ABSTRACT

Title of Document: LOCAL ENSEMBLE TRANSFORM KALMAN FILTER WITH REALISTIC OBSERVATIONS

Hong Li, Doctor of Philosophy, 2007

Directed By: Professor Eugenia Kalnay
Department of Atmospheric and Oceanic Science

The main goal of my research is to improve the performance of the EnKF in assimilating real observations in order to accelerate the development of EnKF systems towards operational applications. A Local Ensemble Transform Kalman Filter (LETKF, Hunt et al. 2007) is used as an efficient representative of other EnKF systems. This dissertation has addressed several issues relating to the EnKF for assimilating real data.

The first issue is model errors. We assimilated observations generated from the NCEP/NCAR reanalysis fields into the SPEEDY model. The performance of the LETKF without accounting for model errors is seriously degraded compared with that in the perfect model scenario. We then investigated several methods to handle model errors including model bias and system-noise. Our results suggest that the pure bias removal methods (DdSM and LDM) are not able to beat the multiplicative or additive inflation schemes that account for the effects of total model errors. By contrast, when the bias removal methods (DdSM+ and LDM+) are supplemented by additive noise for representing the system-noise, they outperform the inflation schemes. Of these

augmented methods, the LDM+, where the constant bias, diurnal bias and state-dependent errors are estimated from a large sample of 6-hour forecast errors, gives the best results.

The other two issues addressed are the estimation of the inflation factor and of observation error variance. Without the accurate observation error statistics, a scheme for adaptively estimating inflation alone does not work, and vice versa. We propose to estimate simultaneously both the adaptive inflation and observation error variance. Our results for the Lorenz-96 model examples suggest that the simultaneous approach works perfectly in the perfect model scenario and in the presence of random model errors. For the case of systematic model bias, although it underestimates the observation error variance, our algorithm produces analyses that are comparable with the best tuned inflation value. SPEEDY model experiments indicate that our method is able to retrieve the true error variance for different types of instrument separately when applied to a more realistic high-dimension model.

Our research in this dissertation suggests the need to develop a more advanced LETKF with both bias correction and adaptive estimation of inflation within the system.

LOCAL ENSEMBLE TRANSFORM KALMAN FILTER WITH REALISTIC OBSERVATIONS

By

Hong Li

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor Eugenia Kalnay, Chair/Advisor
Professor Brian Hunt
Professor James Carton
Professor Ernesto Hugo Berbery
Professor Istvan Szunyogh

Acknowledgements

I am grateful to a number of people for supporting me in all kinds of ways. First of all, I would like to thank my advisor Prof. Eugenia Kalnay for her inmeasurable support, valuable guidance and warm encouragement. I am deeply indebted to her for helping me through the accomplishment of this manuscript during the difficult time when she lost her husband last month. I truly appreciate my committee members: Prof. Istvan Szunyogh, Prof. Brian Hunt, Prof. James Carton, and Prof. Ernesto Hugo Berbery for their kind support and insightful suggestions. I thank Ricardo Todling for helpful discussions and suggestions on model bias estimation.

I owe a debt of gratitude to my parents Yuzhu Li and Wei Wei, my husband Renzhi Lu and my son Jiawei Lu for their invaluable help and patience to encourage me to achieve my goal even in living alone in America for five years. This dissertation could not have been possible without their countless support.

My special thanks go to Eugenia's husband, Malise Dick who passed away last month. He always helped improve my English and revised part of this manuscript. Malise and Eugenia have given me their warm care and shown me many kindnesses.

