009) (see also Alam and Gustafsson (2020), Nicely and Wells (2019)), we draw $r_\ell \sim U([0,1])$, $\ell = 1, ..., L$, set $R_\ell = \ell - 1 + r_\ell$ and define the transform matrix for the particles by

$$\breve{\mathbf{W}}_{i,\ell} = \begin{cases} 1, & \text{if } R_\ell \in \; (w_{ac_{\ell-1}}, w_{ac_\ell}], \\ 0, & \text{otherwise,} \end{cases} \tag{3.1.11}$$

$i, \ell = 1, .., L$ with $\breve{\mathbf{W}} \in \mathbb{R}^{L \times L}$, where $(s, t]$ denotes the interval of values $s < \eta \leq t$. This is carried out for each analysis grid point $p \in \mathcal{G}$, for brevity however, we use $\breve{\mathbf{W}}$ instead of $\breve{\mathbf{W}}(p)$.

3.1.4 Spread Control

In ensemble data assimilation systems the spread of the ensemble evolves as a result of model dynamics, model errors (represented by additive perturbations

or multiplicative inflation) and active observations[2], and thus relies on a correct specification of model and observational errors. As it is very difficult to properly estimate and model these errors the spread of the ensemble is adjusted. In the operational LETKF an adaptive inflation factor ρ is estimated, based on statistics of observations minus first guess (compare Desroziers et al. (2005) and Li et al. (2009)). For this purpose we use adaptive Gaussian resampling with parameters based on the estimate of ρ in the LAPF. It is derived from the observation minus background ($o-b$) statistics, the current ensemble spread and the assumed observation error. Its determination is based on

$$\mathbf{d}_{o-b} = \mathbf{y}^o - H(\mathbf{x}^b) = \mathbf{y}^o - H(\mathbf{x}^t) + H(\mathbf{x}^t) - H(\mathbf{x}^b) \approx \boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^b \tag{3.1.12}$$

with the true background state $\mathbf{x}^t$, the background state $\mathbf{x}^b$, the linearisation $\mathbf{H}$ of H, the vector of observation errors $\boldsymbol{\epsilon}^o$ and the vector of background errors $\boldsymbol{\epsilon}^b$. Then, if the observation errors and background errors are uncorrelated, we obtain

$$\mathbb{E}[\mathbf{d}_{o-b}\mathbf{d}_{o-b}^T] = \mathbb{E}[\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^o)^T] + \mathbf{H}\mathbb{E}[\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^T]\mathbf{H}^T \tag{3.1.13}$$

(see Desroziers et al. (2005)). To estimate the inflation factor we substitute the expectation values of the background and observation error with the actual ensemble covariance matrix $\mathbf{P}^b$ multiplied by the inflation factor ρ and the nominal covariance of observation error $\mathbf{R}$, respectively: $\mathbb{E}[\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^T] \approx \rho\,\mathbf{P}^b$ and $\mathbb{E}[\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^o)^T] \approx \mathbf{R}$ resulting in

$$\mathbb{E}[\mathbf{d}_{o-b}\mathbf{d}_{o-b}^T] \approx \mathbf{R} + \rho\,\mathbf{H}\mathbf{P}^b\mathbf{H}^T. \tag{3.1.14}$$

Now by taking the trace $\mathrm{Tr}(A) = \sum_{j=1}^m a_{jj}$ of the matrices on both sides, using $\mathrm{Tr}\big(A + \tilde{A}\big) = \mathrm{Tr}(A) + \mathrm{Tr}\big(\tilde{A}\big)$, $\mathrm{Tr}\big(\rho\tilde{A}\big) = \rho\,\mathrm{Tr}\big(\tilde{A}\big)$ and $\mathrm{Tr}\big(vv^T\big) = \mathrm{Tr}\big(v^Tv\big)$, the

[2] Active observations are those which passed the quality control, see section 2.2.

inflation factor ρ is estimated by

$$\rho = \frac{\mathbb{E}[\mathbf{d}_{o-b}^T \mathbf{d}_{o-b}] - \mathrm{Tr}(\mathbf{R})}{\mathrm{Tr}\Big(\mathbf{H}\mathbf{P}^b\mathbf{H}^T\Big)} \tag{3.1.15}$$

(following Desroziers et al. (2005) and Li et al. (2009)).

(3.1.15) computes a scalar inflation factor ρ based on a set of observations and the corresponding ensemble spread in observation space. It is carried out locally as a localised ensemble data assimilation method is employed, i.e. we calculate

$$\rho(p) = \frac{\mathbf{d}_{o-b}^T \mathbf{d}_{o-b} - r^2}{q^2} \tag{3.1.16}$$

at each point $p \in \mathcal{G}$ with the local innovation vector $\mathbf{d}_{o-b}$, the observation error $r^2 = \mathrm{Tr}(\mathbf{R})$ and the local estimate $q^2 := \mathrm{Tr}\Big(\mathbf{H}\mathbf{P}^b\mathbf{H}^T\Big)$ of the background error covariance in observation space. The factor $\rho(p)$ is the estimate for the local variance inflation at the analysis point p in the LETKF.

Due to the localisation procedure the number of observations used in this method may be small and the estimated value of ρ may be based on limited statistics. To make the estimate more robust we first limit ρ by lower and upper bounds of 0.9 and 1.5 and afterwards perform a temporal smoothing: A weighting factor $\alpha = 0.05$ is chosen to combine the $\tilde{\rho}_k$ estimated by (3.1.16) in the current cycle k and the ρ_{k-1} used in the previous analysis cycle (3 h in the past) to get the ρ_k to be applied in the current cycle k:

$$\rho_k = \alpha\tilde{\rho}_k + (1-\alpha)\rho_{k-1}, \quad k = 1, 2, 3, ... \tag{3.1.17}$$

In the LETKF, ρ is used at each analysis grid-point to calculate the filter transformation matrix $\mathbf{W}_{\text{infl}}$ by

$$\mathbf{W}_{\text{infl}}(p) = \sqrt{\rho(p)}\mathbf{W}(p), \quad p \in \mathcal{G}, \tag{3.1.18}$$

where $\mathbf{W}$ is the transform matrix defined in (2.2.6). For the localised adaptive

particle filter pure multiplicative inflation is not appropriate, since it would just inflate the distribution of the remaining duplicate ensemble members. Instead we apply the Gaussian resampling based on (3.1.15) - (3.1.16) as described below.

3.1.5 Gaussian Resampling or Rejuvenation

The LAPF first calculates the ensemble weights according to the ensemble space projection (3.1.7) and the resampling (3.1.11) at each of the analysis points $p \in \mathcal{G}$. Usually this resampling leads to a part of the total number of particles only getting the majority of the weights. Often, a rejuvenation step (see Doucet et al. (2001) or van Leeuwen et al. (2015), equation (2.39)) is carried out around each of the remaining particles, i.e., new particles are generated based on a pseudo-random draw in ensemble space. We note that this rejuvenation can be considered as classical resampling from some posterior represented by a superposition of Gaussian functions in the spirit of classical MCMC methods (Nakamura and Potthast (2015), Chapters 4 and 5). Note also that jittering is a special case of rejuvenation.

In contrast to Lang et al. (2022), who are using Metropolis Hasting for jittering, we draw from a Gaussian distribution around each remaining particle in ensemble space, used with appropriate multiplicity as constructed by $\breve{\mathbf{W}}$ in (3.1.11). Using a Gaussian with mean given by the column vector $\breve{\mathbf{W}}_\ell$ of $\breve{\mathbf{W}}$ and with a covariance matrix $\sigma^2\mathbf{I} \in \mathbb{R}^{L\times L}$, this leads to a draw from a distribution in physical space with mean given by the ensemble $\{x^{(\ell)}, \ell = 1, ..., L\}$ and the re-scaled ensemble covariance matrix $\sigma^2\mathbf{P}^b$.

Global Rejuvenation

A pseudo-random matrix $\mathbf{N} \in \mathbb{R}^{L\times L}$ with each element draw from a Gaussian distribution with mean zero and covariance 1 is chosen globally to ensure the best possible continuity of the meteorological variables in physical space. Since resampling is global, without modulating the random matrix $\mathbf{N}$ the resulting

perturbations would be scaled superpositions of the original ensemble members, keeping linear balances completely and non-linear balances to some degree. However, adaptivity of the pseudo-random draws turns out to be crucial to avoid filter divergence and filter collapse. Only after adapting the spread of the rejuvenation or resampling carefully as follows, the filter started to be stable.

Adaptivity

For the adaptive resampling step the size of the draw (given by $\mathbf{N}$) is modulated by applying a scalar perturbation factor σ for each analysis grid point. Scaling of the draw around each member at time t_k is carried out by

$$\mathbf{W}(p) = \breve{\mathbf{W}}(p) + \mathbf{N} \cdot \sigma(\rho_k(p)). \tag{3.1.19}$$

The specification of the factor σ is based on the inflation parameter ρ_k estimated in (3.1.17):

$$\sigma(\rho) := \begin{cases} c_0, & \rho < \rho^{(0)}, \\ c_0 + (c_1 - c_0) \cdot \frac{\rho - \rho^{(0)}}{\rho^{(1)} - \rho^{(0)}}, & \rho^{(0)} \leq \rho \leq \rho^{(1)} \\ c_1, & \rho > \rho^{(1)}, \end{cases} \tag{3.1.20}$$

where elementary tuning tests lead to the values $c_0 = 0.02$, $c_1 = 0.2$, $\rho^{(0)} = 1.0$ and $\rho^{(1)} = 1.4$ (which might not be an optimal choice). The function continuously depends on ρ with $\sigma = c_0$ if $\rho \leqq \rho_0$ and $\sigma = c_1$ if $\rho \geqq \rho_1$.

We summarize that we always perturb each of the remaining members of the filter after the classical resampling step by a Gaussian with standard deviation of at least c_0 and at most c_1. The scaling is a continuous function of the input parameters of our estimator of ρ, i.e. it depends continuously on space, on the observations and the ensemble members. The only discontinuities can occur when the matrix $\breve{\mathbf{W}}$ in dependence of the spatial point p itself is discontinuous according to a change in the weights resulting from the classical resampling

procedure.

Finally, the full analysis ensemble (Hunt et al., 2007) is given by

$$\begin{aligned} \mathbf{X}^{(a,full)} &= \overline{x}^{(b)} + \mathbf{X}^{(b)}\mathbf{W} \\ &= \overline{x}^{(b)} + \underbrace{\mathbf{X}^{(b)}\breve{\mathbf{W}}}_{class.\ resampling} + \underbrace{\mathbf{X}^{(b)}\mathbf{N}\sigma}_{rejuvenation} \end{aligned} \tag{3.1.21}$$

Number of Surviving Particles

We remark that the adaptive Gaussian rejuvenation ensures that with probability one the analysis ensemble consists of L distinct particles. However, in the first classical resampling step, the number of distinct particles can be significantly smaller with the limiting case where only one particle gains all the weight, such that after resampling and before rejuvenation we have L identical copies of this particle in a given localisation point. It is of course highly interesting to study the statistics of how many particles remain at each analysis grid point and in what way the ensemble projection (3.1.3) helps the filter to stay away from collapse.

In Figure 3.1 for three dates close to the end of the experimental period we show some histograms of the number of particles at each analysis grid point surviving the first resampling step, i.e. those particles with weights $w_{k,l} \geq 1$ before rejuvenation when normalising the total weight to L. The results show that at 100 hPa (high level with few data, first column) mainly five to twenty particles obtain most of the weight. There are considerably fewer cases where twenty to forty particles survive the classical resampling step.

In the middle of the troposphere (500 hPa, column in the middle) it can clearly be seen that the first date (20.05.2016, 03 UTC) and the third date (31.05.2016, 21 UTC) show a quite similar distribution. The number of cases with one up to thirty particles with weight larger than one is very similar. For the second date (second row - 25.05.2016, 12 UTC) there are only a few cases with more than twenty surviving particles. This is probably due to the larger amount of synoptic

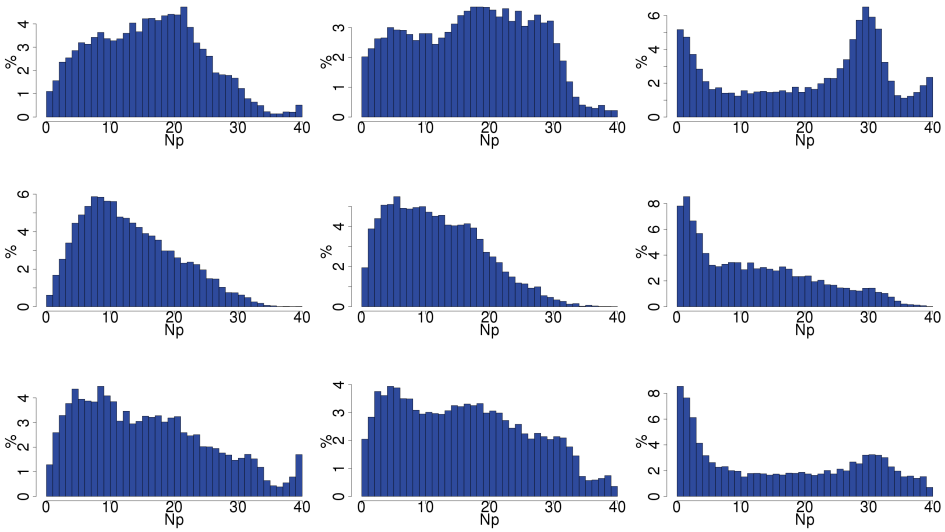

Figure 3.1: Here we show global histograms of the number of particles with weights above 1 (when the total weight is given by the number of particles L) for three dates and three pressure levels. In the first row we show results for May 20, 2016 at 03 UTC, the second row displays results for May 25, 2016 at 12 UTC and in the third row the results for May 31, 2016 at 21 UTC are shown. From left to right we show the levels 100 hPa, 500 hPa and 1000 hPa. The x-axis shows the number of particles (Np) with weight larger than one, the y-axis the percentage of analysis grid points with these numbers.

data at 12 UTC compared to 03 and 21 UTC.

The result for the bottom level (1000 hPa, last column) differs from the other two levels. In all these sub-figures there exists one peak at one to five particles. For the first and the third row there is also a peak at thirty particles. The first peak might be due to model biases in the boundary layer in combination with a high number of observations.

A high number of observations leads to a small number of particles surviving the classical resampling, which is the well-known filter divergence phenomenon. We remark, however, that due to the ensemble transform and projection step of Section 3.1.1, this divergence does only occur in a part of the localisation boxes. Further, adaptive Gaussian rejuvenation in ensemble space guarantees the calculation of L distinct analysis ensemble members with a controlled spread and distribution.

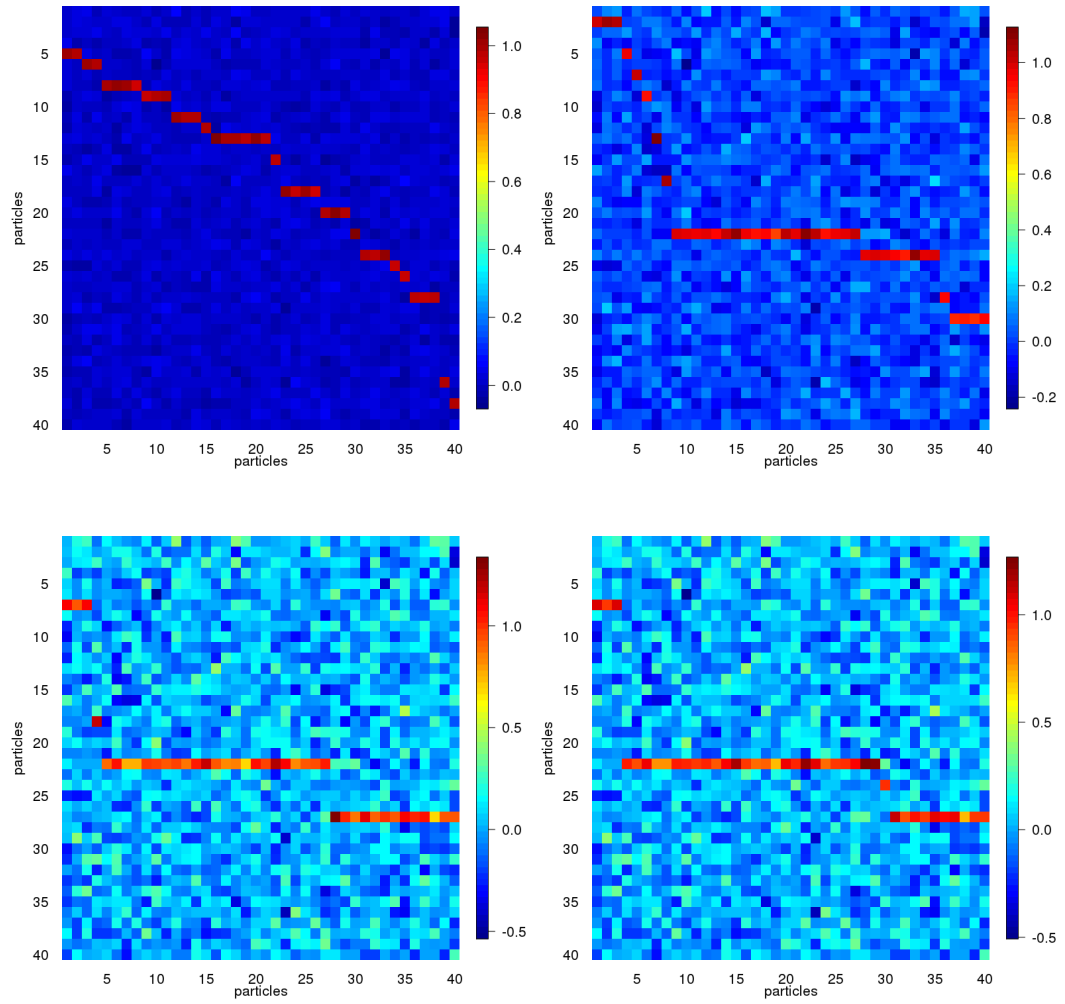

Figure 3.2: Examples for transformation matrices **W** of the LAPF after Gaussian resampling. We show **W** at one analysis grid point (60 N, 90 E (Siberia)), at levels of 100 hPa, 500 hPa, 750 hPa and 1000 hPa (from left to right and top to bottom) for May 26, 2016, 00 UTC. The x-axis shows the analysis ensemble index, the y-axis the first guess ensemble index.

Examples

Examples for the matrix **W** from (3.1.19) are displayed in Figure 3.2. The **W** matrices show how the analysis ensemble is constructed from the first guess ensemble. Entries close to one indicate that an analysis ensemble member is sampled from the respective first guess member. In general some first guess members lead to multiple analysis members whereas others are dismissed totally (only entries close to zero in that row). The deviations from zero or one are consequences of the Gaussian resampling step.

In the first case (top left, in the lower stratosphere with few observations) the analysis members are sampled from a large number of first guess particles. Going further to the ground fewer first guess particles are selected, but are replicated multiple times due to their large weight. This is caused by the larger number of observations further down in the atmosphere putting more constraints on the prior distribution.

It may also be noted that in the first sub-figure the strength of the Gaussian resampling is weak so that entries remain close to zero or one, respectively. In the subsequent figures, the estimated inflation factor is larger so that final weights almost fill the range in between -0.5 and 1.

3.2 Numerical Tests in the Operational Framework and Results

Here, we present tests of the localised adaptive particle filter for the global ICON model for a typical experimental setup and for a time period of one month. In particular, we investigate the assimilation cycle and forecast scores in some detail, diagnosing the development of spread, bias, scores and the stability of the system.

We start with a short description of the experimental setup in section 3.2.1. Afterwards we evaluate the assimilation cycle of the LAPF experiment in Section 3.2.2. As reference we use the LETKF implementation which is operational at DWD. With the help of spread control, the LAPF provides reasonable results and is stable over the full experimental period. In Section 3.2.3 we investigate the quality of forecast runs and compare it to the performance of the system in the operational setup.

3.2.1 Development Environment, Experimental Setup and Period

The main goal of our work is to investigate the feasibility and performance of a stable particle filter for global numerical weather prediction with the ICON model. For our experimental test we chose the period of May 1-31, 2016.

Since quality control is carried out based on the deterministic run, it is always part of experiments. Here for the development of the LAPF, we choose the established 52 km experimental resolution for the ensemble and 26 km for the deterministic run. In standard DWD experiments, Ensemble Data Assimilation tests are usually run with the operational ensemble size $L = 40$ members, which is also our choice for the LAPF development. In this setup, we study the performance and stability of the particle filter in direct comparison to the LETKF based operational setup. Here, we use the term stable to denote a filter which does not degenerate, i.e. the spread of the ensemble does nor become too small or too large, providing some realistic estimate for the forecast uncertainty (Brett et al., 2013).

The development of the Localised Adaptive Particle Filter takes place in the Data Assimilation Coding Environment of DWD. The suite includes modules for snow analysis (SNOW) every three hours, sea surface temperature analysis (SST) and soil moisture analysis (SMA) once per day. The surface analysis consists of separate modules in which amongst others random perturbations are added to the ensemble members. For our experiments this part has been kept identical to the operational setup (compare Reinert et al. (2018)).

3.2.2 Assimilation cycle

For the experimental diagnostics, a spin-up period of one week is excluded from the observation minus first guess[3] statistics, as bias correction algorithms and

[3]First guess, also known as background, in the context of the DWD's data assimilation system, is the three-hour forecast based on the previous analysis.

ensemble spread have to adapt to the new method.

We first investigate upper-air observation minus first guess (obs-fg) statistics based on radiosonde observations (TEMP). Some selection of results is displayed in Figure 3.3, in particular bias (ME), root mean square error (RMSE) and standard deviation (SD) for 3-hour-forecasts (first guesses), generated by the reference (LETKF) and the particle filter (LAPF). These statistics are based on observations which passed the quality control, i.e. which were actually used in the assimilation. Figure 3.3 visualises statistics for the global domain and for the time period May 8, 2016 to May 31, 2016.

The results show that the root mean square error and standard deviation of the current LAPF is about 10%-15% worse than those of the LETKF. It also shows that the system is functioning and shows comparable features to the LETKF based statistics. The LAPF shows better results for upper air temperature than for relative humidity. For the bias, the results for relative humidity determined by the LAPF are slightly better than the results of the LETKF (in the sense that they are closer to zero). The values of RMSE and SD in the vertical column show similar shape, but with higher values for LAPF. Overall, this first implementation of the LAPF scheme, where we were not yet able to carry out the time-consuming full tuning which is usually done for an operational system, shows a very reasonable behaviour in comparison to the LETKF.

Since filter collapse and filter divergence are of high interest, we next investigate the behavior of corresponding diagnostics, investigating spread behaviour and the number of surviving particles in each resampling step.

Figure 3.4 shows the spread averaged over all ensemble members for May 31, 2016 at 00 UTC, i.e. at the end of the first month of cycling. Displayed is the spread at ICON-level 64 (approximately 500 hPa) for upper air temperature and specific humidity. In this context spread is the point wise variance of the ICON-EPS.

- The left panels in the second and third row show the fields for the LAPF,

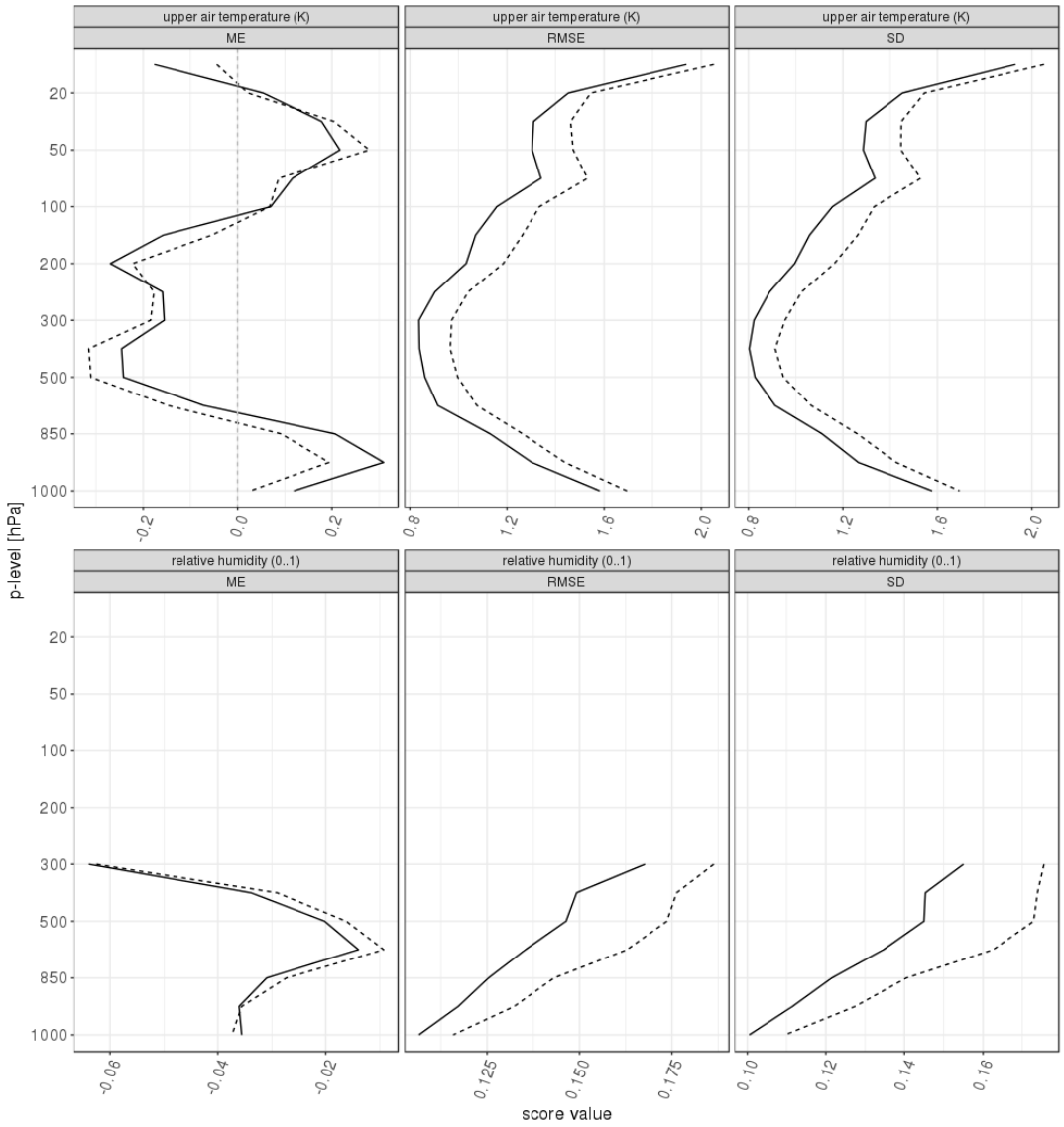

Figure 3.3: Observation minus ensemble mean first guess statistics in the 3-hourly assimilation cycle for the period May 8, 2016 to May 31, 2016 in the global domain: mean deviation (ME - left column), root mean square error (RMSE - middle column) and standard deviation (SD - right column). The first row shows upper air temperature $[K]$, and the second row relative humidity [0..1]. Solid lines indicate the reference (LETKF), dashed lines the experiment (LAPF).

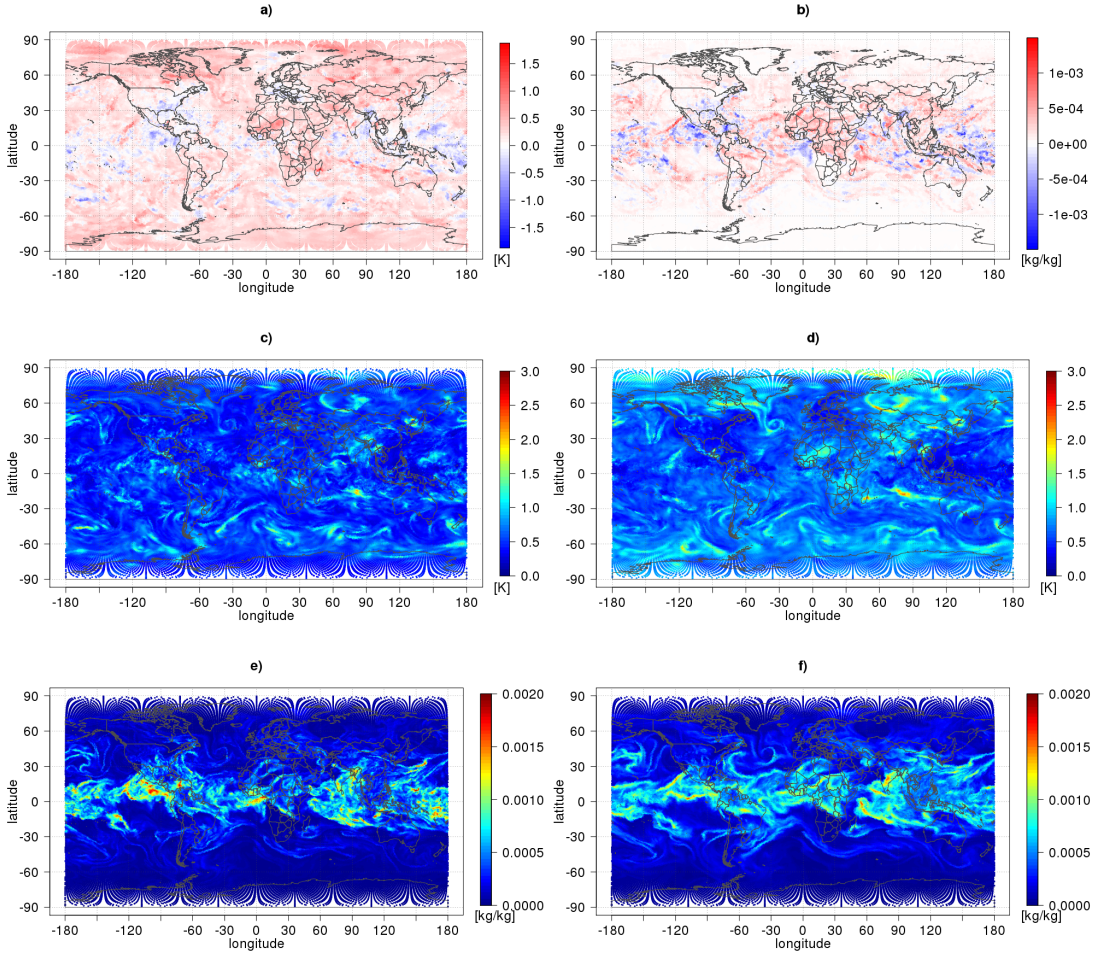

Figure 3.4: Spread at model level 64 (approx 500 hPa) for May 31, 2016, 00 UTC. The first row shows differences for a) upper air temperature $[K]$ and b) specific humidity $[\frac{kg}{kg}]$ between LETKF and LAPF. The second row shows fields for upper air temperature (LAPF in panel c), LETKF in panel d)), and the third row shows the fields for specific humidity (LAPF in panel e), LETKF in panel f)).

the right panels those for the LETKF. In the first row the differences of the spread between LETKF and LAPF are displayed for upper air temperature in a) and specific humidity spread in b). It can clearly be seen that the structures are similar, but the spread for both variables is higher for the LETKF ($\overline{\sigma_T} = 0.6653\ K$ and $\overline{\sigma_q} = 0.0003\ \frac{kg}{kg}$) than for the LAPF ($\overline{\sigma_T} = 0.5037\ K$ and $\overline{\sigma_q} = 0.0002\ \frac{kg}{kg}$).

- The main structures of the spread of LAPF and LETKF show similar physical features. The differences between the two filters are more random and linked to the stochastic parts of the methods.

- The regions with the maximum of spread are also the places where the LAPF and the LETKF differ most. One example for this is the temperature over Madagascar. Here the spread of the LETKF is much larger than that of the LAPF. Also, the vortex over western Siberia shows a big difference between the two filters.

- For the specific humidity the biggest differences are situated in the tropics where humidity values are large. The spread difference plot clearly resembles the patterns of the spread in specific humidity itself. It indicates that the differences between LAPF and LETKF are often situated at the borders of the big vortices. The largest differences are situated for example in the region around the west coast of Mexico, where the LAPF has a maximum but the LETKF does not.

For studying the development of the spread and the stability of the filter, time series from May 1, 2016 to May 31, 2016 for both filters are plotted in Figure 3.5. The spread starts from zero and needs a couple of days to settle, because we started all members from the same state duplicated $L = 40$ times. The first row shows the mean of the spread for upper air temperature and specific humidity, calculated at each point in time and for one horizontal level of the atmospheric grid. For both quantities the LAPF shows lower values for mean-spread than

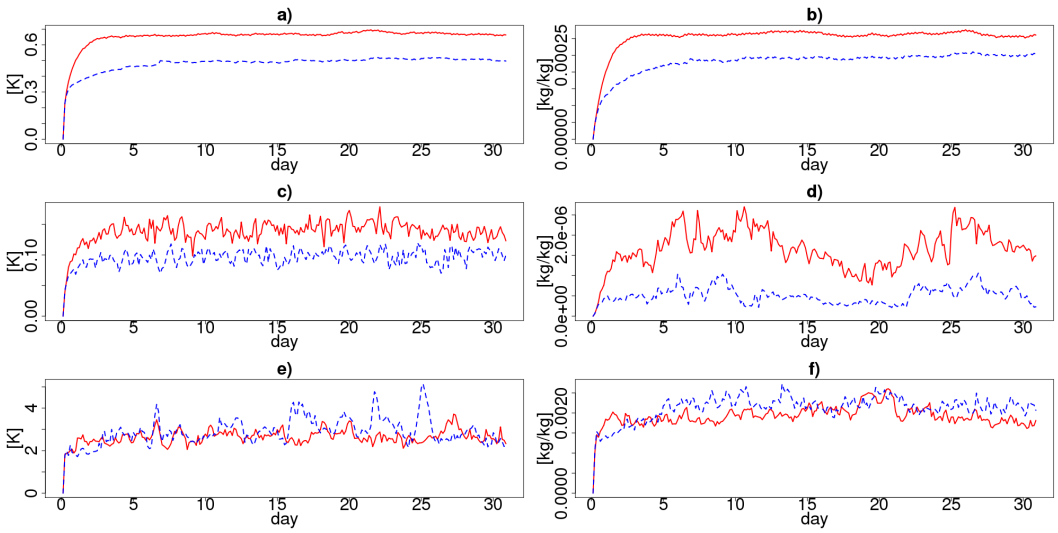

Figure 3.5: Time series of the first guess ensemble spread at ICON-level 64 (approx 500 hPa) averaged globally. The left column shows the spread for upper air temperature $[K]$, the right one for specific humidity $[\frac{kg}{kg}]$. The upper row shows the mean, the middle row the minimum and the lower one the maximum of the spread (red (solid) line = LETKF, blue (dashed) line = LAPF).

the LETKF. The same holds for the minimum value of the spread. However, the maximum values of the spread of the LETKF and the LAPF show quite similar values.

3.2.3 Forecast Verification

Forecasts were run twice a day at 00 UTC and 12 UTC. In Figure 3.6 and 3.7 a verification of temperature, relative humidity and wind components compared to radiosonde observations is shown: continuous ranked probability score (CRPS), standard deviation (SD), root mean square error (RMSE) and bias (ME) for lead times of 24, 48, 72, 96, 120 and 144 hours for the forecasts based on the LETKF and the LAPF analyses. The determination of SD, RMSE and MEAN is based on the ensemble mean. At this stage, the results show that in the current development stage the LAPF does not outperform the LETKF. However, the shapes of CRPS, SD and RMSE are comparable, indicating that the LAPF is stable and has a reasonable behaviour. The humidity bias of the LAPF is reduced in comparison to the LETKF at all heights above 850 hPa. It is consistent with

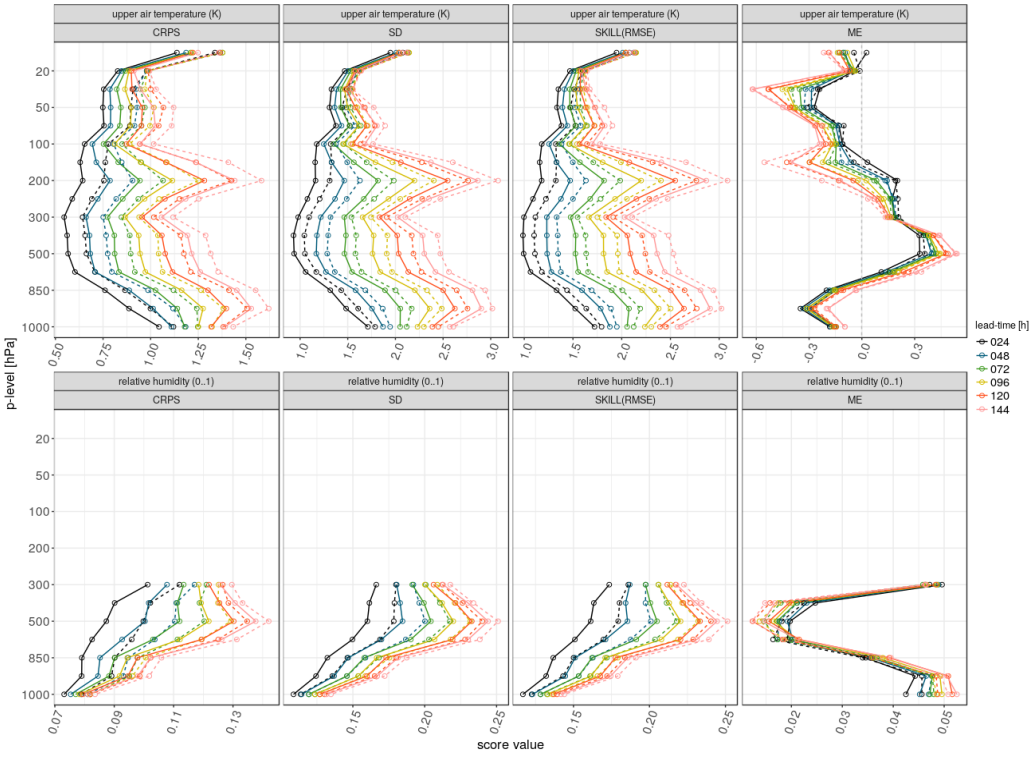

Figure 3.6: Global statistics of forecasts against observations for the period of May 2, 2016 to May 24, 2016. Displayed is the continuous ranked probability score (CRPS, left column), the standard deviation (SD, column two), the root mean square error (RMSE, column three) and the mean (ME, right column). The first row shows the statistics for upper air temperature $[K]$, the second row for relative humidity [0..1]. The solid line indicates the reference (LETKF), the dashed line the experiment (LAPF). Colours indicates different lead-times (from 24 h to 144 h).

theory and implementation that the particle filter does not draw the model fields to the observations as strongly as the LETKF, a distance which is maintained throughout all forecast lead times and explains the behaviour of the scores.

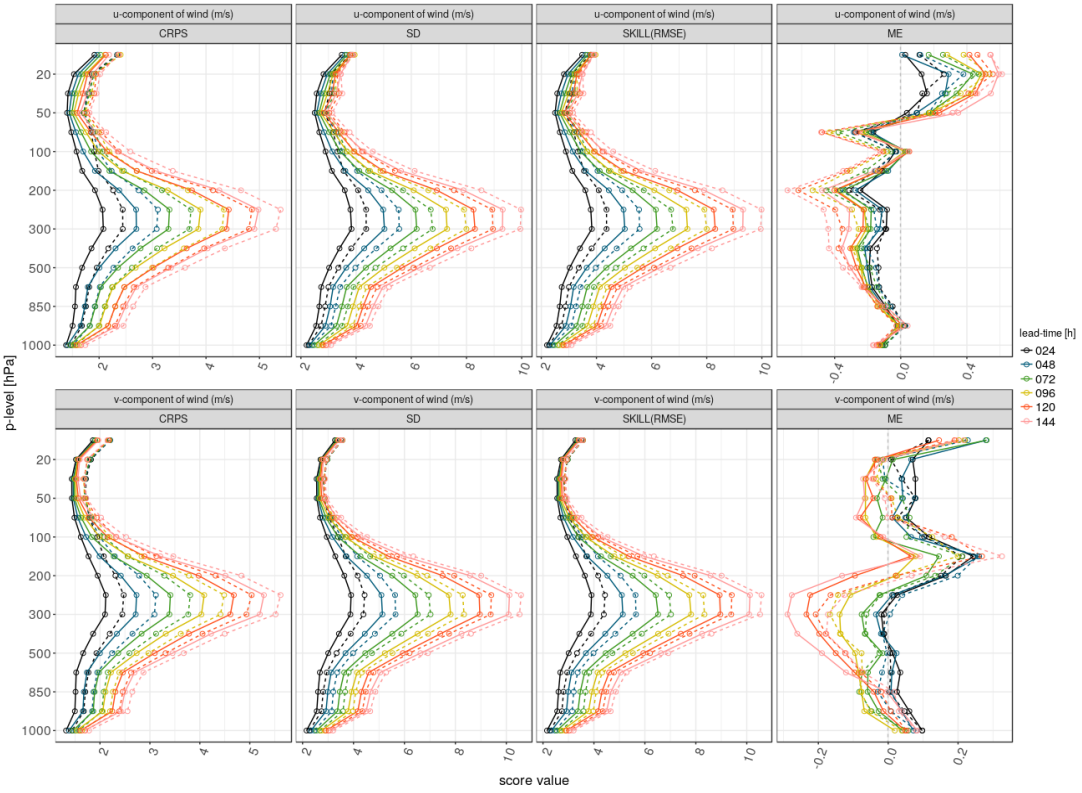

Figure 3.7: Same as in Figure 3.6 but for u-component of wind $[\frac{m}{s}]$ (upper row) and for v-component of wind $[\frac{m}{s}]$ (lower row).

3.3 Conclusions

Standard algorithms for data assimilation used for large-scale atmospheric analysis in operational centres include the Ensemble Kalman Filter and 4d-Var. These data assimilation methods are either inherently or practically based on the assumption that the underlying distribution is Gaussian. If the ensemble distribution is not Gaussian, these methods are not optimal. In the case of non-Gaussianity more general Bayesian methods such as the particle filter have been proposed. The core idea of the particle filter is to realise the Bayesian approach giving a weight to each particle depending on its distance to the observations.

The adaptation of particles is carried out in different ways (van Leeuwen et al., 2019), either by resampling (Potthast et al., 2019; Rojahn et al., 2023), nudging particles towards some proposal distribution (Zhu et al., 2016; van Leeuwen, 2010) or by optimal transport processes (Reich, 2013).

Classical particle filter in high dimensional dynamical systems suffer from filter divergence or filter collapse due to the curse of dimensionality. In this chapter we have developed and implemented a Localised Adaptive Particle Filter (LAPF) in ensemble space with spread control and Gaussian resampling or rejuvenation. With the help of modulated rejuvenation we prevent the filter divergence as well as filter collapse. It has been implemented for global atmospheric data assimilation to fit into the framework of the global operational weather prediction model ICON of DWD.

The LAPF was tested over a period of one month with 40 ensemble members, a global horizontal resolution of 52 km and 90 vertical layers in an operational setup with slightly reduced resolution. A comparison of the scores with those of the operational system of DWD (with some modest reduction of resolution) shown in Section 3.2 demonstrate that the localised adaptive particle filter is able to provide reasonable atmospheric analysis in a large-scale environment. We have shown that for this first attempt the RMSE-quality of forecasts based on the LAPF is 10-15% behind the forecast quality of the LETKF for forecasts up to several days (compare Figure 3.6 and 3.7) and BIAS is partly improved, in particular for humidity. Altogether, for the assimilation cycle and forecasts the LAPF shows promising results (Figure 3.3). Furthermore, we are able to demonstrate the stability of the LAPF over a period of one month (compare Figures 3.4 and 3.5) and show that atmospheric data assimilation within an operational modelling environment is possible based on a localised adaptive particle filter approach.

Chapter 4

Particle Filtering and Gaussian Mixtures - On a Localised Mixture Coefficients Particle Filter (LMCPF) for global NWP

The contents of this chapter have been accepted for publication in Rojahn et al. (2023). Since the reader should again be able to read this as a standalone chapter, some parts of Chapter 2 are recapitulated.

Here, our starting point is the investigation of the behaviour of the Localised Adaptive Particle Filter (LAPF) with respect to errors in the prior distribution $p_k^{(b)}$. By studying the statistics of the observations vector mapped into ensemble space, we will show that in many cases the model forecasts show significant distance to the observations, and the particle filter based on a limited number of δ-distributions does not pull the particles close enough to the observations when the move of particles is only achieved through adaptive resampling.

To allow individual particles to move towards the observations, we further develop the LAPF by bringing ideas from Gaussian Mixtures into its framework. We reach this goal by including model and forecast uncertainty for each parti-

cle, as for example suggested by the Low-Rank Kernel Particle Kalman Filter (LRKPKF) of Hoteit et al. (2008), compare also Liu et al. (2016a) and Liu et al. (2016b). The basic idea is to consider each particle to be a Gaussian where its width is representing its uncertainty. This means we study a prior distribution given by a Gaussian mixture. Note that this is a special case of Radial Basis Functions (RBFs). Then, the prior has the form

$$p^{(b)}(x) := c \sum_{\ell=1}^{L} c_\ell e^{-\frac{1}{2}(x-x^{(b,\ell)})^T \mathbf{G}^{-1}(x-x^{(b,\ell)})}, \quad x \in \mathbb{R}^n, \tag{4.0.1}$$

with constants $c_\ell = 1/\sqrt{(2\pi)^n \det(\mathbf{G})}$ for the individual Gaussian basis functions with mean $x^{(b,\ell)}$ and covariance $\mathbf{G}$ and a normalization constant c, which in this case is given by $c = 1/L$. For this approximation, and when the observation operator H is linear, we can explicitly calculate the posterior distribution as a corresponding Gaussian Mixture, i.e.

$$p^{(a)}(x) := \tilde{c} \sum_{\ell=1}^{L} c_\ell w_\ell e^{-\frac{1}{2}(x-\tilde{x}^{(a,\ell)})^T \tilde{\mathbf{G}}^{-1}(x-\tilde{x}^{(a,\ell)})}, \quad x \in \mathbb{R}^n, \tag{4.0.2}$$

with some matrix $\tilde{\mathbf{G}}$ (calculated e.g. in Chapter 5.4 of Nakamura and Potthast (2015)), constants w_ℓ given by

$$\begin{aligned}
w_\ell &= \int_{\mathbb{R}^n} \tilde{c}_\ell e^{-\frac{1}{2}(x-x^{(b,\ell)})^T \mathbf{G}^{-1}(x-x^{(b,\ell)})} e^{-\frac{1}{2}(y-H(x))^T \mathbf{R}^{-1}(y-H(x))}\, dx \\
&= \int_{\mathbb{R}^n} \tilde{c}_\ell e^{-\frac{1}{2}(x-\tilde{x}^{(a,\ell)})^T \tilde{\mathbf{G}}^{-1}(x-\tilde{x}^{(a,\ell)})} e^{-\frac{1}{2}(y-H(x^{b,\ell}))^T (\mathbf{H}\mathbf{G}\mathbf{H}^T+\mathbf{R})^{-1}(y-H(x^{b,\ell}))}\, dx \\
&= \tilde{c}_\ell \sqrt{(2\pi)^n \det\left(\tilde{\mathbf{G}}\right)}\; e^{-\frac{1}{2}(y-H(x^{b,\ell}))^T (\mathbf{H}\mathbf{G}\mathbf{H}^T+\mathbf{R})^{-1}(y-H(x^{b,\ell}))} \\
&= e^{-\frac{1}{2}(y-H(x^{b,\ell}))^T (\mathbf{H}\mathbf{G}\mathbf{H}^T+\mathbf{R})^{-1}(y-H(x^{b,\ell}))}
\end{aligned} \tag{4.0.3}$$

with $\tilde{c}_\ell = 1/\sqrt{(2\pi)^n \det\left(\tilde{\mathbf{G}}\right)}$ explicitly calculated by equation (40) in Schenk et al. (2022), with temporary analysis states $\tilde{x}^{(a,\ell)}$, $\ell = 1, ..., L$, with

$$\tilde{c} := \frac{1}{\sum_{\ell=1}^{L} c_\ell w_\ell \sqrt{(2\pi)^n \det\left(\tilde{G}\right)}},$$

and with the components

$$q^{(a,\ell)}(x) := \tilde{c} c_\ell w_\ell e^{-\frac{1}{2}(x-\tilde{x}^{(a,\ell)})^T \tilde{\mathbf{G}}^{-1}(x-\tilde{x}^{(a,\ell)})}, \quad x \in \mathbb{R}^n. \tag{4.0.4}$$

The constant $\tilde{c}$ will normalize the integral of $p^{(a)}$ to one, but not individual terms $q^{(a,\ell)}$ given by (4.0.4).