I appreciate Drs. Takemasa Miyoshi and Chris Danforth for providing source code and suggestions. My version of LETKF on the SPEEDY model is built on the work of Takemasa, who first implemented the LEKF to the SPEEDY model. I also thank students in the weather chaos group, especially Dr. Shu-Chih Yang, Junjie Liu, Dr. Elana Fertig, Seung-Jong Baek and Dagmar Merkova for fruitful discussions. Finally I would like to acknowledge Emily Becker for reading and improving this manuscript.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Data assimilation algorithms seek to find the optimal combination of model forecast ("background") and the available observations to generate improved initial conditions ("analysis") for numerical weather predictions. Most assimilation schemes are based on the linear estimation theory in which the background and the observations are given a weight proportional to the inverse of their corresponding specified error covariances. As such, the accuracy of a data assimilation scheme relies highly on the knowledge of the error statistics of both the model background and the observations. The observation error covariance is usually assumed to be diagonal and time invariant. It is not a good approximation to assume the background error covariance is also stationary since it is actually flow-dependent. In practice, however, estimating the time-dependent background error covariance is difficult. In 3DVAR (e.g. Parrish and Derber, 1992), a data assimilation scheme used in many operational centers, the background error covariance is assumed to be isotropic and stationary. In the Kalman filter (KF, Kalman 1960), the background error covariance is propagated explicitly with a linear model. Ensemble-based Kalman filter (EnKF) techniques instead estimate the background error covariances from an ensemble of forecasts which allows them in theory to include information on the flow-dependent error of day (both temporally and spatially variant). For more background of the EnKF, the readers are recommended to refer, for example, Evensen (1994), Houtekamer and

Mitchell (1998), Anderson (2001), Whitaker and Hamill (2002), Tippett et al. (2003), Ott et al. (2004), Evensen (2003).

Most studies to date have tested EnKF systems under perfect model assumptions and with simulated observations (their error statistics are perfectly known). Only within the last few years have EnKF methods been tested in assimilating real observations. In the real-world applications, several issues have to be dealt with, such as: 1) model errors, 2) inconvenience or infeasibility of manually tuning the inflation factor when it is regional and/or variable dependent, 3) erroneously specified observation error statistics, 4) imperfect forward observation operator, 5) non-Gaussianity of forecast and observation errors, etc. In this dissertation, we focus on the first three issues. Methods to deal with them are investigated and tested.

1.1 EnKF in the presence of model errors

EnKF methods have been shown to be more accurate than 3D-Var under the assumption of a perfect model (Miyoshi 2005, Liu et al 2006). However, in the real world, forecast errors derive not only from errors in the initial conditions but also from errors due to the model deficiencies. The latter type of error is usually called *model error*. The sources of model error can be due to lack of resolution, approximate parameterizations of physical processes, numerical dispersion, etc. For assimilation of real observations, the assumption of a perfect model must be dropped. As a result, there is no guarantee that the EnKF will be still better than 3D-Var data assimilation systems when assimilating real observations. In fact, Miyoshi (2005) has shown that model error has a stronger negative influence on the performance of the EnKF than

on the 3D-Var. Accounting for model errors is important but is also the most difficult issue for EnKF systems.

Methods to deal with model errors were introduced decades ago in the engineering community. In the meteorological community, until recently, solving this problem has been receiving widespread attention. Dee and da Silva (1998, hereafter referred as DdS) proposed a method for the on-line estimation and correction of model bias where the prior estimate of the bias is updated by using the Kalman filter. This bias correction method (DdSM, hereafter) has been successfully tested, for example, by Dee and Todling (2000), Carton et al. (2000), Martin et al. (2002), Chepurin et al. (2005). Recently, Baek et al (2006) developed another bias correction method which is similar to the DdSM except for accounting for the cross-correlation of uncertainties in model state and bias that had been ignored in the DdSM. They successfully tested this approach with the Lorenz-96 model. However, both of these two methods assume a steady forecast model for the bias. This results in being limited to estimate only the slowly varying component of forecast errors. In reality, model error may be time-variant. To correct the evolving model error, an empirical correction that depends on the instantaneous state of the model has been proposed to correct the 'state-dependent error' (Leith 1978; DelSole and Hou 1999). This method relies on the cross covariance between the model error and the model state. Direct computation of the bias correction term on each model grid point is in practice prohibitive for the operational forecast model. Building on works of Leith (1978), Danforth et al (2007) proposed an alternative approach based on SVD to estimate the state-dependent error with much lower computational cost. Danforth et al. also