If there are no further constraints to the variables, the ℓ-th posterior particle can be directly drawn with relative probability w_ℓ from the distribution component $q^{(a,\ell)}(x)$ leading to an analysis ensemble member $x^{(a,\ell)}$. This drawing process is carried out based on localisation, adaptivity and the transformation into ensemble space as developed for the LAPF (see chapter 3); details will be described in sections 4.1.1 and 4.1.2. As for other particle filters, the posterior particles will be calculated by an ensemble transform matrix, with details worked out in Section 4.1.2. For each posterior ensemble member, based on the prior Gaussian mixture, this matrix defines transformation coefficients arising from the weights of each particle. The name Localised Mixture Coefficients Particle Filter (LMCPF) has been used to distinguish from other localised particle filter methods. For example, Reich and Cotter (2015) present Localised Particle Filter (LPF) versions, which include sophisticated optimal transport properties. A further LPF method is introduced by Penny and Miyoshi (2016). We note that the choice for $\mathbf{G}$ of formula (4.0.1) as a scaled version of the ensemble correlation matrix of Hunt et al. (2007), i.e. $\mathbf{G} = \kappa\mathbf{B}$, with $\mathbf{B} = \frac{1}{(L-1)}\mathbf{X}\mathbf{X}^T$, resembles the choices made for the LETKF (Hunt et al., 2007) and leads to very efficient code. We will investigate the usefulness of the Gaussian uncertainty within the particle filter in very high-dimensional systems, leading to moves or shifts of the particles towards the observations. Statistics of these shifts will be shown, demonstrating that for this global atmospheric NWP system the uncertainty plays an important role, since it controls the spread of the ensemble and therefore the stability of the filter. Further, our numerical results show that the LMCPF is a particle filter with a quality comparable to the LETKF for state-of-the-art real-world opera-

tional global atmospheric NWP forecasting systems. This will be demonstrated by numerical experiments based on an implementation of the particle filter in the operational data assimilation software suite of DWD.

The LMCPF is introduced in section 4.1, where we start in section 4.1.1 with a summary of the ingredients we use from the LAPF of chapter 3.1. Then, an elementary Gaussian filtering step in ensemble space is described in Section 4.1.2. Finally, the full LMCPF method is presented in Section 4.1.3. We describe the high-dimensional experimental environment for our development and evaluation framework for numerical tests in Section 4.2. The numerical results for the global weather forecasting model ICON are shown in Section 4.3. We study the statistics of the relationship of observations and the ensemble as well as the corresponding statistics of the shift vectors of the Gaussian particles of the LMCPF. We show the large improvements with respect to standard NWP scores which the LMCPF can achieve compared to the LAPF. Additionally, we present case studies comparing the LMCPF forecast scores to the operational LETKF.

4.1 The Localised Mixture Coefficients Particle Filter (LMCPF)

Our setup for data assimilation is to employ an ensemble $\{x^{(b,\ell)} \in \mathbb{R}^n, \ell = 1, ..., L\}$ of states, which are used to estimate or approximate $p^{(b)}(x)$ (corresponding to Equation (2.1.5) (page 17), see Chapter (2.1) for an Introduction into Bayes Theorem). The basic analysis step of data assimilation is to construct an analysis ensemble $\{x^{(a,\ell)} \in \mathbb{R}^n, \ell = 1, ..., L\}$ of analysis states, which approximate $p^{(a)}(x)$ in a way consistent with the approximation of $p^{(b)}(x)$ by $x^{(b,\ell)}$, $\ell = 1, ..., L$. The above idea is common to both the Ensemble Kalman Filter (EnKF) and to particle filters. We employ the notation

$$\mathbf{X}^{(b)} := \left(x^{(b,1)} - \overline{x}, ..., x^{(b,L)} - \overline{x} \right) \in \mathbb{R}^{n \times L} \tag{4.1.1}$$

for the matrix of ensemble differences to the ensemble mean $\overline{x}$ defined by

$$\overline{x} := \frac{1}{L}\sum_{\ell=1}^{L} x^{(b,\ell)} \in \mathbb{R}^n. \tag{4.1.2}$$

For the ensemble differences in observation space we employ

$$\mathbf{Y}^{(b)} := \left(y^{(b,1)} - \overline{y}, ..., y^{(b,L)} - \overline{y}\right) \in \mathbb{R}^{m\times L} \tag{4.1.3}$$

with the mean $\overline{y}$ defined by

$$\overline{y} := \frac{1}{L}\sum_{\ell=1}^{L} y^{(b,\ell)} \in \mathbb{R}^m \tag{4.1.4}$$

and

$$y^{(b,\ell)} := H(x^{(b,\ell)}). \tag{4.1.5}$$

From now on we will use $\mathbf{X}$ for $\mathbf{X}^{(b)}$ and $\mathbf{Y}$ for $\mathbf{Y}^{(b)}$ for brevity. In the case of a linear observation operator we have $\overline{y} = \mathbf{H}\overline{x}$ and $\mathbf{Y} = \mathbf{HX}$. Usually, for EnKFs, the approximation of the covariance matrix is chosen to be based on the estimator

$$\mathbf{B} := \frac{1}{L-1}\sum_{\ell=1}^{L}(x^{(\ell)} - \overline{x}) \cdot (x^{(\ell)} - \overline{x})^T \in \mathbb{R}^{n\times n}. \tag{4.1.6}$$

The estimator $\mathbf{B}$ can also be written as $\mathbf{B} = \frac{1}{L-1}\mathbf{X}\mathbf{X}^T$. Usually, in this case the prior is approximated by

$$p^{(b)}(x) = c_B e^{-\frac{1}{2}(x-\overline{x})^T\mathbf{B}^{-1}(x-\overline{x})} \tag{4.1.7}$$

with $\mathbf{B}^{-1}$ well defined[1] for all $x = \overline{x}+\mathbf{X}\beta$ with some vector $\beta \in \mathbb{R}^L$. We note that, in general $\mathbf{B}$ is rank deficient and can't be inverted. However, by the arguments

[1]The standard arguments, see Lemma 3.2.1 of Nakamura and Potthast (2015), show injectivity of XX^T on $R(X)$: $XX^TX\beta = 0$ with $\beta \in \mathbb{R}^L$ yields $X^TX\beta \in N(X) \cap R(X^T) = R(X^T)^\perp \cap R(X^T)$, thus $X^TX\beta = 0$. The same argument for $X\beta \in N(X^T)$ yields $X\beta = 0$, thus XX^T is injective on $R(X)$. For surjectivity we consider $v \in R(X)$, i.e. $v = Xw$ with $w \in \mathbb{R}^L = N(X) \oplus N(X)^\perp = N(X) \oplus R(X^T)$, such that $w = w_1 + w_2$ with $w_1 \in N(X)$ and $w_2 = X^T\beta$ with some $\beta \in \mathbb{R}^n = R(X) + R(X)^\perp$. Repeating the last argument leads to a $\beta_1 \in R(X)$ with $w = X^T\beta_1$ and thus surjectivity. Invertibility of B is thus shown.

of Nakamura and Potthast (2015) on the given subspace inversion is possible. The normalisation constant c_B can be calculated based on a matrix $\Phi \in \mathbb{R}^{L\times\tilde{L}}$ which consists of an orthonormal basis of $N(X)^\perp \subset \mathbb{R}^L$ of dimension $\tilde{L} < L$ by

$$c_B := \left(\int_{\mathbb{R}^{\tilde{L}}} e^{-\frac{1}{2}(X\Phi\alpha)^T \mathbf{B}^{-1}(X\Phi\alpha)} \sqrt{\det(\Phi^T X^T X \Phi)}\, d\alpha \right)^{-1}, \tag{4.1.8}$$

where $\det\left(\Phi^T X^T X \Phi\right)$ is the Gramian of the injective mapping $X\Phi : \mathbb{R}^{\tilde{L}} \to \mathbb{R}^n$, i.e. the determinant of the Gram matrix $\Phi^T X^T X \Phi$. The approximation of the classical particle filter is

$$p^{(b)}(x) = c \sum_{\ell=1}^{L} \delta(x - x^{(b,\ell)}), \quad x \in \mathbb{R}^n, \tag{4.1.9}$$

with the δ-distribution $\delta(\cdot)$ and a normalisation constant $c = 1/L$. A well-known idea is to employ Gaussian mixtures (c.f. Hoteit et al. (2008); Liu et al. (2016a,b)), i.e. use the approximation

$$p^{(b)}(x) = c \sum_{\ell=1}^{L} c_\ell e^{-\frac{1}{2}(x-x^{(b,\ell)})^T \mathbf{G}_\ell^{-1}(x-x^{(b,\ell)})}, \tag{4.1.10}$$

where $\mathbf{G}_\ell \in \mathbb{R}^{n\times n}$ is some symmetric and positive definite matrix which describes the uncertainty of the individual particle, $c_\ell = 1/\sqrt{(2\pi)^n \det(\mathbf{G}_\ell)}$ is a normalisation constant for each of the Gaussians under consideration and c is an overall normalisation constant.

- The matrix $\mathbf{G}_\ell$ is the covariance of each Gaussian and can be seen as a measure for the short-range forecast error consisting of model error and some of the uncertainty in the initial conditions beyond the distribution of the ensemble of particles itself. We will discuss the important role of $\mathbf{G}_\ell$ in several places later, when we describe the LMCPF and its numerical realisation. In particular, we will investigate the situation where $\mathbf{G}_\ell$ is a multiple of the covariance matrix $\mathbf{B}$ defined above.

- The Gaussian mixture filter can be seen as a generalisation of the classical

particle filter, where instead of a δ-distribution a Gaussian around each prior particle is employed to calculate the posterior distribution and draw from it. Here, we will employ localisation and adaptivity as developed for the LAPF in combination with the mixture concept within the LMCPF.

4.1.1 The Localised Adaptive Particle Filtering Ingredients and Preparations

The goal of this section is to collect, prepare and summarize all components employed for the LMCPF. For the following derivation we assume linearity of $\mathbf{H}$, we will discuss the form of the equations in the case of non-linear $\mathbf{H}$ later. Then, we have $\mathbf{Y}^T = \mathbf{X}^T\mathbf{H}^T$ and with $\gamma = \frac{1}{L-1}$ the standard estimator for the covariance matrix is given by $\mathbf{B} = \gamma\mathbf{X}\mathbf{X}^T$. We will later use $\mathbf{B}$ as measure of uncertainty of individual particles, then using the scaling

$$\gamma = \frac{\kappa}{(L-1)} \tag{4.1.11}$$

with a parameter $\kappa > 0$ scaling the standard covariance matrix. Following standard arguments as in Hunt et al. (2007); Nakamura and Potthast (2015) or Potthast et al. (2019), this leads to the Kalman gain

$$\begin{aligned} \mathbf{K} &= \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \\ &= \gamma\mathbf{X}\mathbf{X}^T\mathbf{H}^T(\mathbf{R} + \gamma\mathbf{H}\mathbf{X}\mathbf{X}^T\mathbf{H}^T)^{-1} \\ &= \gamma\mathbf{X}\mathbf{Y}^T(\mathbf{R} + \gamma\mathbf{Y}\mathbf{Y}^T)^{-1} \end{aligned} \tag{4.1.12}$$

with invertible observation error covariance matrix $\mathbf{R} \in \mathbb{R}^{m\times m}$. We note that we have

$$(\mathbf{I} + \gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})\mathbf{Y}^T = \mathbf{Y}^T\mathbf{R}^{-1}(\mathbf{R} + \gamma\mathbf{Y}\mathbf{Y}^T) \tag{4.1.13}$$

by elementary calculations. We also note that $\mathbf{I} + \gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}$ is invertible on $\mathbb{R}^L$ and $\mathbf{R}+\gamma\mathbf{Y}\mathbf{Y}^T$ is invertible on $\mathbb{R}^m$ by assumption on the invertibility of $\mathbf{R}$. Then,

multiplying (4.1.13) by $(\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}$ from the left and by $(\mathbf{R}+\gamma\mathbf{Y}\mathbf{Y}^T)^{-1}$ from the right we obtain

$$\mathbf{Y}^T(\mathbf{R}+\gamma\mathbf{Y}\mathbf{Y}^T)^{-1} = (\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{R}^{-1}. \tag{4.1.14}$$

Now, (4.1.14) can be used to transform (4.1.12) into

$$\mathbf{K} = \gamma\mathbf{X}(\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{R}^{-1}. \tag{4.1.15}$$

This can be used to calculate the covariance update step of the Kalman filter in ensemble space as follows. We derive

$$\begin{aligned}
\mathbf{B}^{(a)} &= (\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{B}^{(b)} \\
&= \Big(\mathbf{I}-\gamma\mathbf{X}(\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{H}\Big)\gamma\mathbf{X}\mathbf{X}^T \\
&= \mathbf{X}\Big(\mathbf{I}-\gamma(\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}\Big)\gamma\mathbf{X}^T \\
&= \mathbf{X}\Big((\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\Big[\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}-\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}\Big]\Big)\gamma\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\gamma\mathbf{X}^T \\
&= \gamma\mathbf{X}(\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\mathbf{X}^T.
\end{aligned} \tag{4.1.16}$$

For collecting the formulas we now move back to using $\mathbf{X}^{(b)}$ for $\mathbf{X}$. The analysis ensemble $\mathbf{X}^{(a)}$ which generates the correct posterior covariance by $\mathbf{B}^{(a)} = \gamma\mathbf{X}^{(a)}(\mathbf{X}^{(a)})^T$ is given by

$$\mathbf{X}^{(a)} := \mathbf{X}^{(b)}\Big(\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}\Big)^{-\frac{1}{2}} \in \mathbb{R}^{n\times L}, \tag{4.1.17}$$

where the matrix $\mathbf{I}+\gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y} \in \mathbb{R}^{L\times L}$ lives in ensemble space, it is symmetric and invertible by construction, for all $\gamma > 0$.

The localized ensemble transform Kalman filter (LETKF) following Hunt et al. (2007) based on the square root filter for calculating the analysis ensemble can

be written as

$$\overline{x}^{(a)} := \overline{x}^{(b)} + \gamma \mathbf{X}^{(b)} w = \overline{x}^{(b)} + \mathbf{K}(y - \overline{y}) \tag{4.1.18}$$

with

$$w := (\mathbf{I} + \gamma \mathbf{Y}^T \mathbf{R}^{-1} \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{R}^{-1} (y - \overline{y}) \in \mathbb{R}^L \tag{4.1.19}$$

and

$$\mathbf{X}^{(a)} := \mathbf{X}^{(b)} \mathbf{W} \tag{4.1.20}$$

with

$$\mathbf{W} := (\mathbf{I} + \gamma \mathbf{Y}^T \mathbf{R}^{-1} \mathbf{Y})^{-\frac{1}{2}} \in \mathbb{R}^{L \times L}. \tag{4.1.21}$$

The above equations are carried out at each analysis grid point where the matrix $\mathbf{R}$ is localized by multiplication of each entry with a localization function depending on the distance of the variable to the analysis grid point Hunt et al. (2007). Using

$$\mathbf{X}^{(a,full)} := \left(x^{(a,1)}, ..., x^{(a,L)} \right) = (\overline{x}^{(a)} + x^{(a)}) \in \mathbb{R}^{n \times L} \tag{4.1.22}$$

the full update of the LETKF ensemble can be written as

$$\mathbf{X}^{(a,full)} = \overline{x}^{(b)} + \gamma \mathbf{X}^{(b)} w + \mathbf{X}^{(b)} \mathbf{W}, \tag{4.1.23}$$

where we define the sum of a vector (here $\overline{x}^{(b)}$ or $\gamma \mathbf{X}^{(b)} w$) plus a matrix (here $\mathbf{X}^{(b)} \mathbf{W}$) by adding the vector to each column of the matrix.

For non-linear observation operator H as in (18) of Hunt et al. (2007) the operator $\mathbf{K}$ is defined by the last line of (4.1.12), see also (4.1.15) and the ensemble transform by (4.1.20) with $\mathbf{W}$ by (4.1.21). This basically corresponds to an approximate linearisation of H in observation space based on the differences $y^{(b,\ell)} - \overline{y}$.

4.1.2 An Elementary Gaussian Filtering Step in Ensemble Space

Let us consider a Bayesian assimilation step (2.1.5) based on the approximation of the prior $p^{(b)}(x)$ as a Gaussian mixture (4.1.10). We first describe the steps in general, then derive the ensemble space version of the equations. To each particle, we attribute a distribution with covariance $\mathbf{G}$, i.e. we define

$$p^{(b,\ell)}(x) := \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{G})}} e^{-\frac{1}{2}(x-x^{(b,\ell)})^T \mathbf{G}^{-1}(x-x^{(b,\ell)})}, \quad x \in \mathbb{R}^n, \tag{4.1.24}$$

which is normalised according to equation (4.5.28) of Nakamura and Potthast (2015). Then, the full prior is a Gaussian mixture

$$p^{(b)}(x) = c \sum_{\ell=1}^{L} c_\ell e^{-\frac{1}{2}(x-x^{(b,\ell)})^T \mathbf{G}^{-1}(x-x^{(b,\ell)})}, \quad x \in \mathbb{R}^n, \tag{4.1.25}$$

with $c_\ell := 1/\sqrt{(2\pi)^n \det(\mathbf{G})}$ (i.e. we choose the variance uniform for all ℓ) and with some normalisation constant $c = \frac{1}{L}$ in this case. Bayes formula leads to the posterior distribution

$$p^{(a)}(x) = \tilde{c} \sum_{\ell=1}^{L} c_\ell \Big(e^{-\frac{1}{2}(x-x^{(b,\ell)})^T \mathbf{G}^{-1}(x-x^{(b,\ell)})} e^{-\frac{1}{2}(y-H(x))^T \mathbf{R}^{-1}(y-H(x))} \Big), \tag{4.1.26}$$

$x \in \mathbb{R}^n$, with a normalisation constant $\tilde{c}$, here different from the normalisation constant in (4.1.25). We note that the terms in round brackets constitute individual Gaussian assimilation steps. In the case where H is linear or approximated by its linearisation $\mathbf{H}$, the posterior of each of these terms can be explicitly calculated the same way as for the Ensemble Kalman Filter. Following Nakamura and Potthast (2015), Section 5.4, we define

$$x^{(a,\ell)} := x^{(b,\ell)} + \mathbf{G}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{G}\mathbf{H}^T)^{-1}(y - H(x^{(b,\ell)})), \quad \ell = 1, ..., L, \tag{4.1.27}$$

and

$$\mathbf{K} = \mathbf{G}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{G}\mathbf{H}^T)^{-1}, \;\; \mathbf{G}^{(a)} := (I - \mathbf{K}\mathbf{H})\mathbf{G}. \tag{4.1.28}$$

Then, we know that

$$\begin{aligned} q^{(a,\ell)}(x) &= \tilde{c}c_\ell e^{-\frac{1}{2}(x-x^{(b,\ell)})^T\mathbf{G}^{-1}(x-x^{(b,\ell)})}e^{-\frac{1}{2}(y-H(x))^T\mathbf{R}^{-1}(y-H(x))} \\ &= \tilde{c}c_\ell e^{-\frac{1}{2}(x-x^{(a,\ell)})^T[\mathbf{G}^a]^{-1}(x-x^{(a,\ell)})}e^{-\frac{1}{2}(y-H(x^{(b,\ell)}))^T(\mathbf{H}\mathbf{G}\mathbf{H}^T+\mathbf{R})^{-1}(y-H(x^{(b,\ell)}))} \\ &= \tilde{c}c_\ell w_\ell e^{-\frac{1}{2}(x-x^{(a,\ell)})^T[\mathbf{G}^{(a)}]^{-1}(x-x^{(a,\ell)})}, \;\; x \in \mathbb{R}^n, \end{aligned} \tag{4.1.29}$$

with constants $c_\ell := 1/\sqrt{(2\pi)^n \det(\mathbf{G})}$ (see also equation (40) in Schenk et al. (2022)) and w_ℓ given by (4.0.3). Since c_ℓ does not depend on ℓ, the constants are irrelevant for the resampling step and will be removed by the normalisation step. Note that the constants w_ℓ, $\ell = 1, ..., L$, are extremely important, since they contain the relative weights of the individual posterior particles with respect to each other. They should not be ignored! Here, we first describe the full posterior distribution, which is now given by

$$p^{(a)}(x) = \tilde{c}\sum_{\ell=1}^{L} c_\ell w_\ell e^{-\frac{1}{2}(x-x^{(a,\ell)})^T[\mathbf{G}^{(a)}]^{-1}(x-x^{(a,\ell)})}, \;\; x \in \mathbb{R}^n. \tag{4.1.30}$$

In the case of the classical particle filter, the Gaussians $c_\ell e^{-\frac{1}{2}(x-x^{(b,\ell)})^T\mathbf{G}^{-1}(x-x^{(b,\ell)})}$ become δ-distributions $c_\ell\delta(x - x^{(b,\ell)})$ with weights $c_\ell = 1$. In this case, the individual posterior weights w_ℓ are given by the likelihood of observations

$$w_\ell := e^{-\frac{1}{2}(y-H(x^{(b,\ell)}))^T\mathbf{R}^{-1}(y-H(x^{(b,\ell)}))}, \;\; \ell = 1, ..., L. \tag{4.1.31}$$

This choice will also be a reasonable approximation in the case of small variance $\mathbf{G}$ of the Gaussians under consideration in comparison with the distance $y - H(x^{(b,\ell)})$. In the general Gaussian case, the weights can be calculated from (4.0.3). For our numerical experiments we use non-zero G with some positive variance, and tested both the exact weights (4.0.3) or approximate weights w_l given by (4.1.31).

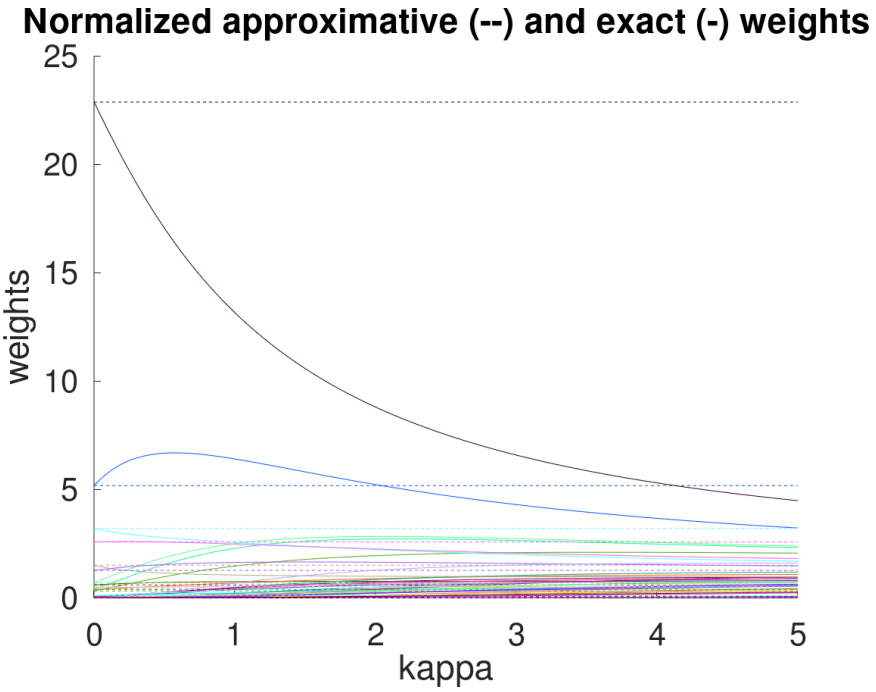

Figure 4.1: We show a comparison between the normalised approximative weights calculated as in (4.1.31) versus the normalised exact calculated weights (4.0.3). The solid lines show the normalised exact determined weights and the dashed lines the normalised approximative weights. The colors vary for different ensemble members (L=40). On the x-axis we show the value for κ of equation (4.1.11), on the y-axis the values of the weights.

In Figure 4.1 we show a comparison of the normalised approximative weights (4.1.31) as dashed lines and the normalised exact determined weights (4.0.3) as solid lines, for a selected point of the full NWP model described in Sections 4.2 and 4.3. Here, each ensemble member (L=40) is described by a different color. For this plot we varied the parameter κ, described in equation (4.1.11), between 0 and 5. Figure 4.1 shows how the normalised approximative weights differ from the normalised exact weights. We have carried out experiments both with the exact and approximate weights, we found that overall the results with exact weights show a better performance.

Let us now describe the ensemble space transformation of the above equations. The ensemble space as a subset of the state space is spanned by $\mathbf{X}$ given in (4.1.1). Our approach for the model error covariance is $\gamma\mathbf{X}\mathbf{X}^T$ with some scaling factor γ. We note that for the LETKF, $\gamma = \frac{1}{L-1}$. Here, $\gamma > 0$ can be any real number. We will provide some estimates for what γ can be in a global NWP model setup in our numerical part in Section 4.3. In the transformed space this leads to the covariance $\gamma\mathbf{I} \in \mathbb{R}^{L\times L}$ to be used for the ensemble transform version of (4.1.26). Recall the ensemble transformation $x - \overline{x} = \mathbf{X}\beta$, $x^{(\ell)} - \overline{x} = \mathbf{X}e_\ell$ and $x - x^{(\ell)} = \mathbf{X}(\beta - e_\ell)$ for $\ell = 1, ..., L$, where e_ℓ is the standard unit vector with one in its ℓ-th component and zero otherwise leading to

$$\begin{aligned}(x - x^{(\ell)})^T(\gamma\mathbf{X}\mathbf{X}^T)^{-1}(x - x^{(\ell)}) &= (\beta - e_\ell)^T\gamma^{-1}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}(\beta - e_\ell) \\ &= (\beta - e_\ell)^T\gamma^{-1}\mathbf{I}(\beta - e_\ell). \qquad (4.1.32)\end{aligned}$$

We note that $\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X} = \mathbf{I}$ is true only on the subspace $N(\mathbf{X})^\perp$, but we can employ the arguments used to justify equation (15) of Hunt et al. (2007) to use the covariance $\gamma^{-1}\mathbf{I}$ in ensemble space for the prior term. For the observation

error term of (4.1.26) in ensemble space $\mathbb{R}^L$ we use (3.1.3), i.e. we have

$$q^{(a,\ell)}(\beta) = \hat{c}c_\ell e^{-\frac{1}{2}(\beta-e_\ell)^T(\gamma^{-1}\mathbf{I})(\beta-e_\ell)} e^{-\frac{1}{2}[P(y-\bar{y}-\mathbf{Y}\beta)]^T\mathbf{R}^{-1}[P(y-\bar{y}-\mathbf{Y}\beta)]}, \quad \beta \in \mathbb{R}^L, \tag{4.1.33}$$

with norming constant $\hat{c}$, for $\ell = 1, ..., L$, where P is the orthogonal projection onto span$\{\mathbf{Y}\}$ with respect to the scalar product in $\mathbb{R}^m$ weighted by $\mathbf{R}^{-1}$; it is defined in (3.1.2) and Lemma 3.2.3 of Nakamura and Potthast (2015) to be given by

$$P = \mathbf{Y}(\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{R}^{-1}. \tag{4.1.34}$$

As in (3.1.5) - (3.1.7) the right-hand side of (4.1.33) can be transformed into

$$q^{(a,\ell)}(\beta) = \hat{c}c_\ell e^{-\frac{1}{2}(\beta-e_\ell)^T(\gamma^{-1}\mathbf{I})(\beta-e_\ell)} e^{-\frac{1}{2}[C-\beta]^T\mathbf{A}[C-\beta]}, \quad \ell = 1, ..., L, \tag{4.1.35}$$

with

$$\mathbf{A} := \mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}, \qquad C := \mathbf{A}^{-1}\mathbf{Y}^T\mathbf{R}^{-1}(y - \overline{y}). \tag{4.1.36}$$

We now carry out (4.1.27) and (4.1.28) in ensemble space based on (4.1.12) and (4.1.13), leading to the new mean of the posterior distribution for the ℓ-th particle prior distribution

$$\beta^{(a,\ell)} = e_\ell + \gamma(\mathbf{I} + \gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}(C - e_\ell) \tag{4.1.37}$$

and the new covariance matrix of this distribution

$$\mathbf{G}_{ens}^{(a)} = (\frac{1}{\gamma}\mathbf{I} + \mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1} \in \mathbb{R}^{L\times L} \tag{4.1.38}$$

independent of ℓ when $\mathbf{G} = \gamma\mathbf{X}\mathbf{X}^T$ is independent of ℓ. This means that we obtain

$$q^{(a,\ell)}(\beta) = \hat{c}c_\ell w_\ell e^{-\frac{1}{2}(\beta-\beta^{(a,\ell)})^T\mathbf{G}_{ens}^{(a)}(\beta-\beta^{(a,\ell)})}, \quad \beta \in \mathbb{R}^L \tag{4.1.39}$$

with $\beta^{(a,\ell)}$ given by (4.1.37) and $\mathbf{G}_{ens}^{(a)}$ given by (4.1.38) for the posterior distri-

bution of the ℓ-th particle in ensemble space. We denote the term

$$\beta^{(shift,\ell)} := \gamma(\mathbf{I} + \gamma\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}(C - e_\ell) \tag{4.1.40}$$

as the shift vector for the ℓ-th particle in ensemble space, i.e. $\beta^{(a,\ell)} = e_\ell + \beta^{(shift,\ell)}$ in (4.1.37). The use of the model error $\gamma\mathbf{I}$ corresponding to $\gamma\mathbf{X}\mathbf{X}^T$ for this particle in ensemble space leads to this shift in the analysis. The shift has important effects:

1. it moves the particle towards the observation in ensemble space,
2. by the use of particle uncertainty, it constitutes a further degree of freedom which can be used for tuning of a real system.

One of the major advantages and problems at the same time of the LAPF as well as a classical particle filter is that the particles are taken as they are. If the model has some local bias, i.e. if all particles have a similar behaviour and do not fit the observation well, then there is no inherent tool in the classical particle filter or the basic LAPF to move the particles towards the observation - this move is only achieved by selection of the best particles, closest to the observation. By resampling and rejuvenation, effectively the whole ensemble is moved towards the observation. Here, with the introduction of uncertainty of individual particles into the assimilation step, this is already carried out for each individual particle by calculating a posterior mean $\beta^{(a,\ell)}$ in (4.1.37) of the posterior component $q^{(a,\ell)}(\beta)$ given by (4.1.39) for the model error prior distribution $q^{(b,\ell)}(x)$ attributed to each particle (4.1.24).

4.1.3 Putting it all together: the full LMCPF

Here, we now collect all steps to describe the full LMCPF assimilation step and data assimilation cycle. The LMCPF assimilation cycle is run analogously to the LETKF or LAPF assimilation cycle, i.e. we start with some initial ensemble $x_0^{(a,\ell)}$ at time t_0. Then, for time steps t_k, $k = 1, 2, 3, ...$ we

(1) carry out a propagation step, i.e. we run the model forward from time t_{k-1} to t_k for each ensemble member, leading to the background ensemble $x_k^{(b,\ell)}$ at time t_k.

(2) Then, at each localisation point ξ on a coarser analysis grid $\mathcal{G}$ we carry out the localised ensemble transform (4.1.36), calculating C and $\mathbf{A}$. Localisation is carried out as for the LETKF and LAPF, i.e. the matrix $\mathbf{R}$ is weighted depending of the distance of each of its observations to the analysis point.

(3) We now carry out a classical resampling step following Section 3.1.3. This leads to a matrix

$$\breve{\mathbf{W}}_{i,\ell} = \begin{cases} 1, & \text{if } R_\ell \in (w_{ac_{i-1}}, w_{ac_i}], \\ 0, & \text{otherwise,} \end{cases} \tag{4.1.41}$$

$i, \ell = 1, ..., L$, draw $r_\ell \sim U([0,1])$, set $R_\ell = \ell - 1 + r_\ell$, with accumulated weights w_{ac}, $w_{ac_0} = 0, w_{ac_i} = w_{ac_{i-1}} + w_{k,i}$, $w_{k,i} := p(\mathbf{y}_k | x^{(b,i)})$ and $\breve{\mathbf{W}} \in \mathbb{R}^{L \times L}$ defined by (4.1.41) with entries one or zero reflecting the choice of particles. As for the LETKF and LAPF this is carried out at each localisation point ξ on a coarser analysis grid $\mathcal{G}$ to ensure that the weight matrices only change on scales on the order of the localisation length scale. Here, we use $\breve{\mathbf{W}}$ instead of $\breve{\mathbf{W}}(\xi)$ for brevity.

(4) The posterior matrix $\mathbf{G}_{ens}^{(a)}$ given by (4.1.38) and the shift vectors $\beta^{(shift,\ell)}$ given by (4.1.40) for $\ell = 1, ..., L$ are calculated for each localisation point ξ. We define

$$\mathbf{W}^{(shift)} := \left(\beta^{(shift,1)}, ..., \beta^{(shift,L)} \right) \in \mathbb{R}^{L \times L}. \tag{4.1.42}$$

Then, if we want the shift given by the lth-particle, we obtain it by the product $\mathbf{W}^{(shift)} e_\ell$. If we have a selection matrix $\breve{\mathbf{W}}$ for which each column with index ζ, $\zeta = 1, ..., L$, contains some particle e_ℓ with $\ell = \ell(\zeta)$, which

has been chosen to be the basis for the corresponding new particle, we obtain the shifts for these particles by the product $\mathbf{W}^{(shift)}\breve{\mathbf{W}}$. According to the analysis equation (4.1.37) the new coordinates in ensemble space are calculated by

$$\left(\beta^{(a,1)}, ..., \beta^{(a,L)}\right) = \breve{\mathbf{W}} + \mathbf{W}^{(shift)}\breve{\mathbf{W}}. \tag{4.1.43}$$

(5) For each particle we now carry out an adaptive Gaussian resampling or rejuvenation step. The rejuvenation is carried out the same way as described in Sections 3.1.4 and 3.1.5, i.e. we first calculate

$$\rho = \frac{\mathbf{d}_{o-b}^T \mathbf{d}_{o-b} - Tr(\mathbf{R})}{Tr(\mathbf{H}\frac{1}{L-1}\mathbf{X}\mathbf{X}^T\mathbf{H}^T)} \tag{4.1.44}$$

at each localisation point, with the actual ensemble covariance matrix $\frac{1}{L-1}\mathbf{X}\mathbf{X}^T$ and with the observation minus background statistics $\mathbf{d}_{o-b} = y_k - \bar{y}_k$ where $\bar{y}_k$ denotes the ensemble mean in observation space described in (4.1.4) at time t_k[2]. Then we scale ρ by some function

$$\sigma(\rho) := \begin{cases} c_0, & \rho < \rho^{(0)}, \\ c_0 + (c_1 - c_0)\frac{\rho - \rho^{(0)}}{\rho^{(1)} - \rho^{(0)}}, & \rho^{(0)} \leq \rho \leq \rho^{(1)}, \\ c_1, & \rho > \rho^{(1)}, \end{cases} \tag{4.1.45}$$

where the constants $\rho^{(0)}, \rho^{(1)}, c_0, c_1$ are tuning constants. We note that temporal smoothing is applied to ρ as usual for LETKF or LAPF. Let $\mathbf{N} \in \mathbb{R}^{L\times L}$ be a matrix with entries drawn from a normal distribution, i.e. each entry is taken from a Gaussian distribution with mean zero and variance 1. This is chosen uniformly for all localisation points ξ on the analysis grid $\mathcal{G}$. Then, the rejuvenation plus shift step is carried out by

$$\mathbf{W} := \breve{\mathbf{W}} + \mathbf{W}^{(shift)}\breve{\mathbf{W}} + [\mathbf{G}_{ens}^{(a)}]^{\frac{1}{2}}\mathbf{N}\sigma. \tag{4.1.46}$$

[2]The R matrix is taken from operations, where it is estimated based on standard Desrozier statistics. Usually ρ is kept between a minimal and maximal positive value, e.g. 0.7 and 1.5 for operations to account for statistical outliers in the estimator.

Again, we note that $\mathbf{W} = \mathbf{W}(\xi)$, $\mathbf{W}^{(shift)} = \mathbf{W}^{(shift)}(\xi)$, $\breve{\mathbf{W}} = \breve{\mathbf{W}}(\xi)$, $[\mathbf{G}_{ens}^{(a)}]^{\frac{1}{2}} = [\mathbf{G}_{ens}^{(a)}]^{\frac{1}{2}}(\xi)$ and $\sigma = \sigma(\xi)$ are functions of physical space with $\xi \in \mathcal{G}$ chosen from the analysis grid $\mathcal{G}$.

(6) The matrices $\mathbf{W}$ are calculated at each analysis point ξ on a coarser global analysis grid $\mathcal{G}$. We now interpolate the matrices onto the full model grid $\mathcal{G}_{model}$.

(7) Finally we calculate the analysis ensemble (4.1.22) by

$$\begin{aligned} \mathbf{X}^{(a,full)} &= \overline{x}^{(b)} + \mathbf{X}^{(b)}\mathbf{W} \qquad (4.1.47) \\ &= \overline{x}^{(b)} + \underbrace{\mathbf{X}^{(b)}\breve{\mathbf{W}}}_{class.\ resampling} + \underbrace{\mathbf{X}^{(b)}\mathbf{W}^{(shift)}\breve{\mathbf{W}}}_{shift} + \underbrace{\mathbf{X}^{(b)}[\mathbf{G}_{ens}^{(a)}]^{\frac{1}{2}}\mathbf{N}\sigma}_{adapt.\ Gauss.\ resampling} \end{aligned}$$

Comparing (4.1.47) with (4.1.23) we observe some similarities and some differences. The LETKF does not know the selection reflected by the matrix $\breve{\mathbf{W}}$, instead it transforms the ensemble by its matrix $\mathbf{W}$. Both know a shift term, for the LETKF it is given by w, for the LMCPF by $\mathbf{W}^{(shift)}\breve{\mathbf{W}}$, shifting each particle according to model error (here taken proportional to ensemble spread), where the LETKF shifts according to the full ensemble spread. The LMCPF also takes into account that part of the ensemble spread which is kept during the selection process. Further, it employs adaptive resampling around each remaining shifted particle. This helps to keep the filter stable and achieve an appropriate uncertainty described by $o - b$ statistics.

4.2 Experimental Environment: the Global ICON Model

4.2.1 The ICON Model

We have carried out experiments testing the LMCPF algorithm in the global ICON model, i.e. the operational global NWP model of DWD, compare Zängl

et al. (2014) for further details on the systems. ICON is based on an unstructured grid of triangles generated by subdivision from an initial icosahedron. The operational resolution is 13 km for the deterministic run and 26 km for the ensembles both for the data assimilation cycle and the ensemble prediction system (EPS). The upper air prognostic variables such as wind, humidity, cloud water, cloud ice, temperature, snow and precipitation live on 90 terrain-following vertical model levels from the surface up to 75 km height. In the operational setup, we have 265 million grid points. We also note that there are further prognostic variables on the surface and on eight soil levels, in particular soil temperature and soil water content, as well as snow variables, sea ice fraction, ice thickness and ice surface temperature of ICON's integrated sea-ice model.

The data assimilation for the operational ensemble is carried out by an LETKF based on Hunt et al. (2007). We run a data assimilation cycle with an analysis every 3 hours. Forecasts are calculated based on the analysis for 00 and 12 UTC, with 180 hours forecast lead time. For the operational system, forecasts with shorter lead times of 120 hours for 06 and 18 UTC and 30 hours for 03, 09, 15 and 21 UTC are calculated. The ensemble data assimilation cycle is run with L=40 members.

For the experimental setup of our study, we employ a slightly lower horizontal resolution of 52 km for the ensemble and 26 km for the deterministic run (in the operational setup a part of the observations quality control is carried out within the framework of the deterministic run, we keep this feature for our particle filter experiments). An incremental analysis update with a window of $t \in [-90\,\mathrm{min}, 90\,\mathrm{min}]$ around the analysis time for starting the model runs is used. The analysis is carried out for temperature, humidity and two horizontal wind components, i.e. for four prognostic variables per grid point. This leads to $n = 6.6 \cdot 10^6$ free variables at each ensemble data assimilation step. Forecasts are only carried out for 00 and 12 UTC. We employ L=40 members for the experimental runs as well. As we want to be able to compare the LMCPF

with the LAPF and take into account the latest improvements in the operational setup, we have run the LMCPF for two experimental periods - May 2016 for the comparison with the LAPF and January 2022 for the comparison with the newer near operational setup of the ICON model.

4.2.2 Comparison in an Operational Framework

For testing and developing algorithms in the operational framework, the tuning of basic algorithmic constants is a crucial part. The task of testing in a real-world operational setup is much more intricate than for what is usually done when algorithms are compared in a simulation-only small-scale environment. In particular for new algorithms, the whole model plus assimilation cycled NWP system needs a retuning and it is difficult to compare one algorithmic layer only within a very complex system with respect to its performance. To compare two algorithms A and B, there are two important points to be taken into account:

(1) **Tuning Status of the Methods.** There might be a raw or default version of the algorithms, but when you compare scores with the task of showing that some algorithm is better than the other, you need to compare tuned algorithms. In principle, you have to tune algorithm A to give the best results and then you have to tune algorithm B to give the best results and then compare the results of tuned A and tuned B. If A has been tuned for several years, but B is raw, the results give you insight into the tuning status of A and B, but not necessarily of the algorithms as such! So we have to be very careful with generic conclusions.

(2) **Quality Control of Observations.** When you compare two algorithms for assimilation or two models, verification provides a variety of scores. But verification with real data needs quality control of these data, since otherwise scores are mainly determined by outliers, and one broken device can make the whole verification result completely useless. But how is the data quality controlled? Usually we employ observation minus first guess

$(o-f)$ statistic and remove observations which are far away from the model first guess. This leads to an important point: each algorithm A and B needs to use its own quality control. If model biases change between A and B, you will have a different selection of 'good' observations.

But how do you compare two systems which employ different observations? One solution can be to use observations for comparison which passed both quality controls. A second method is to verify each algorithm separately and then compare the scores (this is what is done with World Meteorological Organization (WMO) score comparisons between global models). A third method is to try to use 'independent' observations. But these also need some quality control, and since they are not linked to any of the forecasting systems, it is unclear in what way their use in verification helps to judge a particular algorithm or to compare two algorithms.

For our experiments, we compare the LMCPF with the LAPF and the LETKF. The LETKF has a relatively advanced tuning status. LAPF has been mildly tuned and the LMCPF is relatively new. We carried out several tuning steps to try to make LMCPF and LETKF comparable. Further, we employ quality control for the observations in each system separately. Verification of the $o - f$ statistics is based on each system independently. Here, one important performance measure is the number of observations which passes the quality control. If these number is larger for B than for A, we can conclude that the system fits better to the observations, which is a good indicator for the quality of a short-range forecast. For comparison of forecasts the joint set of observations is used, those which pass both the quality control of algorithm A and algorithm B.

4.3 Numerical Results

The goal of this numerical part is, firstly, to investigate the relationship between the observation vector mapped into ensemble space and the ensemble distribu-

tion. Secondly, we show since the LMCPF moves particles based on the Gaussian uncertainty of individual particles, it bridges the gap between forecast ensemble and observations. Furthermore we study its distribution. The third part shows results of observation - first guess (o-f) statistics for the LMCPF with different choices for $\kappa > 0$ compared to the LETKF and the LAPF. Fourthly, we investigate the evolution of ensemble spread with different parameter settings. In the last part we demonstrate the feasibility of the LMCPF as a method for atmospheric analysis and subsequent forecasting in a very high-dimensional operational framework, demonstrating that it runs stably for a month of global atmospheric analysis and forecasting.

4.3.1 Distributions of Observations in Ensemble Space

In a first step, we study (a) the distance between the observation and the ensemble mean and (b) the minimum distance between the observation and the ensemble members. In ensemble space, for distance calculations an appropriate metric needs to be used. Recall that $\mathbb{R}^m$ with dimension m is the observation space and $\mathbb{R}^L$ with dimension L the ensemble space. Given a vector $\beta \in \mathbb{R}^L$ in ensemble space, the distance corresponding to the physical norm $\|\cdot\|_{R^{-1}}$ in observation space, which is relevant to the weight calculation of the particle filter, is calculated by

$$\begin{aligned} \|\mathbf{Y}\beta\|^2_{R^{-1}} &= \langle \mathbf{Y}\beta, \mathbf{Y}\beta \rangle_{R^{-1}} \\ &= \langle \mathbf{Y}\beta, \mathbf{R}^{-1}\mathbf{Y}\beta \rangle \\ &= (\mathbf{Y}\beta)^T \mathbf{R}^{-1}\mathbf{Y}\beta \\ &= \beta^T (\mathbf{Y}^T \mathbf{R}^{-1}\mathbf{Y})\beta \\ &= \langle \beta, \mathbf{A}\beta \rangle \\ &= ||\beta||^2_{\mathbf{A}} \end{aligned} \tag{4.3.1}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard L^2-scalar product in $\mathbb{R}^m$ or $\mathbb{R}^L$, respectively. The notation $\langle \cdot, \cdot \rangle_{\mathbf{D}}$ with some positive definite matrix $\mathbf{D}$ denotes the weighted scalar product $\langle \cdot, \mathbf{D}\, \cdot \rangle$ and $\|\cdot\|_{\mathbf{D}} = \langle \cdot, \cdot \rangle_{\mathbf{D}}$, here with either $\mathbf{R}^{-1}$ in $\mathbb{R}^m$ or $\mathbf{A}$ in $\mathbb{R}^L$. Note that for A to be positive definite we need $L \leq m$.

The matrix $\mathbf{A}$ including the standard LETKF localisation in observation space has been integrated into the data assimilation coding environment. Here, we show results from an LMCPF one month experiment studying one assimilation step at 0 UTC of May 6, 2016. The cycle has been started May 1, such that the results illustrate a situation where the spin-up period is over and LMCPF spread has reached a steady state (compare Figure 4.8).

At each analysis grid point ξ of some coarse global analysis grid $\mathcal{G}$ we have a matrix $\mathbf{A}$ (see Eq. (4.1.36)), $L = 40$ ensemble members and one projected observation vector $C \in \mathbb{R}^L$ (see Eq. (4.1.36)). This leads to a total of $N_\omega = 10890$ samples ω numbering the analysis grid points in a given height layer, e.g. for 850 hPa. The distance of the observations to the ensemble mean is given by

$$d_C(\omega) := \|C(\omega)\|_{\mathbf{A}(\omega)}, \tag{4.3.2}$$

where the metric A is chosen to be consistent with (4.1.35). The minimal distance of the observations vector to the ensemble members is given by

$$d_{min}(\omega) := \min_{j=1,\ldots,L} \|C(\omega) - e_j\|_{\mathbf{A}(\omega)}, \tag{4.3.3}$$

with $\omega = 1, ..., N_\omega$, where we employed (4.3.1) and where we note that in ensemble space the ensemble members $x^{(b,j)} - \overline{x}$ are given by the standard unit normal vectors e_j, $j = 1, ..., L$.