corrected the state-independent model error by expanding model errors into a bias and low order EOFs that could correct other errors such as those associated with the diurnal cycle. They found this low-dimensional method (LDM, hereafter) to be very successful and efficient computationally. However, in their experiments there is no data assimilation involved, since the initial conditions are assumed to be perfect. It is worthwhile to investigate whether the application of the LDM can be expanded to more realistic situations where the forecast-analysis is cycled and, as a result, forecast error includes both model error and dynamical growing error due to the imperfect initial condition.

All of the methods mentioned above are meant to estimate and remove model bias from the forecasts, which can be applied to any data assimilation system. For EnKF systems, enlarging background error variance is another way to handle model errors. Anderson and Anderson (1999) introduced the idea of multiplicative inflation. Corazza et al. (2002) added random perturbations to bred vectors and found it improved their ability to represent background errors. Corazza et al (2007) found the additive inflation worked much better than the multiplicative inflation in the LEKF (Ott et al. 2004) within a perfect model experiment. Although the multiplicative and additive inflation were initially introduced to ameliorate sampling errors due to small ensemble size in the perfect model scenario, they have been recently used to account for model errors in assimilating real observations. Whitaker et al. (2004) used multiplicative inflation and obtained a better reanalysis than the NCEP 3D-Var in the middle and low troposphere from real surface pressure observations. Houtekamer et al. (2005) parameterized model errors with 'additive error' by adding random noise to

each forecast ensemble member. The covariance of the random noise has the same structure and scaled-down amplitude as the 3D-Var background error covariance. The results showed that the quality of their ensemble Kalman filter was comparable to 3D-Var using real observations. Hamill et al. (2005) compared different ways to parameterize additive error. Whitaker et al (2007) generated additive noise by randomly selecting samples from the 6-hour tendencies field of the NCEP/NCAR reanalysis (NNR, Kalnay et al. 1996) states and found it gave more accurate analysis than the multiplicative inflation when assimilating real observations using the NCEP GFS model. All of these methods are specific to EnKF systems. They enlarge the background ensemble spread without changing the ensemble mean. To our best knowledge, there have been no comparisons so far between these variance-enlarging methods and those estimating and removing model error from the ensemble mean (e.g. DdSM and LDM).

1.2 Adaptive estimation of inflation factor

Though multiplicative and additive inflation schemes are widely used in EnKF systems in both the perfect and imperfect model experiments, these adjunct algorithms require considerable tuning for good performance. Manually tuning the inflation parameter is common in the EnKF experiments but is expensive, since, on its own, the forecast-analysis cycle does require many ensemble members. Even worse, it becomes infeasible if the inflation factor is regional and/or variable dependent. Wang and Bishop (2003) adopt the maximum likelihood parameter estimation theory of Dee (1995) to estimate the inflation factor from the innovation

observation-minus-background statistics $\mathbf{d}^T\mathbf{d} = trace[(1+\Delta)\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}]$ in their ensemble forecast scheme. Miyoshi (2005) reported the use of a similar method to estimate the background error inflation factor within EnKF. Although the results were satisfactory it is obvious that this inflation estimation method relies on the assumption of a perfect knowledge of the observational error covariance $\mathbf{R}$. This assumption is valid for the simulated observations but may not for real observations. When assimilating real data, we need a method to obtain the correct statistics of observation errors if we want to apply the scheme for on-line estimating inflation factor.