To analyse the role of moving particles towards the observation in ensemble space, in Figure 4.2 we show global histograms for d_C and d_{min} for three height levels of approximately 500 hPa, 850 hPa and 1000 hPa. When the distribution of both d_C and d_{min} are similar, i.e. the distribution of the minimal distance of

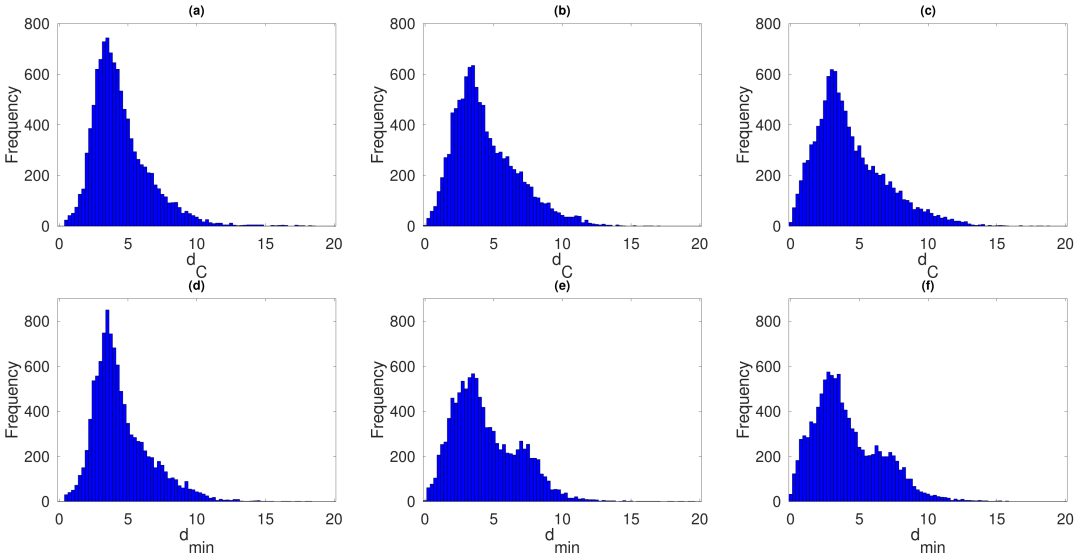

Figure 4.2: We show global histograms of d_C and d_{min} defined in (4.3.2) and (4.3.3) for three different pressure levels: 500 hPa in (a) and (d), 850 hPa in (b) and (e) and 1000 hPa in (c) and (f), with d_C in (a)-(c) and d_{min} in (d)-(f). Shown are statistics for the LMCPF with $\kappa = 25$ for May 6th, 0 UTC.

the observation to the ensemble members and the distribution of the distance of observations to the ensemble mean are comparable, it indicates that we have a well-balanced system. To understand the particular form of the distributions, we compare it with simulations of random draws of a Gaussian distribution in a 40 dimensional space shown in Figure 4.3. When you draw from a Gaussian with mean zero and standard deviation $\sigma = 4$, we obtain Figure 4.3 (a). The behaviour of the histograms of the norms of the points drawn changes significantly if we consider mixtures with different variances in different space directions. Figure 4.3 (a)-(e) shows different distributions with variances given by

$$\sigma_j = \frac{\eta}{j^\nu}, \quad j = 1, ..., L \tag{4.3.4}$$

where the constant $\eta \in (4, 15, 30, 40, 50)$ has been chosen to achieve a maximum around 4 and different decay exponents $\nu \in (0, 0.5, 1, 2, 3)$ have been tested. The distributions of Figure 4.2 correspond to a decay exponent between $\nu = 1$ and $\nu = 2$. How much is this reflected by the eigenvalue distributions for the matrices $\mathbf{A}$? We have carried out a fit to the eigenvalue decay of $\mathbf{A}$ for a selection

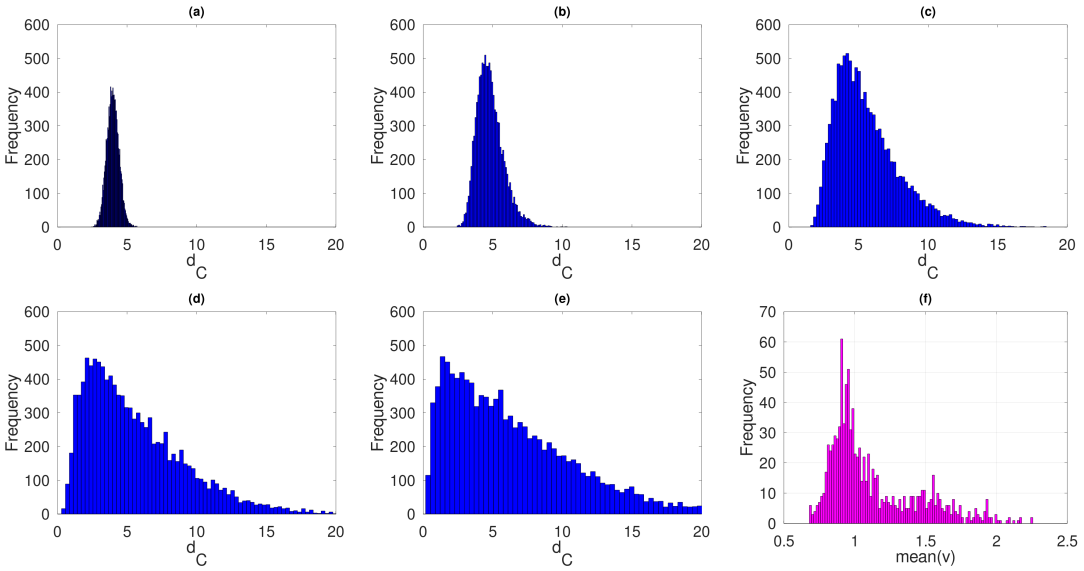

Figure 4.3: We show simulations of distributions of random draws in an $L = 40$ dimensional space, with different mixtures of variances given by formula (4.3.4), here with $\eta \in (4, 15, 30, 40, 50)$ and $\nu \in (0, 0.5, 1, 2, 3)$ in (a) to (e). A histogram of the fit of exponents ν as in (4.3.4) to the eigenvalue decay of the matrices $\mathbf{A}$ for a selection of 1000 points is shown in (f). The fit is obtained from the mean of exponents derived from formula (4.3.5).

of analysis points. The constant η is obtained by using $j = 1$, which leads to $\sigma_1 = \eta$. Taking the logarithm on both sides now yields

$$\nu \log(j) = \log(\eta) - \log(\sigma_j), \quad j = 2, ..., L. \tag{4.3.5}$$

A fit of ν can be obtained for example by division through $\log(j)$ and taking the mean of the remaining right-hand side. The distribution of the resulting exponents is displayed in Figure 4.3 (f). The results find exponents between 0.7 and 2.2. The corresponding distributions are those shown in Figure 4.3(c) and (d), which are quite close to the distributions of d_C found in the empirical particle-filter generated NWP ensemble in Figure 4.2.

4.3.2 The Move of Particles

At a second step, we want to investigate the capability of the LMCPF to move particles towards the observation by testing different choices of $\kappa > 0$ given by

(4.1.11). In Figure 4.4 we compare histograms of the norm of the mean ensemble shift in ensemble space for pressure level 500 hPa, determined for May 6th, 0 UTC. The four histograms show the statistics for the three filters in different settings: a) LAPF, b) LMCPF with $\kappa = 1$, c) LMCPF with $\kappa = 2.5$ and d) LMCPF with $\kappa = 25$.

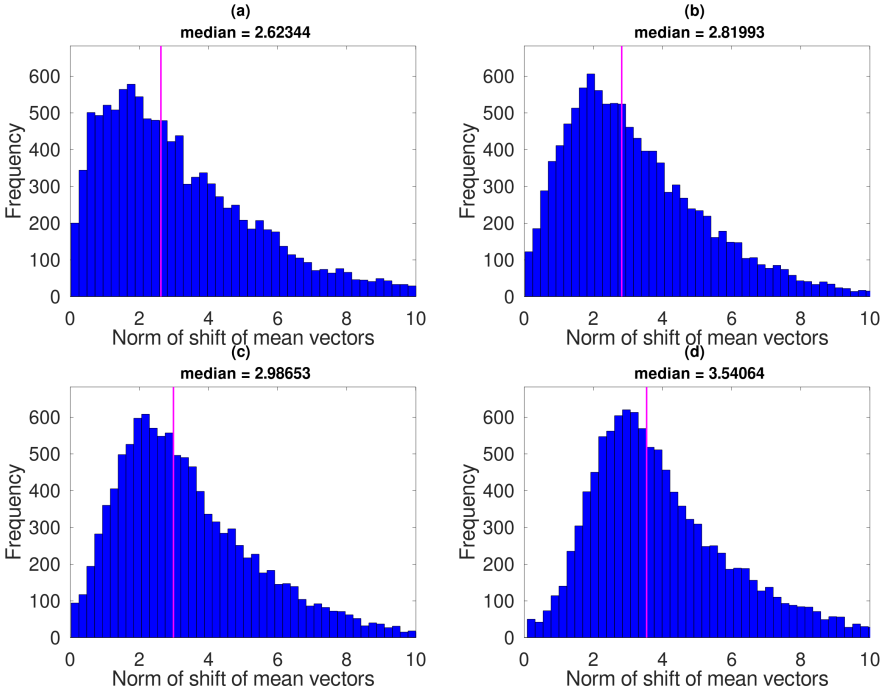

Figure 4.4: We show global histograms of the norm of the mean ensemble shift at pressure level at 500 hPa. On the x-axis we show the norm of shift of mean vectors in ensemble space and on the y-axis we show the frequency. We display the histogram for (a) the LAPF, (b) the LMCPF with $\kappa = 1$, (c) the LMCPF with $\kappa = 2.5$ and (d) shows the LMCPF with $\kappa = 25$. The pink line displays the median, which is also shown on the top of each plot. Shown are the statistics for May 6th, 0 UTC.

There are two effects seen in Figure 4.4. First, we see the distribution of average shifts or moves of the ensemble mean generated by the LAPF and the LMCPF with three different choices κ controlling the size of the uncertainty used for each particle. The mean shift increases if the uncertainty increases, i.e. from $\kappa = 1$ to $\kappa = 2.5$ and $\kappa = 25$. To develop an understanding of the relative size of this shift let us look at the one-dimensional version of formula (4.1.40) given by

$$s(\kappa) = \frac{\kappa b}{r + \kappa b}, \tag{4.3.6}$$

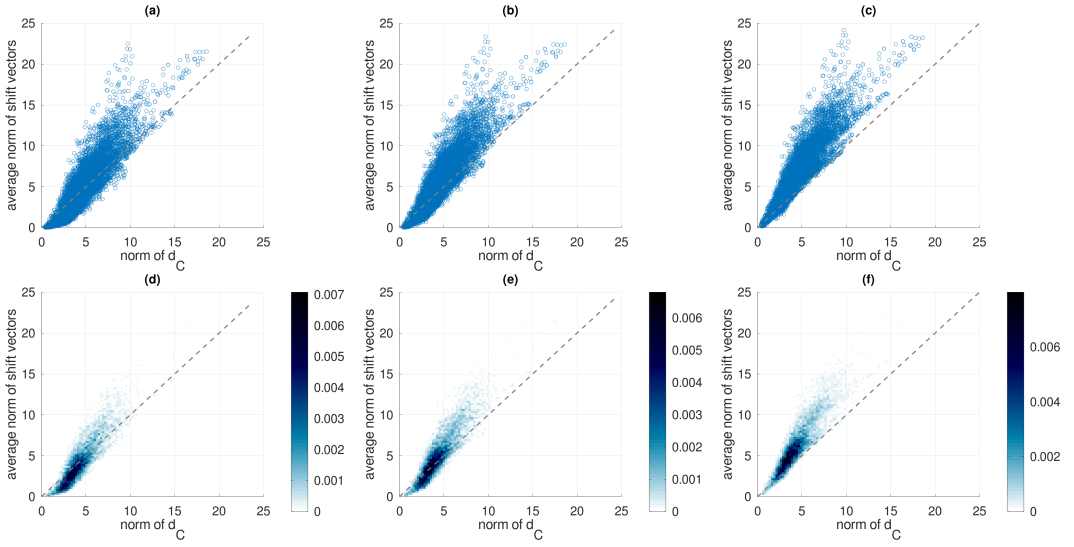

Figure 4.5: We show scatter and density plots of the average particle move versus the distance of the observation vector to the ensemble mean, all for the pressure level 500 hPa in ensemble space. On the x-axis we can see the norm of the observation distance to ensemble mean and on the y-axis we show the average size of the corresponding particle move. We display statistics for the LMCPF with different particle uncertainty, for each setting a scatter plot and a density plot which shows high density of points in a better way. (a) and (d) show the statistics for $\kappa = 1$, (b) and (e) for $\kappa = 2.5$ and (c) and (f) for $\kappa = 25$, all for May 6th, 0 UTC.

with background variance b and observation error variance r, reflecting the size of the particle move. When we, for example, choose $r = 4$ and $b = 16$, as we would get with typical values for the error of 2 $\frac{m}{s}$ for wind measurements and an ensemble standard deviation of 4 $\frac{m}{s}$, and then study $\kappa \in (1, 2.5, 10, 25)$, we obtain factors of size $s(\kappa) \in (0.8, 0.9, 0.97, 0.99)$. If the observation has a distance of 3.6 to the ensemble mean, as seen in Figure 4.2, this would make the means observed in Figure 4.4 plausible. For small $\kappa = 1$ here the particle move is 0.8 times the innovation, for large $\kappa = 25$ it is 0.99 times the innovation $y - H(x^{(b)})$. In Figure 4.4 we observe this behaviour with the median of the ensemble increments being `median` $= 2.62$ in (a) to `median` $= 3.54$ in (d).

As a final step of this part, we want to investigate not only the overall distribution of the particle moves, but relate the size of the average particle move to the distance of the observation to the ensemble mean. Figure 4.5 shows scatter and density plots for the LMCPF with different particle uncertainty. We employ

the same values for κ as in Figure 4.4, (a) and (d) with $\kappa = 1$, (b) and (e) with $\kappa = 2.5$, (c) and (f) show results for $\kappa = 25$. Displayed are statistics for the average particle move vs. the difference of the observation vectors from the ensemble mean, all for the pressure level at 500 hPa.

The results of Figure 4.5 show that clearly the move of the particles is related to the necessary correction as given by the distance of the observation to the individual particle. There is a clear correlation of the average move to the observation discrepancy with respect to the ensemble mean. If we would investigate each particle individually in one dimension, all points would be on a straight line with slope given by (4.3.6). The situation in a high-dimensional space with non-homogeneous metric is more complicated as reflected by Figure 4.5. The figure confirms that the method is working as designed.

4.3.3 Assimilation Cycle Quality Assessment of the LMCPF

Here, studying standard global atmospheric scores for the analysis cycle we investigate the quality of the LMCPF by testing different choices of $\kappa > 0$, investigate the interaction effects between particle uncertainty, ensemble spread and adaptive spread control and compare it to the way the LETKF moves the mean of the ensemble. For this aims we show two figures.

Figure 4.6 shows the functionality of the LMCPF by a display of the analysis and the first guess errors for upper air temperature for an ICON assimilation step, comparing the LETKF and the LMCPF with exact and approximate weights, respectively. Here, in the first line we show statistics for the LMCPF (blue line) with exact weights and $\kappa = 2.5$ compared to the LETKF (red line). The left panel shows the number of observations which passed quality control, the middle panel shows the root mean square error (RMSE) of observation minus first guess statistics $(o - f)$ (also known as observation - background $(o - b)$ statistics) and the right panel shows the RMSE for observations minus analysis statistics $(o - a)$.

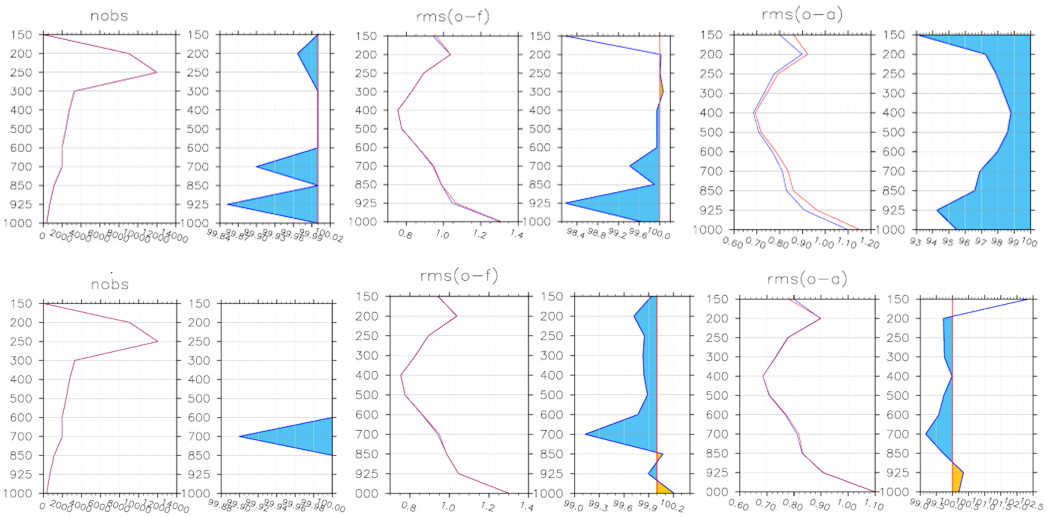

Figure 4.6: We show the observation verification of upper air temperature measured by aircrafts, in particular the first guess and analysis scores. The three columns show the number of observations which passed quality control, the RMSE for $o - f$ statistics and the RMSE for $o - a$ statistics for the LMCPF with exact weights (blue line) compared to the LETKF (red line) in the first row and the LMCPF (blue line) with exact weights compared to the LMCPF with approximate weights (red line) in the second row. We display results for one global assimilation step at 01.01.2022 00 UTC.

The blueish shading shows areas with lower values for the LMCPF in comparison to the LETKF. The second row shows the comparison of the LMCPF with exact (blue lines) and approximate (red lines) weights.

It can clearly be seen that with respect to $o - f$ scores the LMCPF is able to outperform the LETKF in case studies with one assimilation step when an appropriate size of the uncertainty of each particle, here given by the size of κ, is found. The experiments demonstrate that the exact weights yield better results than the approximate weights.

The numerical experiments prove that the particle uncertainty enables the LMCPF to move the background ensemble towards the observation in a way comparable to or even better than the LETKF. This effect remains active during model propagation and can also be observed for the first guess statistics and for forecasts with short lead times. Here, the LMCPF is able to outperform the operational version of the LETKF.

In Figure 4.7 we show a comparison of analysis cycle verification for a full one

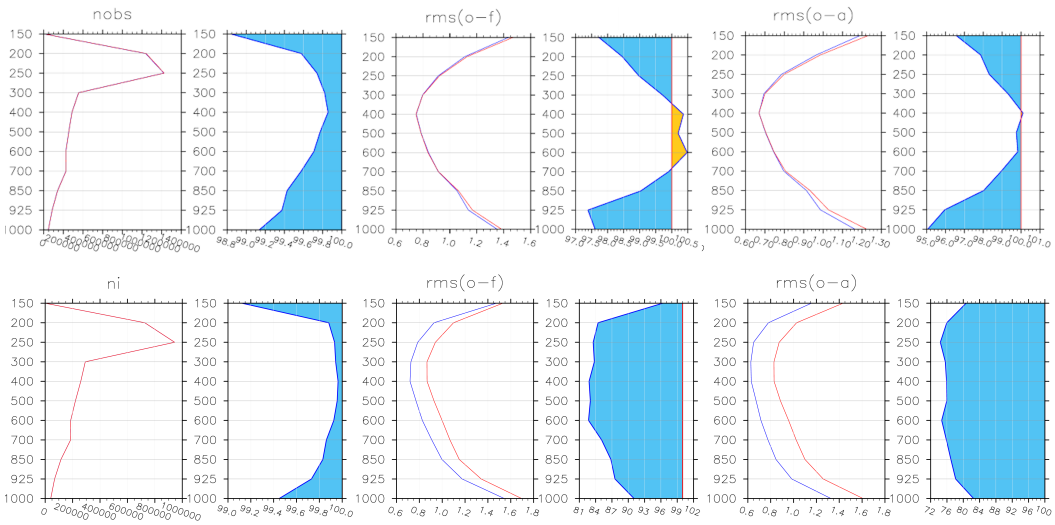

Figure 4.7: Again, we show some observation verification statistics for upper air temperature measured by aircrafts. We show the same statistics as in Figure 4.6 but for experiments carried out for the period one month each. In the upper row the comparison between LETKF (red line) and LMCPF (blue line) with exact weights is shown for Jan 2022, in the lower row we show the comparison between LAPF (red line) and LMCPF (blue line) in May 2016.

month period of LMCPF, LAPF and LETKF experiments. The columns are showing the same statistics as in Figure 4.6. The first row in Figure 4.7 shows the differences between LETKF (red line) and LMCPF with exact weights and $\kappa = 2.5$ (blue line) for a full month of cycling (Jan 2022). The second row shows the comparison of LAPF (red line) and LMCPF with approximate weights and $\kappa = 2.5$ (blue line) for one month (May 2016). Again, the blueish shading indicates lower numbers or RMSE values for the experiment (LMCPF), the yellowish shading indicates lower values for the reference (LETKF resp. LAPF).

Row one shows that the LMCPF with particle uncertainty given by $\kappa = 2.5$ can outperform the LETKF for short lead times, which is very important for practical applications. Here the LMCPF is up to 2.5% better than the LETKF for the $o - f$ statistics. In this experiment, for some levels in the atmosphere the $o - a$ and $o - f$ statistics of the LMCPF are up to 0.5% worse than the LETKF. The amount of data which passes quality control is quite similar for all methods under consideration, however, at some levels we loose up to 1.1% of observations in comparison with the LETKF. This is an effect of quality control based on the

ensemble spread - a smaller ensemble spread as we observe for the particle filter leads to less observations passing quality control. In the second row of Figure 4.7 we show the statistics of LAPF vs. LMCPF. Here we can clearly see that the LMCPF shows much better upper air scores than the LAPF. It clearly shows the importance of allowing a movement of particles towards the observations by using particle uncertainty.

Overall we conclude that with respect to the verification of the analysis cycle the LMCPF with particle uncertainty given by $\kappa = 2.5$ is comparable to the LETKF, with some levels to be better, some to be worse, overall differences mostly below 3%. The upper air verification for the analysis cycle of the LMCPF in operational setup is more than 10% better than for the LAPF.

4.3.4 The Evolution of the Ensemble Spread

It is an important evaluation step to investigate the stability of the LMCPF for global NWP over longer periods of time. We compare the particle spread evolution of the LMCPF, the LAPF and the LETKF in Figure 4.8. All experiments were started with an ensemble which consists of 40 identical copies of the particles, i.e. with an ensemble in degenerate state. Thus, here the tests also evaluate the capability of the whole system to resolve degeneracy and return to an ensemble with reasonable stable spread.

In a sequence of experiments we have tested the ability of the LMCPF to reach and maintain a particular ensemble spread using a combination of the choice of κ with a posterior covariance inflation

$$\tilde{\mathbf{G}}_{ens}^{(a)} = \kappa_{post}\mathbf{G}_{ens}^{(a)} \tag{4.3.7}$$

for each particle with $\tilde{\mathbf{G}}_{ens}^{(a)}$ replacing $\mathbf{G}_{ens}^{(a)}$ in equation (4.1.47), which is used to generate the analysis ensemble by random draws. We also note that for the random draw of equation (4.1.45) we employed bounds given by c_0 and c_1. The

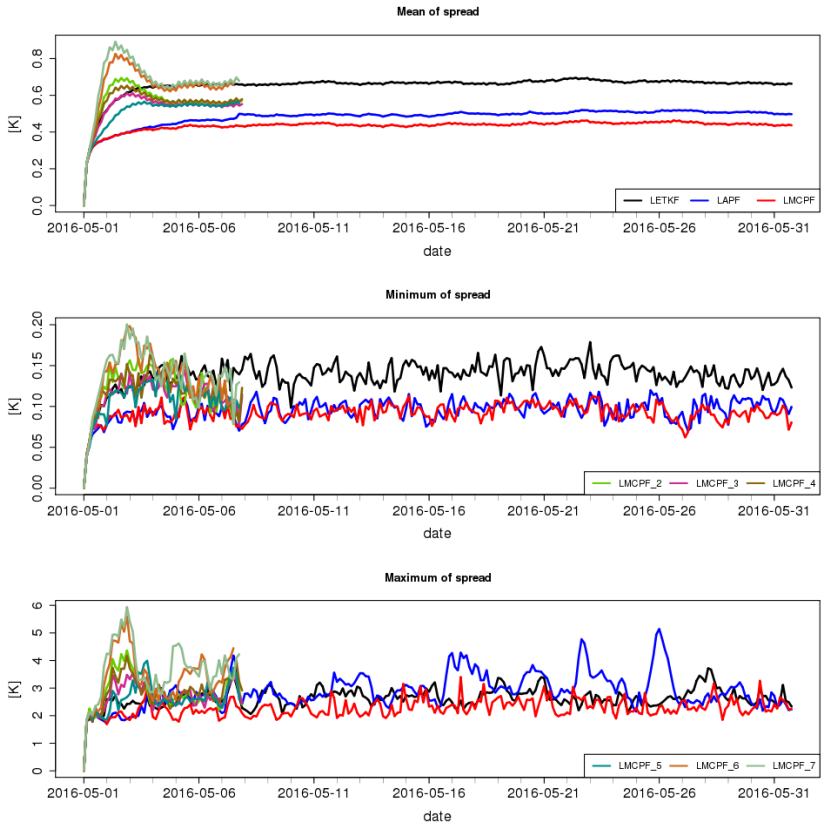

Figure 4.8: The evolution of the ensemble spread is shown for three filters and six different parameter choices for the LMCPF for a time period of both one month (LETKF - black, LAPF - blue, LMCPF - red) and for one week for different parameter choices for the LMCPF (see Table 4.1). The x-axis shows the period in one day steps. The y-axis shows the upper air temperature at ICON model level 64 ($\approx$ 500 hPa) in Kelvin. The first row shows the mean of the spread, the second row the minimum and the third row the maximum.

Exp No.	κ	κ_{post}	c_1	$\rho^{(1)}$
2	0.5	5	0.5	1.5
3	0.5	3	0.5	1.5
4	0.3	5	0.5	1.5
5	1	1	0.3	3.0
6	0.5	3	0.5	3.0
7	0.3	5	0.5	3.0

Table 4.1: Parameter choices for the six one week experiments of Figure 4.8. Further, we used $c_0 = 0.02$ and $\rho^{(0)} = 1.0$ for all experiments.

parameter combinations chosen for six different experiments over one week are compiled into Table 4.1. The corresponding spread evolution is visualised in Figure 4.8. The results show that, starting with an initial ensemble of identical particles, after some spin-up phase of 2-3 days all particle filters reach their particular spread level and keep it stable over a longer period of time. We carried out selected longer term studies comparing the behaviour of the LMCPF (red), the LAPF (blue) and the LETKF (black) over a period of one month.

The control of the ensemble spread is a delicate topic. A larger ensemble spread does not necessarily lead to better forecast scores, measured by RMSE (Skill) of the ensemble mean or its standard deviation (SD), defined as the RMSE after the bias has been subtracted. With the ability to control separately the strength of the adaptive resampling and the ability of the filter to pull the particles towards the observations, we have independent parameters at hand to adapt the approximations to a real-world situation. At the same time, the way the assimilation step of the LMCPF pulls the ensemble to the observations is based on both the size of the particle uncertainty, which itself is depending on the ensemble spread, and within the cycled environment on the adaptive resampling. Of course, it would be desirable to develop tools to estimate the real uncertainty adequate for each particle, and to keep all parts of the system consistent. We expect this to lead to much further research and discussions, which are beyond the scope of this work.

4.3.5 Forecast Quality of the LETKF and LMCPF experiments

As the last part of the numerical results, we study the quality of longer forecasts based on the analysis cycle of the LMCPF with $\kappa = 2.5$ and compare it to the LETKF based forecasts in Figure 4.9 and to forecasts based on the LAPF analysis cycle in Figure 4.10. For this purpose, forecasts were run twice a day at 00 UTC and 12 UTC. In Figure 4.9 we display upper air verification for the LMCPF (dashed lines) with exact weights and for the LETKF (solid lines). The different colors identify the different lead times, from one day up to one week. The first row shows the upper air temperature and the second row shows the verification of pressure forecasts. The first panel shows the Continuous Ranked Probability Score (CRPS), the second panel the Standard Deviation (SD), the third panel the Root Mean Square Error (RMSE) and the last panel shows the Mean (ME). For CRPS, SD and RMSE it is the aim to receive statistics as low as possible; for the Mean (=Bias) it is the goal to reach zero. We used the same observations for verification in both experiments.

Studying the results shown in Figure 4.9, we observe that forecast scores are very similar for LMCPF and LETKF for the upper air temperature. For pressure forecast the bias (ME) for the LMCPF is mostly smaller than for the LETKF below $50hPa$.

In Figure 4.10 we show the same statistics as in Figure 4.9 focussing on relative humidity and upper air temperature for the comparison of LMCPF and LAPF, where here we used the approximate weights or both to study the effect of the shifts only. Here, it can be clearly seen that the LMCPF shows lower RMS errors than the LAPF for both variables and for all levels. For relative humidity the LMCPF is clearly better for the shorter lead times up to three days, but with less prominence it still outperforms the LAPF for the longer lead times up to one week. For the upper air temperature the RMSE statistics are clearly better for the LMCPF for all lead times. It is worth noting that the biases for the two

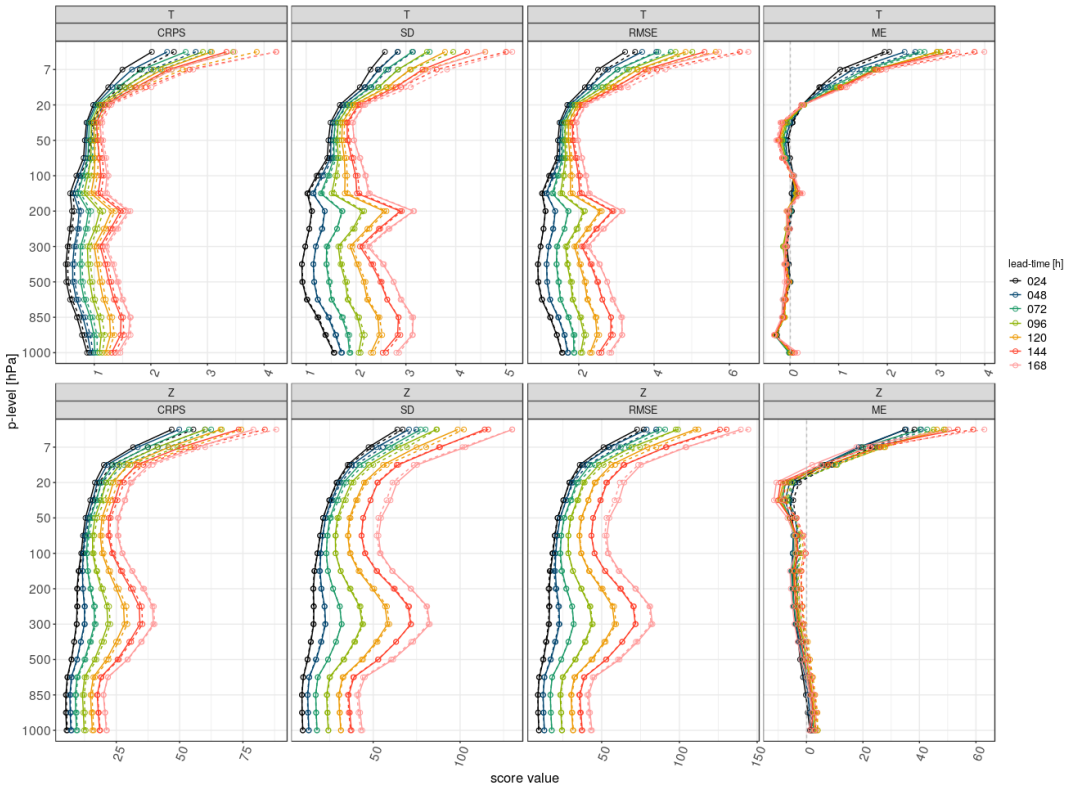

Figure 4.9: We display forecast scores for the LMCPF (dashed) with exact weights and the LETKF (bold lines) calculated for January 2022. Shown are the continuous rank probability score (CRPS), the standard deviation (SD), the RMSE and the mean (ME). First row shows the upper air temperature, the second row shows pressure forecast verification. The colors indicate the different lead times from one day to 7 days.

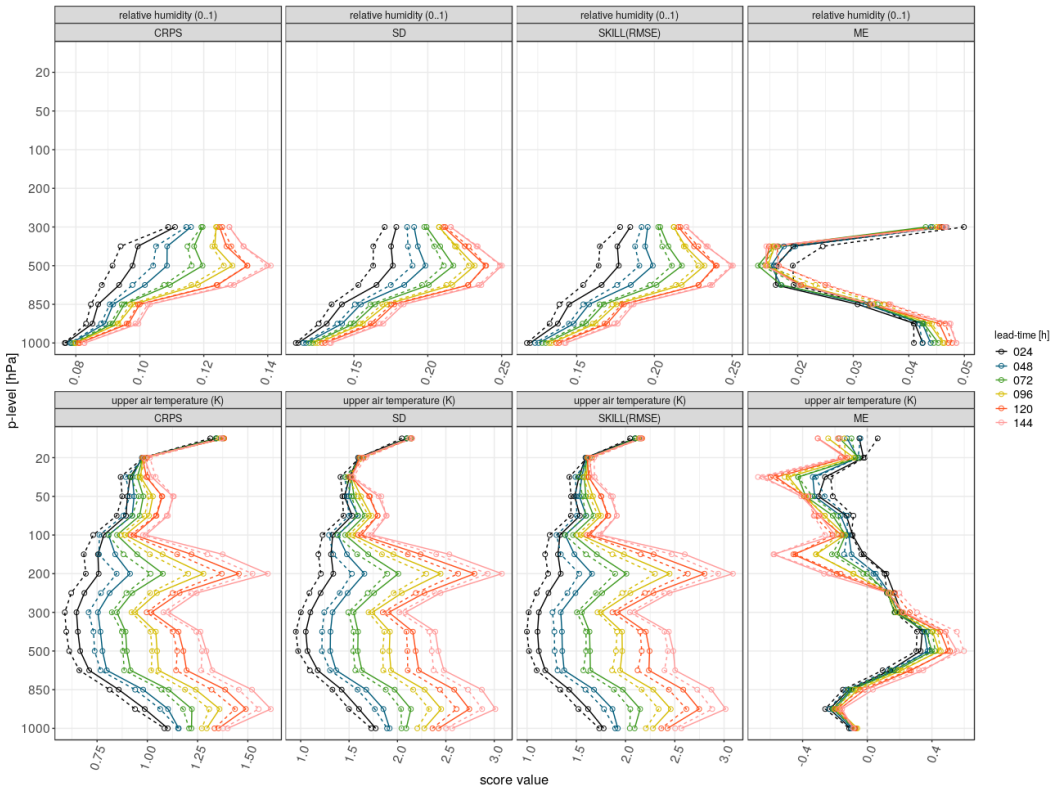

Figure 4.10: Exemplary for relative humidity and upper air temperature we show the improvement of the LMCPF with approximate weights (dashed) compared to the the LAPF (bold lines) for May 2016.

particle filters show a quite similar behaviour.

These results altogether demonstrate that using particle uncertainty is an important ingredient for improving first guess and forecast scores of the particle filter.

4.4 Conclusions

In this work we develop the use of a Gaussian mixture within the framework of the Localized Adaptive Particle Filter (LAPF) introduced in chapter 3, as an approximation to model and forecast particle uncertainty in the prior and posterior distributions. The filter, following earlier ideas of Hoteit et al. (2008) and Liu et al. (2016a,b) constructs an analysis distribution based on localised Gaussian mixtures, whose posterior coefficients, covariances and means are calculated based on the prior mixture given by the ensemble first guess and the observa-

tions. The analysis step is completed by resampling and rejuvenation based on the LAPF techniques, leading to a Localised Mixture Coefficients Particle Filter (LMCPF). In contrast to the LAPF the LMCPF is characterised by a move or shift of the first guess ensemble towards the observations, which is consistent with the non-Gaussian posterior distribution based on a Bayesian analysis step, and where the size of the move is controlled by the size of the uncertainty of individual particles.

We have implemented the LMCPF in the framework of the global ICON model for numerical weather prediction, operational at Deutscher Wetterdienst. Our reference system to test the feasibility of ideas and demonstrate the quality of the LMCPF is the LETKF implementation operational at DWD, which generates initial conditions for the global ICON Ensemble Prediction System ICON-EPS. We have shown that the LMCPF runs stably for a month of global assimilations in operational setup and for a wide range of specific LMCPF parameters. Our investigation includes a study of the distribution of observations with respect to the ensemble mean and statistics of the distance of ensemble members to the projection of the observations into ensemble space. We also study the average size of particle moves when uncertainty is employed for individual Gaussian particles within the LMCPF and provide an analytic explanation of the histogram shapes with a comparison to the eigenvalue distribution of the matrices $\mathbf{A}$ on which the particle weights are based.

We show that the upper air first guess errors of the LMCPF and LETKF during the assimilation cycle are very similar within a range of plus-minus 1-3%, with the LMCPF being better below 850 hPa and the LETKF being better in some range above. Forecast scores for a time-period of one month have been calculated, demonstrating as well that the RMSE of the ensembles is comparable for upper air temperature, relative humidity, wind fields and pressure (2-3%). The size of the mean spread of the LMCPF strongly depends on parameter choices and is usually stable after a spin-up period.

In several shorter case studies we demonstrate that by varying the parameter choices, we can achieve better first guess RMSE for the LMCPF in comparison to the LETKF, which shows that for very short range forecasts the quality of the method can be comparable to or better than that of the LETKF. While reaching a break-even point for operational scores with a new method establishes an important mile-stone, we need to note that there are many open and intricate scientific questions here with respect to the choice of parameters for the Gaussian mixture and their inter-dependence as well as about the control of an optimal and correct ensemble spread both in the analysis cycle and for the forecasts.

Chapter 5

Interpretation of the LMCPF

This chapter will show that the LMCPF of chapter 4 in the case of Gaussian mixtures can be interpreted as Implicit Equal-Weights Particle Filter (Zhu et al., 2016).

5.1 Interpretation of the LMCPF - The Localised Implicit Equal-Weights Particle Filter

The Implicit Equal-Weights Particle Filter (IEWPF) of Zhu et al. (2016), which we were working with since the beginning of the particle filter development, is another promising particle filter version. During the work on this thesis we investigated the following: The Localised Mixture Coefficients Particle Filter (LMCPF) can be interpreted as a special case of the IEWPF where the proposal distribution is used to draw from an analytical posterior distribution. In particular, we replace $q(\xi_j)$ in equation (5.1.4) with $p(y|\alpha_j\xi_j)p(\alpha\xi_j)$. It then reduces to $1/p(y)$. In this case by construction we obtain equal weights with $\alpha = 1$.
During the following sections we will now present the details, starting with the basic idea of the IEWPF, investigating a Newton method and showing results for a one dimensional example to transparently demonstrate the IEWPF and its special case.

5.1.1 The basic idea of the Implicit Equal-Weights Particle Filter (IEWPF)

The Implicit Equal-Weights Particle Filter (IEWPF) of Zhu et al. (2016) employs the use of a proposal distribution to draw from the posterior. The basic idea is to make the proposal distribution dependent on a parameter α and then choose α such that the posterior weights of the particles become equivalent.

In more detail, the Bayesian step to calculate the posterior distribution is given by

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{5.1.1}$$

with the prior distribution $p(x)$, the probability $p(y)$ of the observation which serves as a norming constant and the conditional likelihood $p(y|x)$ of the observation given the model state x. When we draw from a proposal distribution $q(x)$, the equation (5.1.1) becomes

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)q(x)} \cdot q(x). \tag{5.1.2}$$

When we scale $q(x)$ by a constant α, we obtain

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)q(\frac{x}{\alpha})} \cdot q(\frac{x}{\alpha}). \tag{5.1.3}$$

In the case where q is a Gaussian distribution and the random draw is given by ξ, the corresponding draw from $q(\frac{x}{\alpha})$ is given by $\alpha\xi$ where ξ is drawn from $q(x)$. Its probability with respect to $q(\frac{x}{\alpha})$ is given by $q(\xi)$.

Let us assume that we draw from $q(\frac{x}{\alpha_j})$ for $j = 1, ..., L$, with L is the number of particles, i.e. we obtain the posterior weight w_j of $\alpha_j\xi_j$ to be

$$w_j = \frac{p(y|\alpha_j\xi_j)p(\alpha_j\xi_j)}{p(y)q(\xi_j)}, \quad j = 1, ..., L. \tag{5.1.4}$$

The Implicit Equal-Weights Particle Filter now solves the equation system (5.1.4)

with respect to $\alpha = (\alpha_1, ..., \alpha_L)$ such that $w_1 = ... = w_L$.

5.1.2 A Newton Method for the IEWPF

We employ a Newton method for solving (5.1.4). To this end, let us write the equation system in the form

$$F_j(\alpha_j, w) := \frac{p(y|\alpha_j\xi_j)p(\alpha_j\xi_j)}{q(\xi_j)} - w, \quad j = 1, ..., L, \tag{5.1.5}$$

where we removed the unknown normalization factor $p(y)$. If the probability distributions $p(x)$ and $p(y|x)$ are continous in x and nonzero, for sufficiently small w the equation 5.1.5 for $j = 1, ..., L$ do have a solution $F_j(\alpha_j, w) = 0$, which follows from the intermediate value theorem applied to the interval between minimal and maximal values of the function $F_j(\alpha) := \frac{p(y|\alpha\xi_j)p(\alpha\xi_j)}{p(y)q(\xi_j)}$.
Newton's method searches for zeros of a function, i.e. now we solve

$$F_j(\alpha_j, w) \stackrel{!}{=} 0, \quad j = 1, ..., L \tag{5.1.6}$$

by linearisation and iteration. We linearise at an iteration point $\alpha_j^{(i)}$ with iteration index i to obtain

$$\begin{aligned} F_j(\alpha_j, w^{(i)}) &\approx F_j(\alpha_j^{(i)}, w^{(i)}) + \frac{dF_j(\alpha_j, w^{(i)})}{d\alpha_j}\Big(\alpha_j - \alpha_j^{(i)}\Big) && (5.1.7) \\ &\stackrel{!}{=} 0 && (5.1.8) \end{aligned}$$

Solving the Newton equation leads to the update

$$\alpha_j^{(i+1)} = \alpha_j^{(i)} - \Big(\frac{dF_j(\alpha_j, w^{(i)})}{d\alpha_j}\Big)^{-1} F_j(\alpha_j^{(i)}, w^{(i)}) \tag{5.1.9}$$

for all particles with index $j = 1, ..., L$.

Solving (5.1.4), the weight w will finally be identical to $w = \frac{1}{L}$. However, since $p(y)$ is usually not known, we cannot calculate the exact function F. Usually, F

is known only up to a constant, which is calculated by normalisation using the weights of all particles, i.e.

$$p(y) = \sum_{j=1}^{L} w_j. \tag{5.1.10}$$

with w_j given in equation (5.1.4).

This leads to

$$w = \frac{1}{L} \sum_{j=1}^{L} w_j, \tag{5.1.11}$$

which we employ in our Newton method.

The calculation of the derivative can be done analytically when the functions $p(y|x)$ and the prior $p(x)$ is given analytically, for example in the case where we study a Gaussian mixture approach. In general, we can employ numerical differentiation by e.g. finite differences.

5.1.3 A one-dimensional example of the IEWPF analysis step

For study and working towards a full-scale implementation of a localised IEWPF in the large-scale framework of the ICON model, here we study a one-dimensional example in detail. In Figure 5.1 we show the prior distribution $p(x)$ in blue and the likelihood $p(y|x)$ in green for our 1d example. The blue dashed curve shows the probability density function, the blue dots below the curves are the particles drawn from this function and the blue solid curve shows the histogram of this particles plotted as a curve. The vertical blue solid line indicates the mean of the distribution. The green solid curve shows the data distribution $p(y|x)$ with its mean indicated by the vertical solid green line. The vertical green and blue lines are also shown in Figure 5.2 where the analysis $x^{(a)}$, i.e. the mean of the posterior, is shown as vertical orange line. The dashed orange curve shows the posterior analytically calculated in the case of Gaussian prior and Gaussian likelihood.

The IEWPF starts with draws from a Gaussian distribution with variance $\sigma = 1$ centered at $x^{(a)}$. Here, we carry this out and display the result as magenta

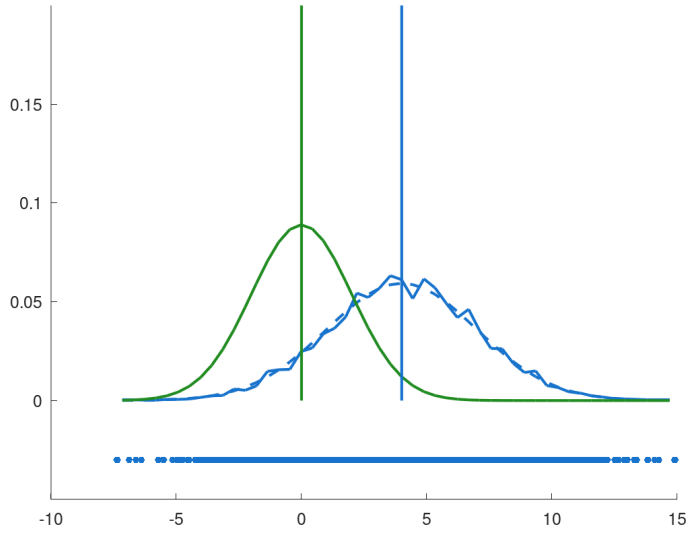

Figure 5.1: Shown is the prior distribution ($p(x)$) in blue and the likelihood ($p(y|x)$) in green. The probability density function is shown as blue dashed curve, the blue dots below indicate the particles drawn from this function and the blue solid curve shows the histogram of the particles. The vertical lines indicate the corresponding mean.

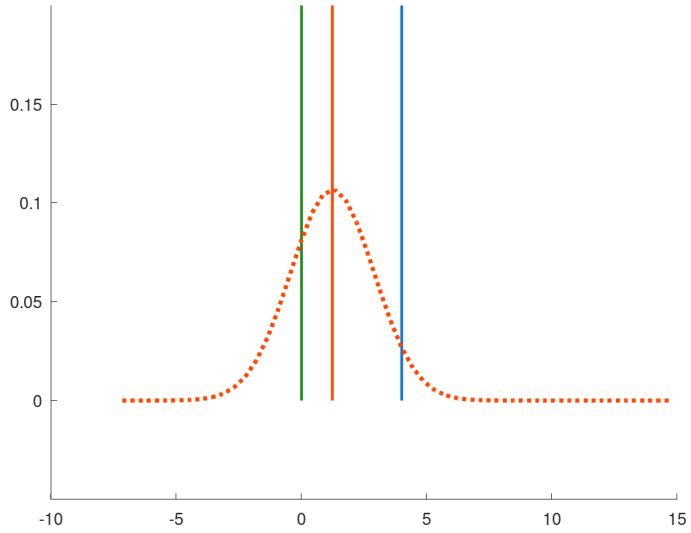

Figure 5.2: Shown are the mean of the prior (blue) and the data distribution (green) and the mean of the posterior distribution (vertical solid orange line). The dashed orange curve shows the posterior distribution $x^{(a)}$.

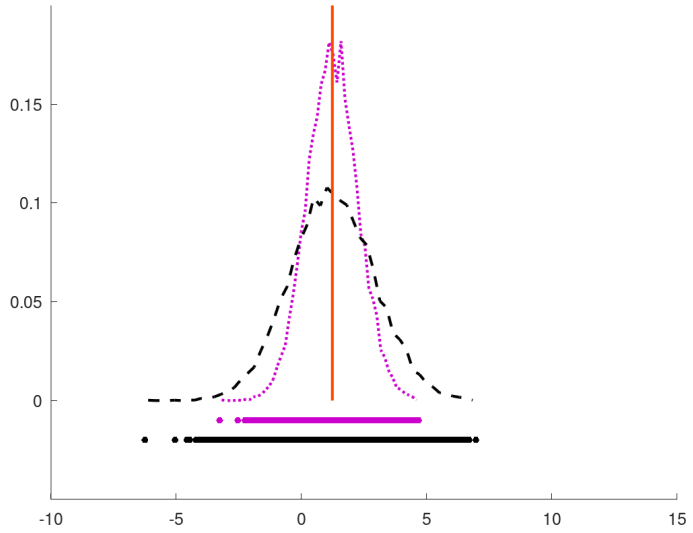

Figure 5.3: The magenta dots show the result of drawing with the IEWPF from a Gaussian with it's centre at the analysis $x^{(a)}$ (vertical solid orange line). The magenta curve shows the histogram of this magenta points. The dashed black curve shows the posterior distribution after drawing with the IEWPF from the proposal, with matching particles as black dots below.

points. Their histogram is shown as dotted magenta curve in Figure 5.3. Next, we use Newton's method to determine α and then use the IEWPF to draw from the proposal distribution with the determined $\alpha's$ for each particle and get the dashed black line - the posterior distribution determined by the IEWPF, the black dots below the curves are indicating the particles.

Finally, in Figure 5.4 we show all curves in one plot for a better comparison.

In Figure 5.5 we now modify the proposal distribution of the IEWPF and draw from the analytical version of the posterior calculated for Gaussian mixtures. In this case, we employ

$$q(x) := p(y|x)p(x), \quad x \in \mathbb{R}. \tag{5.1.12}$$

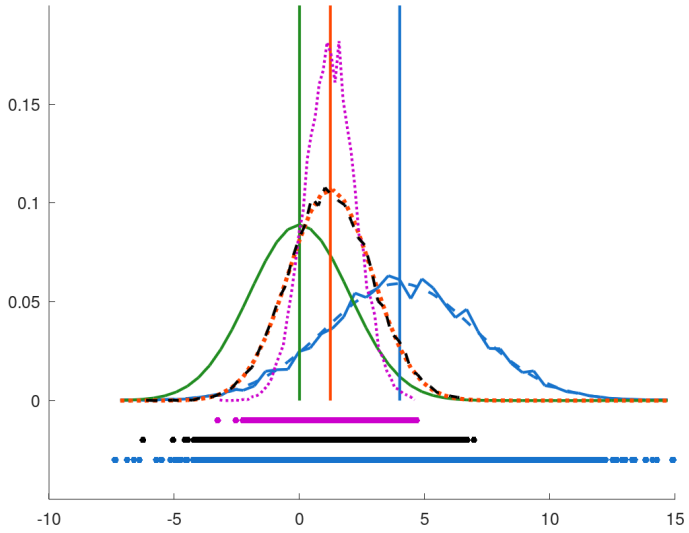

Figure 5.4: Shown are all statistics of Figures 5.1 to 5.3 in one plot for a better comparison.
Again, blue indicates the prior distribution, green the likelihood distribution, orange the analysis, magenta the proposal distribution and black indicates the posterior distribution.

This means that the weights we calculate for $\alpha_j \xi_n$ are given by

$$\begin{aligned} w_j &= \frac{p(y|\alpha_j\xi_j)p(\alpha_j\xi_j)}{p(y)q(\xi_j)} \\ &= \frac{p(y|\alpha_j\xi_j)p(\alpha_j\xi_j)}{p(y)p(y|\alpha_j\xi_j)p(\alpha_j\xi_j)} \\ &= \frac{1}{p(y)} \end{aligned} \tag{5.1.13}$$

for $j = 1, ..., L$. The Newton method will terminate in the first step with $\alpha_j = 1$. The IEWPF now coincides with the analytical calculation of the posterior and draws from its distribution. We show the results in this case, where again the colours are as in Figures 5.1 to 5.3. The dashed black curve visualises the modified posterior distribution which is now equal to the proposal distribution (magenta dashed curve).

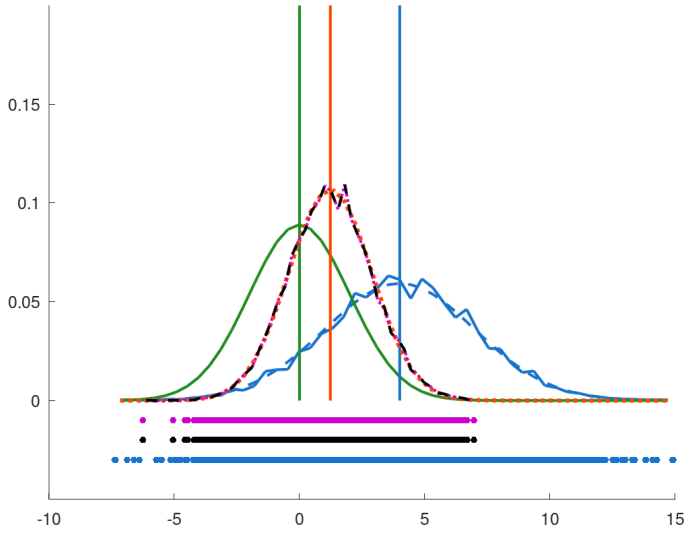

Figure 5.5: Shown are all statistics in one plot, but for the modified proposal distribution calculated for Gaussian mixtures. Again, the blue lines indicating the prior distribution, the green lines the likelihood, the orange lines the analysis. The dashed black curve visualizes the modified posterior distribution which is now equal to the proposal distribution (magenta dashed curve) calculated for Gaussian mixtures.

5.1.4 Conclusions

We have shown that with drawing from the right proposal the LMCPF could be interpret as a special version of the IEWPF and proved this in section 5.1.3 with the help of a one dimensional example.

The IEWPF of Zhu et al. (2016) draws from a proposal distribution determined by drawing from a Gaussian centered at the analysis ($x^{(a)}$), i.e. the mean of the posterior. Afterwards the $\alpha's$ are determined and then used to draw from the proposal. If we use a Gaussian Mixture to determine the proposal, as the LMCPF does, we receive $\alpha_j = 1$, the IEWPF draws from the proposal which is now equal to the analytical calculation of the posterior.

Chapter 6

Summary and Conclusions

This final chapter summarises the two particle filters and gives an outlook on how to proceed with them.

Standard algorithms for data assimilation used for large-scale atmospheric analysis in operational centres include the Ensemble Kalman Filter and 4d-Var. These data assimilation methods are either inherently or practically based on the assumption that the underlying distribution is Gaussian. If the ensemble distribution is not Gaussian, these methods are not optimal. In the case of non-Gaussianity more general Bayesian methods such as the particle filter have been proposed. The core idea of the particle filter is to realise the Bayesian approach giving a weight to each particle depending on its distance to the observations. The adaptation of particles is carried out in different ways, either by resampling, nudging particles towards some proposal distribution or by optimal transport processes.

Classical particle filters in high dimensional dynamical systems suffer from filter divergence or filter collapse due to the curse of dimensionality. In this thesis two particle filters have been developed and implemented into the operational global NWP model ICON of DWD - the Localised Adaptive Particle Filter (LAPF) and the Localised Mixture Coefficients Particle Filter (LMCPF). With the help of modulated rejuvenation, localisation in ensemble space and adaptivity we

prevent the filter divergence as well as filter collapse.

The LAPF and the LMCPF have been tested over a period of one month with 40 ensemble members, a global horizontal resolution of 52 km and 90 vertical layers in an operational setup with slightly reduced resolution. Both particle filters are able to run stably over the test period of one month.

As the LAPF has been the first particle filter which has been implemented in an operational large-scale and high dimensional NWP model it has some disadvantages. The LAPF just chooses the particles closest to the observations and than drawing from them, it can't move them towards the observations if the prior particles are far away. So a large gap in the location between prior particle distribution and observations can't be handled. Therefore, we have developed the second particle filter of this thesis, the LMCPF.

The LMCPF also runs stably over the experimental period of one month and is even able to match the quality of the reference, determined with the LETKF. Because of the shift vectors $\beta^{(shift,\ell)}$ given by (4.1.40) the LMCPF is able to move the selected particles towards the observations and then draw around them from a Gaussian distribution. This leads to a large improvement in comparison with the LAPF and to results comparable to the LETKF.

To complete this thesis, the following sections will summarise both filters, starting with the LAPF (section 6.1), followed by the LMCPF (section 6.2) and then gives a brief outlook on Future Work (section 6.3).

6.1 The LAPF

The first particle filter developed in this thesis is the Localised Adaptive Particle Filter (LAPF, chapter 3). It is based on the LETKF implementation of DWD and stays as parallel as possible to it. It consists of the following five steps:

1 The classical Particle Filter step (section 3.1.1, page 28)

The classical Particle Filter step uses an ensemble $x^{(l)}$ of states which represents the prior probability distribution $p_k^{(b)}$ at time t_k in the form of δ-distributions. To carry out the analysis step at time t_k, weights are calculated by

$$w_{k,\ell} := cp(y_k|x^{(l)}), \quad \ell = 1, ..., L,$$

where c is a normalisation constant.

2 Projection onto ensemble space (section 3.1.2, page 28)

As seen from (2.2.1), the LETKF is based on the projection of the observation onto ensemble space of observation equivalents. We are doing this by calculating $A := \mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}$ and $C := A^{-1}\mathbf{Y}^T\mathbf{R}^{-1}(\mathbf{y}^o - \overline{\mathbf{y}}^b)$. Localisation is performed as for LETKF and LAPF, i.e. the matrix R is weighted according to the distance of each of its observations to the analysis point.

3 Classical Resampling (section 3.1.3, page 31)

For resampling, accumulated weights w_{ac_ℓ}, $\ell = 1, ..., L$, are defined by

$$w_{ac_0} = 0, \;\; w_{ac_\ell} = w_{ac_{\ell-1}} + w_{k,\ell}, \qquad \ell = 1, ..., L,$$

where now we employ normalisation to the total weight of L. Then, we draw $r_\ell \sim U([0,1])$, $\ell = 1, ..., L$, set $R_\ell = \ell - 1 + r_\ell$ and define the transform matrix for the particles by

$$\breve{\mathbf{W}}_{i,\ell} = \begin{cases} 1, & \text{if } R_\ell \in \; (w_{ac_{\ell-1}}, w_{ac_\ell}], \\ 0, & \text{otherwise,} \end{cases}$$

$i, \ell = 1, .., L$ with $\breve{\mathbf{W}} \in \mathbb{R}^{L\times L}$, where $(s,t]$ denotes the interval of values $s < \eta \leq t$.

4 Spread Control (section 3.1.4, page 31)

In ensemble data assimilation systems the spread of the ensemble evolves as a result of model dynamics, model errors and active observations. In the operational LETKF an adaptive inflation factor ρ is estimated. For the LAPF we are calculating

$$\rho = \frac{\mathbb{E}[\mathbf{d}_{o-b}^T \mathbf{d}_{o-b}] - \mathrm{Tr}(\mathbf{R})}{\mathrm{Tr}\big(\mathbf{H}\mathbf{P}^b\mathbf{H}^T\big)}.$$

at each localisation point.

5 Gaussian Resampling or Rejuvenation (section 3.1.5, page 34)

For the adaptive resampling step the size of the draw (given by $\mathbf{N}$) is modulated by applying a scalar perturbation factor σ for each analysis grid point. Scaling of the draw around each member at time t_k is carried out by

$$\mathbf{W}(p) = \breve{\mathbf{W}}(p) + \mathbf{N} \cdot \sigma(\rho_k(p)).$$

The specification of the factor σ is calculated by:

$$\sigma(\rho) := \begin{cases} c_0, & \rho < \rho^{(0)}, \\ c_0 + (c_1 - c_0) \cdot \frac{\rho - \rho^{(0)}}{\rho^{(1)} - \rho^{(0)}}, & \rho^{(0)} \leq \rho \leq \rho^{(1)} \\ c_1, & \rho > \rho^{(1)}, \end{cases}$$

where the constants $\rho^{(0)}, \rho^{(1)}, c_0, c_1$ are tuning constants.

Finally, the full analysis ensemble is given by

$$\mathbf{X}^{(a,full)} = \overline{x}^{(b)} + \underbrace{\mathbf{X}^{(b)}\breve{\mathbf{W}}}_{class.\ resampling} + \underbrace{\mathbf{X}^{(b)}\mathbf{N}\sigma}_{rejuvenation}$$

These five ingredients ensure that the LAPF doesn't suffer from filter divergence

or filter collapse due to the curse of dimensionality

We have implemented the LAPF in the framework of the operational global ICON model for numerical weather prediction of DWD. The reference system to test the feasibility of ideas and demonstrate the quality of the LAPF is the operational LETKF implementation at DWD, which generates initial conditions for the global ICON Ensemble Prediction System ICON-EPS.

With this system we run a global experiment with a resolution of 52 km for the 40 ensemble runs and 26 km for the deterministic run and with 90 vertical layers. The period was taken to be a whole month, May 1-31, 2016.

We are able to show that the LAPF runs stably over the whole period of one month and gives results comparable to the LETKF. It is therefore the first particle filter implemented in an operational global NWP model which runs stably and gives results close to the quality of the reference LETKF as we can see from the results shown in section 3.2. We have shown that for the RMSE the quality of forecasts is 10-15% behind the forecast quality of the LETKF for forecasts up to several days (compare Figures 3.6 and 3.7) and the bias is partly improved, in particular for humidity. Altogether, for the assimilation cycle and forecasts the LAPF shows promising results.

6.2 The LMCPF

The second particle filter developed in this thesis is the Localised Mixture Coefficients Particle Filter (LMCPF, chapter 4). The LMCPF uses an analysis distribution based on localised Gaussian mixtures, whose posterior coefficients, covariances and means are calculated based on the prior mixture given by the ensemble first guess and the observations. The analysis step is completed by resampling and rejuvenation based on the LAPF techniques.

In contrast to the LAPF, which just makes a choice on which particles will survive, the LMCPF is able to move or shift the first guess ensemble towards the observations, which is consistent with the non-Gaussian posterior distribution

based on a Bayesian analysis step. This move is controlled by the size of the uncertainty of the individual particles.

The LMCPF consists of the following six steps (see also section 4.1.3),

1 The classical Particle Filter step

We start with some initial ensemble $x_0^{(a,\ell)}$ at time t_0. We carry out a propagation step, i.e. we run the model forward from time t_{k-1} to t_k for each ensemble member, leading to the background ensemble $x_k^{(b,\ell)}$ at time t_k.

2 Projection onto ensemble space

Then, at each localisation point ξ on a coarser analysis grid $\mathcal{G}$ we carry out the localised ensemble transform (4.1.36), calculating C and $\mathbf{A}$. Localisation is carried out as for the LETKF and LAPF, i.e. the matrix $\mathbf{R}$ is weighted depending of the distance of each of its observations to the analysis point.

3 Classical Resampling

We now carry out a classical resampling step following Section 3.1.3. This leads to a matrix

$$\breve{\mathbf{W}}_{i,\ell} = \begin{cases} 1, & \text{if } R_\ell \in \ (w_{ac_{i-1}}, w_{ac_i}], \\ 0, & \text{otherwise,} \end{cases}$$

$i, \ell = 1, ..., L$, draw $r_\ell \sim U([0,1])$, set $R_\ell = \ell - 1 + r_\ell$, with accumulated weights w_{ac}, $w_{ac_0} = 0, w_{ac_i} = w_{ac_{i-1}} + w_{k,i}$, $w_{k,i} := p(\mathbf{y}_k | x^{(b,i)})$ and $\breve{\mathbf{W}} \in \mathbb{R}^{L \times L}$ with entries one or zero reflecting the choice of particles.

4 Posterior matrix and shift vectors

The posterior matrix $\mathbf{G}_{ens}^{(a)}$ given by (4.1.38) and the shift vectors $\beta^{(shift,\ell)}$ given by (4.1.40) for $\ell = 1, ..., L$ are calculated for each localisation point. We define

$$\mathbf{W}^{(shift)} := \left(\beta^{(shift,1)}, ..., \beta^{(shift,L)} \right) \ \in \mathbb{R}^{L \times L}.$$

We obtain the shifts for these particles by the product $\mathbf{W}^{(shift)}\breve{\mathbf{W}}$. According to the analysis equation (4.1.37) the new coordinates in ensemble space are calculated by

$$\left(\beta^{(a,1)}, ..., \beta^{(a,L)}\right) = \breve{\mathbf{W}} + \mathbf{W}^{(shift)}\breve{\mathbf{W}}.$$

5 Spread Control

For each particle we now carry out an adaptive Gaussian resampling or rejuvenation step. The rejuvenation is carried out the same way as described in Sections 3.1.4 and 3.1.5, i.e. we first calculate

$$\rho = \frac{\mathbf{d}_{o-b}^T \mathbf{d}_{o-b} - Tr(\mathbf{R})}{Tr(\mathbf{H}\frac{1}{L-1}\mathbf{X}\mathbf{X}^T\mathbf{H}^T)}$$

at each localisation point, with the actual ensemble covariance matrix $\frac{1}{L-1}\mathbf{X}\mathbf{X}^T$ and with the observation minus background statistics $\mathbf{d}_{o-b} = y_k - \bar{y}_k$ where $\bar{y}_k$ denotes the ensemble mean in observation space described in (4.1.4) at time t_k.

6 Gaussian Resampling or Rejuvenation

We scale ρ by some function

$$\sigma(\rho) := \begin{cases} c_0, & \rho < \rho^{(0)}, \\ c_0 + (c_1 - c_0)\frac{\rho - \rho^{(0)}}{\rho^{(1)} - \rho^{(0)}}, & \rho^{(0)} \leq \rho \leq \rho^{(1)}, \\ c_1, & \rho > \rho^{(1)}, \end{cases}$$

where the constants $\rho^{(0)}, \rho^{(1)}, c_0, c_1$ are tuning constants. Then, the rejuvenation plus shift step is carried out by

$$\mathbf{W} := \breve{\mathbf{W}} + \mathbf{W}^{(shift)}\breve{\mathbf{W}} + [\mathbf{G}_{ens}^{(a)}]^{\frac{1}{2}}\mathbf{N}\sigma.$$

Finally we calculate the analysis ensemble by

$$\begin{aligned}\mathbf{X}^{(a,full)} &= \overline{x}^{(b)} + \mathbf{X}^{(b)}\mathbf{W} \\ &= \overline{x}^{(b)} + \underbrace{\mathbf{X}^{(b)}\breve{\mathbf{W}}}_{class.\ resampling} + \underbrace{\mathbf{X}^{(b)}\mathbf{W}^{(shift)}\breve{\mathbf{W}}}_{shift} + \underbrace{\mathbf{X}^{(b)}[\mathbf{G}_{ens}^{(a)}]^{\frac{1}{2}}\mathbf{N}\sigma}_{adapt.\ Gauss.\ resampling}\end{aligned}$$

The LETKF does not know the selection reflected by the matrix $\breve{\mathbf{W}}$, instead it transforms the ensemble by its matrix $\mathbf{W}$. The LETKF and the LMCPF know a shift term, for the LETKF it is given by w, for the LMCPF by $\mathbf{W}^{(shift)}\breve{\mathbf{W}}$, shifting each particle according to model error, where the LETKF shifts according to the full ensemble spread. The LMCPF also takes into account that part of the ensemble spread which is kept during the selection process. Further, it employs adaptive resampling around each remaining shifted particle. This helps to keep the filter stable and achieve an appropriate uncertainty described by $o-b$ statistics.

We have implemented the LMCPF in the operational NWP model ICON of DWD. The experiments carried out with the same setup as the experiments with the LAPF - we run a global experiment with 52 km resolution, 90 vertical layer and 40 ensemble members. Again, the experimental periods have been one whole month, the first period May 2016, the second experimental period has been January 2022.

In section 4.3 we have shown that the LMCPF runs stably for both whole experimental periods of one month. Our investigation includes a study of the distribution of observations with respect to the ensemble mean and statistics of the distance of ensemble members to the projection of the observations into ensemble space. We also study the average size of particle moves when uncertainty is employed for individual Gaussian particles within the LMCPF.

Furthermore, we showed that the upper air first guess errors of the LMCPF and LETKF during the assimilation cycle are very similar within a range of

plus-minus 1-3%, with the LMCPF being better below 850 hPa and the LETKF being better in some range above. Forecast scores demonstrating as well that the RMSE of the ensembles is comparable for upper air temperature, relative humidity, wind fields and pressure (2-3%).
In several shorter case studies we demonstrate that by varying the parameter choices, we can achieve better first guess RMSE for the LMCPF in comparison to the LETKF, which shows that for very short range forecasts the quality of the method can be comparable to or better than that of the LETKF.

Overall, with the LMCPF we demonstrate significant progress compared to the LAPF for NWP in an operational setup, demonstrating that the LMCPF has reached a stability and quality comparable to that of the LETKF, while allowing and accounting for various non-Gaussian distributions in its analysis steps.
However, the LMCPF in this thesis is almost untuned, e.g. we haven't had a closer look at the observation handling. There is also a lot of work to be done to calibrate e.g. the tuning constants in (4.1.45) or the parameters of Table 4.1.

6.3 Future Work

Finally, to complete this thesis, what will be done next?

Of course, there is a lot of further work necessary. While reaching a break-even point for operational scores with a new method establishes an important milestone, we need to note that there are many open and intricate scientific questions here with respect to the choice of parameters for the Gaussian mixture and their dependence on the control of an optimal and correct ensemble spread both in the analysis cycle and for the forecasts. Also, in further steps we will take a look at the quality control. Currently, the LMCPF and the LETKF are using the same observation quality control, but the LMCPF seems to need a more accurate approach.

Besides that, the studies carried out for this thesis only cover a the period of one month. It would also be interesting to study the behaviour of the LMCPF over a longer period of time, e.g. a whole year.

Moreover, we have implemented the LAPF and LMCPF in the Lorenz 63 and Lorenz 96 models and have studied the characteristics of the particle filters in low-dimensional systems (Schenk et al., 2022).

Recently, the LMCPF has been implemented in the operational regional ICON-D2 model of the DWD. As the time variable in this regional setup is much more important as in the global setup, implementing a particle filter in this setup will lead into a 4d-particle filter version.

Furthermore, it would also be interesting to test the LMCPF in the ICON model with Aerosols and Reactive Trace gases (ICON-ART) (operational since end of March 2024), which is able to simulate the spatiotemporal evoultion of aerorols and trace gases (Rieger et al., 2015; Schröter et al., 2018). As the LMCPF should be able to handle with extreme events better than the LETKF, because it can give a higher weight to a few particles closer to the observations, the LMCPF should be able to outperform the LETKF in these cases.

In addition to the ICON-ART model, the DWD is also developing an ocean model (ICON-O, see Korn (2017)) and a coupling between this ocean model and the atmospheric ICON model. Again, we could be the first to study the behaviour of the LMCPF in a high-dimensional global ocean model or a fully coupled ocean-atmosphere model and to verify that the LMCPF gives reasonable results even in the case of the slow changes in the ocean.

Of course, there are many more models and applications in which particle filters can be used. For example, outside the weather forecasting community Kalman Filters and particle filters are used to model the yield of crops (compare e.g. Zare et al. (2022); Fattori Junior et al. (2022)). Again, the LMCPF should be able to to outperform the Kalman filters because of its ability to give the most

weight to the particles clostest to the observations and to move them towards the observations.

Finally, we can conclude that we have developed and implemented the LMCPF in the ICON model, which is able to outperform the reference LETKF in some cases. With this basic work done, we are now able to further tune and improve the LMCPF and reach a level where it should be able to completely outperform the LETKF abd where we are able to use the LMCPF operationally.

Bibliography

Ades, M. and van Leeuwen, P. J. An exploration of the equivalent weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, 139:820–840, 2013. doi: 10.1002/qj.1995.

Ades, M. and van Leeuwen, P. J. The equivalent-weights particle filter in a high-dimensional system. *Quarterly Journal of the Royal Meteorological Society*, 141:484–503, 2015. doi: 10.1002/qj.2370.

Alam, S. A. and Gustafsson, O. Improved particle filter resampling architectures. 92:555–568, 6 2020. doi: 10.1007/s11265-019-01489-y.

Anderson, J. L. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, 230: 99–111, 2007. ISSN 0167-2789. URL `http://dx.doi.org/10.1016/j.physd.2006.02.011`.

Anderson, J. L. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129(12):2884–2903, 2001. doi: {10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2}.

Anderson, J. L. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59(2):210–224, 2007. doi: {10.1111/j.1600-0870.2006.00216.x}.

Anderson, J. L. Spatially and temporally varying adaptive covariance inflation

for ensemble filters. *Tellus A*, 61(1):72–83, 2009. doi: 10.1111/j.1600-0870.2008.0361.x.

Bain, A. and Crisan, D. *Fundamentals of Stochastic Filtering*. Stochastic Modelling and Applied Probability. Springer, 2009. ISBN 978-0-387-76896-0. doi: 10.1007/978-0-387-76896-0. URL http://dx.doi.org/10.1007/978-0-387-76896-0.

Bauer, P., Thorpe, A., and Brunet, G. The quiet revolution of numerical weather prediction. *Nature*, 525:47–55, 2015. doi: 10.1038.

Bengtsson, T., Snyder, C., and Nychka, D. Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003. doi: 10.1029/2002JD002900. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD002900.

Bickel, P., Li, B., and Bengtsson, T. *Sharp failure rates for the bootstrap particle filter in high dimensions*, volume 3 of *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, pages 318–329. Institute of Mathematical Statistics, Beachwood, Ohio, USA, 2008. doi: 10.1214/074921708000000228. URL https://doi.org/10.1214/074921708000000228.

Bishop, C. H. The GIGG-EnKF: ensemble kalman filtering for highly skewed non-negative uncertainty distributions. *Quarterly Journal of the Royal Meteorological Society*, 142(696):1395–1412, 2016. doi: 10.1002/qj.2742. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2742.

Bishop, C. H. and Hodyss, D. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 133(629):2029–2044, 2007. doi: {10.1002/qj.169}.

Bishop, C. H. and Hodyss, D. Ensemble covariances adaptively localized with eco-rap. Part 1: tests on simple error models. *Tellus A*, 61(1):84–96, 2009a. ISSN 1600-0870. doi: 10.1111/j.1600-0870.2008.00371.x. URL http://dx.doi.org/10.1111/j.1600-0870.2008.00371.x.

Bishop, C. H. and Hodyss, D. Ensemble covariances adaptively localized with eco-rap. Part 2: a strategy for the atmosphere. *Tellus A*, 61(1):97–111, 2009b. doi: 10.1111/j.1600-0870.2008.00372.x. URL http://dx.doi.org/10.1111/j.1600-0870.2008.00372.x.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly weather review*, 129(3):420–436, 2001. doi: {10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2}.

Bjerknes, V. Das problem der wettervorhersage, betrachtet von standpunkt der mechanik und der physik. *Meteorologische Zeitschrift*, 21:1–7, 1904.

Bloom, S. C., Takacs, L. L., da Silva, A. M., and Ledvina, D. Data assimilation using incremental analysis updates. *Monthly Weather Review*, 124:1256–1271, 1996. URL https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.

Brett, C., Lam, K., Law, K., McCormick, D., Scott, M., and Stuart, A. Accuracy and stability of filters for dissipative pdes. *Physica D: Nonlinear Phenomena*, 245(1):34–45, 2013. ISSN 0167-2789. doi: https://doi.org/10.1016/j.physd.2012.11.005. URL https://www.sciencedirect.com/science/article/pii/S0167278912002953.

Browne, P. A. and van Leeuwen, P. J. Twin experiments with the equivalent weights particle filter and hadcm3. *Quarterly Journal of the Royal Meteorological Society*, 141(693):3399–3414, 2015. doi: https://doi.org/10.1002/qj.2621.

Brusdal, K., Brankart, J. M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P. J., Dombrowsky, E., and Verron, J. A demonstration of ensemble-based assimilation methods with a layered ogcm from the perspective of operational ocean forecasting systems. *Journal of Marine Systems*, 40-41:253–289, 2003. doi: 10.1016/S0924-7963(03)00021-6.

Burgers, G., van Leeuwen, P. J., and Evensen, G. Analysis scheme in the ensemble kalman filter. *Monthly Weather Review*, 126(6):1719–1724, 1998. doi: {10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2}.

Campbell, W. F., Bishop, C. H., and Hodyss, D. Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Monthly Weather Review*, 138(1):282–290, 2010. ISSN 10027-0644. doi: 10.1175/2009MWR3017.1. URL `http://dx.doi.org/10.1175/2009MWR3017.1`.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Climate Change*, 9(5):e535, 2018. doi: https://doi.org/10.1002/wcc.535. URL `https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.535`.

Crisan, D. and Rozovskii, B. *The Oxford Handbook of Nonlinear Filtering*. Oxford handbooks in mathematics. Oxford University Press, 2011. ISBN 9780199532902. URL `http://EconPapers.repec.org/RePEc:oxp:obooks:9780199532902`.

Daley, R. *Atmospheric Data Analysis*. Cambridge Atmospheric and Space Science Series. Cambridge University Press, 1993. ISBN 9780521458252. URL `http://books.google.de/books?id=RHM6pTMRTHwC`.

Damrath, U. Verifikation von wettervorhersagen. *Promet*, 28(1/2):8 – 16, 2002.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P. Diagnosis of observation, background and analysis-error statistics in observation space. *Quar-*

terly Journal of the Royal Meteorological Society, 131:3385–3396, 2005. doi: 10.1256/qj.05.108.

Doucet, A., de Freitas, N., and Gordon, N. *Sequential Monte Carlo Methods in Pratice*. Information Science and Statistics. Springer, 2001. ISBN 9781475734379. doi: 10.1007/978-1-4757-3437-9. URL https://www.springer.com/de/book/9780387951461.

European Centre for Medium-Range Weather Forecasts. Parameter detail: Geopotential height. https://codes.ecmwf.int/grib/param-db/156, 2019. Online, accessed March 01, 2024.

Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994. ISSN 2156-2202. doi: 10.1029/94JC00572. URL http://dx.doi.org/10.1029/94JC00572.

Evensen, G. *Data Assimilation: The Ensemble Kalman Filter*. Earth and Environmental Science. Springer, 2009. ISBN 9783642037115. URL http://books.google.de/books?id=2_zaTb_O1AkC.

Evensen, G. and van Leeuwen, P. J. An ensemble kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128:1852–1867, 2000. ISSN 60027-0644. doi: doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2. URL http://dx.doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Evensen, G., Vossepoel, F. C., and van Leeuwen, P. J. *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. Springer International Publishing, 2022. ISBN 978-3-030-96709-3. doi: 10.1007/978-3-030-96709-3_9. URL https://doi.org/10.1007/978-3-030-96709-3_9.

Eyre, J. R., Bell, W., Cotton, J., English, S. J., Forsythe, M., Healy, S. B., and Pavelin, E. G. Assimilation of satellite data in numerical weather prediction.

part ii: Recent years. *Quarterly Journal of the Royal Meteorological Society*, 148(743):521–556, 2022. doi: https://doi.org/10.1002/qj.4228. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4228.

Farchi, A. and Bocquet, M. Comparison of local particle filters and new implementations. *Nonlinear Processes in Geophysics*, 25:765–807, 2018. doi: 10.5194/npg-25-765-2018. URL https://doi.org/10.5194/npg-25-765-2018.

Fattori Junior, I. M., dos Santos Vianna, M., and Marin, F. R. Assimilating leaf area index data into a sugarcane process-based crop model for improving yield estimation. *European Journal of Agronomy*, 136:126501, 2022. ISSN 1161-0301. doi: 10.1016/j.eja.2022.126501. URL https://www.sciencedirect.com/science/article/pii/S1161030122000491.

Frei, M. and Künsch, H. Bridging the ensemble kalman and particle filters. *Biometrika*, 100(4):781–800, 2013. doi: 10.1093/biomet/ast020. URL https://doi.org/10.1093/biomet/ast020.

Gaspari, G. and Cohn, S. E. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125 (554):723–757, 1999. ISSN 1477-870X. doi: 10.1002/qj.49712555417. URL http://dx.doi.org/10.1002/qj.49712555417.

Gassmann, A. and Herzog, H.-J. Towards a consistent numerical compressible non-hydrostatic model using gerneralized hamiltonian tools. *Quarterly Journal of the Royal Meteorological Society*, 134:1597–1613, 2008. doi: 10.1002/qj.297.

Geer, A. J. and Bauer, P. Observation errors in all-sky data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 137(661):2024–2037, 2011. doi: https://doi.org/10.1002/qj.830. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.830.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. Novel approach to

nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings-F*, 140(2): 107–113, 1993.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R. Balance and ensemble kalman filter localization techniques. *Monthly Weather Review*, 139 (2):511–522, 2011. doi: 10.1175/2010MWR3328.1.

Hoteit, I., Pham, D.-T., Triantafyllou, G., and Korres, G. A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. *Monthly Weather Review*, 136:317–334, 2008. doi: 10.1175/2007MWR1927.1.

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

Houtekamer, P. L. and Mitchell, H. L. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137, 2001. ISSN 10027-0644. doi: doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2. URL `http://dx.doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2`.

Houtekamer, P. L. and Mitchell, H. L. Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3269–3289, 2005. ISSN 1477-870X. doi: 10.1256/qj.05.135. URL `http://dx.doi.org/10.1256/qj.05.135`.

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review*, 133 (3):604–620, 2005. ISSN 30027-0644. doi: doi:10.1175/MWR-2864.1. URL `http://dx.doi.org/10.1175/MWR-2864.1`.

Hu, C.-C. and van Leeuwen, P. J. A particle flow filter for high-dimensional system applications. *Quarterly Journal of the Royal Meteorological Society*,

147(737):2352–2374, 2021. doi: https://doi.org/10.1002/qj.4028. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4028.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1-2):112–126, 2007. ISSN 0167-2789. doi: http://dx.doi.org/10.1016/j.physd.2006.11.008. URL http://www.sciencedirect.com/science/article/pii/S0167278906004647.

Janjić, T., Nerger, L., Albertella, A., Schröter, J., and Skachko, S. On domain localization in ensemble-based Kalman filter algorithms. *Monthly Weather Review*, 139(7):2046–2060, 2011. ISSN 70027-0644. doi: doi:10.1175/2011MWR3552.1. URL http://dx.doi.org/10.1175/2011MWR3552.1.

Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2002. doi: 10.1017/CBO9780511802270.

Kalnay, E., Hunt, B., Ott, E., and Szunyogh, I. *Ensemble forecasting and data assimilation: two problems with the same solution?*, page 157–180. Cambridge University Press, 2006. doi: 10.1017/CBO9780511617652.008.

Kawabata, T. and Ueno, G. Non-gaussian probability densities of convection initiation and development investigated using a particle filter with a storm-scale numerical weather prediction model. *Monthly Weather Review*, 148:3–20, January 2020. doi: 10.1175/MWR-D-18-0367.1.

Korn, P. Formulation of an unstructured grid model for global ocean dynamics. *Journal of Computational Physics*, 339:525–552, 2017. ISSN 0021-9991. doi: 10.1016/j.jcp.2017.03.009. URL https://www.sciencedirect.com/science/article/pii/S0021999117301961.

Kotsuki, S., Miyoshi, T., Kondo, K., and Potthast, R. A local particle filter and its gaussian mixture extension implemented with minor modifications to the letkf. *Geoscientific Model Development*, 15(22):8325–8348, 2022. doi:

10.5194/gmd-15-8325-2022. URL https://gmd.copernicus.org/articles/15/8325/2022/.

Lang, O., van Leeuwen, P. J., Crisan, D., and Potthast, R. Bayesian inference for fluid dynamics: A case study for the stochastic rotating shallow water model. *Frontiers in Applied Mathematics and Statistics*, 8, 2022. ISSN 2297-4687. doi: 10.3389/fams.2022.949354. URL https://www.frontiersin.org/articles/10.3389/fams.2022.949354.

Li, H., Kalnay, E., and Miyoshi, T. Simultaneous estimation of covariance inflation and observation erros within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135:523–533, 2009. doi: 10.1002/qj.371.

Liu, B., Ait-El-Fquih, B., and Hoteit, I. Efficient kernel-based ensemble gaussian mixture filtering. *Monthly Weather Review*, 144(2):781–800, 2016a. doi: 10.1175/MWR-D-14-00292.1.

Liu, B., Gharamti, M. E., and Hoteit, I. Assessing clustering strategies for gaussian mixture filtering a subsurface contaminant model. *Journal of Hydrology*, 535:1–21, 2016b. doi: 10.1016/j.jhydrol.2016.01.048.

Lorenz, E. N. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20(2):130–141, 1963. doi: 10.1175/1520-0469(1963)0202.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/20/2/1520-0469_1963_020_0130_dnf_2_0_co_2.xml.

Lorenz, E. N. *The Essence of Chaos*. University of Washington Press, 1995. ISBN 978-0-2959-7514-6. URL https://uwapress.uw.edu/book/9780295975146/the-essence-of-chaos/.

Miyoshi, T. The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform kalman filter. *Monthly Weather Review*, 139(5):1519–1535, 2011. doi: 10.1175/2010MWR3570.1.

Miyoshi, T. and Kondo, K. A multi-scale localization approach to an ensemble kalman filter. *SOLA*, 9:170–173, 2013. doi: 10.2151/sola.2013-038.

Miyoshi, T. and Sato, Y. Assimilating satellite radiances with a local ensemble transform kalman filter (LETKF) applied to the jma global model (GSM). *SOLA*, 3:37–40, 2007. doi: 10.2151/sola.2007-010.

Miyoshi, T., Yamane, S., and Enomoto, T. Localizing the error covariance by physical distances within a local ensemble transform kalman filter (LETKF). *SOLA*, 3:89–92, 2007. doi: 10.2151/sola.2007-023.

Miyoshi, T., Kondo, K., and Imamura, T. The 10,240-member ensemble kalman filtering with an intermediate agcm. *Geophysical research Letters*, 41(14):5264–5271, 2014. doi: 10.1002/2014GL060863.

Nakamura, G. and Potthast, R. *Inverse Modeling*. 2053-2563. IOP Publishing, 2015. ISBN 978-0-7503-1218-9. doi: 10.1088/978-0-7503-1218-9. URL `http://dx.doi.org/10.1088/978-0-7503-1218-9`.

Nicely, M. A. and Wells, B. E. Improved parallel resampling methods for particle filtering. *IEEE Access*, 7:47593–47604, 2019. doi: 10.1109/ACCESS.2019.2910163.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, 56(5):415–428, 2004. ISSN 1600-0870. doi: 10.1111/j.1600-0870.2004.00076.x. URL `http://dx.doi.org/10.1111/j.1600-0870.2004.00076.x`.

Parrish, D. F. and Derber, J. C. The national meteorological center's spectral statistical- interpolation analysis system. *Monthly Weather Review*, 120 (8):1747–1763, 1992. ISSN 80027-0644. doi: http://dx.doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2. URL `http://dx.doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2`.

Penny, S. G. and Miyoshi, T. A local particle filter for high-dimensional geophysical systems. *Nonlinear Processes in Geophysics*, 23:391–405, 2016. doi: 10.5194/npg-23-391-2016. URL www.nonlin-processes-geophys.net/23/391/2016/doi:10.5194/npg-23-391-2016.

Periáñez, A., Reich, H., and Potthast, R. Optimal localization for ensemble Kalman filter systems. *Journal of the Meteorological Society of Japan*, 92(6): 585–597, 2014. doi: http://doi.org/10.2151/jmsj.2014-605.

Pham, D. T., Verron, J., and Roubaud, M. C. A singular evolutive extended kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16:323–340, 1998.

Pinheiro, F. R., van Leeuwen, P. J., and Geppert, G. Efficient nonlinear data assimilation using synchronization in a particle filter. *Quarterly Journal of the Royal Meteorological Society*, 145(723):2510–2523, 2019. doi: https://doi.org/10.1002/qj.3576. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3576.

Poterjoy, J. A localized particle filter for high-dimensional nonlinear system. *Monthly Weather Review*, 144(1):59–76, 2016. doi: 10.1175/MWR-D-15-0163.1. URL https://doi.org/10.1175/MWR-D-15-0163.1.

Poterjoy, J. and Anderson, J. L. Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter. *Monthly Weather Review*, (144):2007–2020, 2016. doi: 10.1175/MWR-D-15-0322.1.

Poterjoy, J., Sobash, R. A., and Anderson, J. L. Convective-scale data assimilation for the weather research forecasting model using the local particle filter. *Monthly Weather Review*, 145:1897–1918, 2017. doi: 10.1175/MWR-D-16-0298.1.

Poterjoy, J., Wicker, L., and Buehner, M. Progress toward the application of a localized particle filter for numerical weather prediction. *Monthly Weather Review*, 147(4):1107 – 1126, 2019. doi: 10.1175/MWR-D-17-0344.1.

Potthast, R., Walter, A., and Rhodin, A. A localized adaptive particle filter within an operational nwp framework. *Monthly Weather Review*, 147(1):345–362, 2019. doi: 10.1175/MWR-D-18-0028.1.

Reich, S. A nonparametric ensemble transform method for bayesian inference. *SIAM Journal on Scientific Computing*, 35(4):A2013–A2024, 2013. doi: 10.1137/130907367.

Reich, S. and Cotter, C. *Probabilistic Forecatsing and Bayesian Data Assimilation*. Cambridge University Press, 2015. ISBN 9781107706804. doi: 10.1017/CBO9781107706804. URL https://doi.org/10.1017/CBO9781107706804.

Reinert, D., Prill, F., Frank, H., Denhrd, M., and Zängl, G. Dwd doku. https://www.dwd.de/SharedDocs/downloads/DE/modelldokumentationen/nwv/icon/icon_dbbeschr_aktuell.pdf?view=nasPublication&nn=13934, 2018. Online, accessed January 5, 2023.

Richardson, L. F. Weather prediction by numerical process. *Cambridge at the University Press*, pages 1–258, 1922.

Rieger, D., Bangert, M., Bischoff-Gauss, I., Förstner, J., Lundgren, K., Reinert, D., Schröter, J., Vogel, H., Zängl, G., Ruhnke, R., and Vogel, B. Icon–art 1.0 – a new online-coupled model system from the global to regional scale. *Geoscientific Model Development*, 8(6):1659–1676, 2015. doi: 10.5194/gmd-8-1659-2015. URL https://gmd.copernicus.org/articles/8/1659/2015/.

Robert, S., Leuenberger, D., and Künsch, H. R. A local ensemble transform Kalman particle filter for convective-scale data assimilation. *Quarterly Journal*

of the Royal Meteorological Society, 2017. ISSN 1477-870X. doi: 10.1002/qj.3116. URL http://dx.doi.org/10.1002/qj.3116.

Rojahn, A., Schenk, N., van Leeuwen, P. J., and Potthast, R. Particle filtering and gaussian mixtures - on a localized mixture coefficients particle filter (lmcpf) for global nwp. *Journal of the Meteorological Society of Japan. Ser. II*, advpub:2023–015, 2023. doi: 10.2151/jmsj.2023-015.

Schenk, N., Potthast, R., and Rojahn, A. On two localized particle filter methods for lorenz 1963 and 1996 models. *Frontiers in Applied Mathematics and Statistics*, 8, 2022. ISSN 2297-4687. doi: 10.3389/fams.2022.920186.

Schröter, J., Rieger, D., Stassen, C., Vogel, H., Weimer, M., Werchner, S., Förstner, J., Prill, F., Reinert, D., Zängl, G., Giorgetta, M., Ruhnke, R., Vogel, B., and Braesicke, P. Icon-art 2.1: a flexible tracer framework and its application for composition studies in numerical weather forecasting and climate simulations. *Geoscientific Model Development*, 11(10):4043–4068, 2018. doi: 10.5194/gmd-11-4043-2018. URL https://gmd.copernicus.org/articles/11/4043/2018/.

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136:4629–4640, Dec 2008. doi: 10.1175/2008MWR2529.1.

Snyder, C., Bengtsson, T., and Morzfeld, M. Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143:4750–4761, Nov 2015. doi: 10.1175/MWR-D-15-0144.1.

Stordal, A. S., Karlsen, H. A., Naevdal, G., Skaug, H. J., and Vallès, B. Bridging the ensemble Kalman filter and particle filters: the adaptive gaussian mixture filter. *Computational Geosciences*, 15(2):293–305, 2011. doi: 10.1007/s10596-010-9207-1.

the ICON partnership (MPI-M, DWD, KIT, DKRZ, and C2SM). Icon. https://www.icon-model.org/, 2024. Online, accessed April 3, 2024.

van Leeuwen, P. J. A variance-minimizing filter for large-scale applications. *Monthly Weather Review*, 131:2071–2084, 2003a.

van Leeuwen, P. J. Nonlinear ensemble data assimilation for the ocean. *ECMWF Seminar Series*, 8:265–286, 2003b.

van Leeuwen, P. J. Particle filtering in geophysical systems. *Monthly Weather Review*, 137(12):4089–4114, 2009. doi: 10.1175/2009MWR2835.1.

van Leeuwen, P. J. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, (136):1991–1999, 2010. doi: 10.1002/qj.699.

van Leeuwen, P. J., Cheng, Y., and Reich, S. *Nonlinear Data Assimilation*. Frontiers in Applied Dynamical Systems: Reviews and Tutorials. Springer, 2015. ISBN 978-3-319-18347-3. doi: 10.1007/978-3-319-18347-3. URL http://dx.doi.org/10.1007/978-3-319-18347-3.

van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., and Reich, S. Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*, 145(723):2335–2365, 2019. doi: 10.1002/qj.3551. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3551.

Vetra-Carvalho, S., van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., Kirchgessner, P., and Beckers, J.-M. State-of-the-art stochastic data assimilation methods for high-dimensional non-gaussian problems. *Tellus A: Dynamic Meteorology and Oceanography*, 70(1):1–43, 2018. doi: 10.1080/16000870.2018.1445364.

Wang, P., Zhu, M., Chen, Y., and Zhang, W. Implicit equal-weights variational

particle smoother. *Atmosphere*, 11(4), 2020. ISSN 2073-4433. doi: 10.3390/atmos11040338. URL https://www.mdpi.com/2073-4433/11/4/338.

Whitaker, J. S. and Hamill, T. M. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913–1924, 2002. doi: 10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2. URL http://dx.doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Whitaker, J. S. and Hamill, T. M. Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140:3078–3089, 2012. doi: 10.1175/MWR-D-11-00276.1.

Yang, S.-C., Kalnay, E., Hunt, B., and Bowler, N. E. Weight interpolation for efficient data assimilation with the local ensemble transform kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(638):251–262, 2009. doi: 10.1002/qj.353. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.353.

Zängl, G. private communication, April 2024.

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M. The icon (icosahedral non-hydrostatic) modelling framework of dwd and mpi-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Meteorological Society*, (141):563–579, March 2014. doi: 10.1002/qj.2378.

Zare, H., Weber, T. K. D., Ingwersen, J., Nowak, W., Gayler, S., and Streck, T. Combining crop modeling with remote sensing data using a particle filtering technique to produce real-time forecasts of winter wheat yields under uncertain boundary conditions. *Remote Sensing*, 14(6), 2022. ISSN 2072-4292. doi: 10.3390/rs14061360. URL https://www.mdpi.com/2072-4292/14/6/1360.

Zhu, M., van Leeuwen, P. J., and Amezcua, J. Implicit equal-weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, (142):1904–1919, 2016. doi: 10.1002/qj.2784.