1.3 Diagnosis of observation error statistics

Besides the issue of the requirement of correct observation error statistics in the on-line estimation of the inflation factor discussed above, the observation error statistics themselves are very important to the data assimilation, since they directly determine the relative weight given to the observations. However, in the real world, this information is not perfectly known. Recent diagnostic work (Desroziers and Ivanov 2001, Talagrand 1999, Cardinali et al. 2004, Chapnik et al. 2006, and others) suggest that innovation and other statistics can be used to diagnose both observation and background errors. A formulation on the cost function of such diagnostics has been proposed and tested in a variational framework. Building on these works, Desroziers et al. (2005) (DEA05 hereafter) developed a set of diagnostics based on the combinations of *observation-minus-analysis*, *observation-minus-background* and *background-minus-analysis t*o adaptively tune observation and background errors. Here we adapt one of these diagnostics for estimating observation error variance into the EnKF. For more details, the reader is referred to section 2.4.2 in Chapter 2.

1.4 Specific goals of this dissertation

The main goal of this thesis is to improve the performance of the EnKF in assimilating real observations and, as a result, to accelerate the development of EnKF systems towards operational applications. To assimilate the real observations, model error is the most important and difficult issue to deal with. Methods have been proposed to either account for model error in the second moment of the ensemble by enlarging background error variance or to estimate and correct model bias for the ensemble mean. However, there is little research on comparing methods between these two approaches. In addition, the imperfect knowledge of observation error statistics and the inconvenience of manually tuning inflation factor are the other two issues we want to address in this thesis. For these purposes, the present research aims to investigate the following questions:

1. Are bias removal methods generally better than those only enlarging background error variance? What are the relative advantage and disadvantage between two major bias removal methods: DdSM and LDM?
2. Does on-line inflation estimation work without the correct observation error information? If not, could we develop a way to simultaneously estimate both the inflation and observation errors?

To achieve the goal, Local Ensemble Transform Kalman Filter (LETKF, Hunt et al. 2007), is chosen as a proxy of the EnKF in this study due to its efficiency and easy access for us, as it is developed at the University of Maryland. LETKF is developed from Local Ensemble Kalman Filter (Ott et al 2004, Szunyogh et al. 2005) but is computationally much faster. It has been used to assimilate simulated

observations in the NASA fvGCM model (Liu et al. 2006) and the results have shown it outperforms 3DVar everywhere. In simple model, it has been shown to perform similarly to 4D-Var (Kalnay et al. 2007). With real data, Szunyogh et al. (2007) reported that the LETKF is more accurate than the SSI (operational 3DVAR) in the SH extratropics, and comparable in NH extratropics and Tropics by simply using multiplicative inflation to account for model errors. Here we investigate more sophisticated techniques for treating model errors and methods for on-line estimating inflation factor within the LETKF, and to develop a data assimilation system capable of assimilating real weather observations. Though we focus on the LETKF, the results are applicable to other EnKF systems.

1.5 Outline of the thesis

Chapter 2 describes the theory of all the methods used in this entire dissertation, including the LETKF (section 2.2), techniques for treating model errors (section 2.3), and methods for on-line estimating inflation and observation errors (section 2.4). In Chapter 3, the LETKF is implemented in an AGCM model, the SPEEDY model, under the perfect model assumption, and its performance is examined. In Chapter 4, we drop the perfect model assumption by assimilating observations generated from the NCEP/NCAR reanalysis fields. The performance of the LETKF without accounting for model errors is examined and compared with that in the perfect model scenario. Two inflation schemes (multiplicative and additive inflation) and two bias correction methods (DdSM and LDM) are applied to account for and/or correct model errors. Their results are compared and discussed. Chapter 5 investigates the methods of adaptive estimation of the inflation factor with and without the accurate

observational error information. A new method of simultaneous estimation of both inflation and observation error is proposed and investigated in the cases of perfect and imperfect model. At last, the conclusions of the entire dissertation are given in Chapter 6. Some future research directions are also discussed in this chapter.

Chapter 2

Theoretical aspects

2.1 Introduction

Here we give a theoretical review of all the methods used in the rest of the dissertation. These are: (1) a data assimilation scheme, the local ensemble transform Kalman filter (LETKF) (Section 2.2); (2) methods to account for and correct model errors (Section 2.3); and (3) methods to on-line estimate the covariance inflation factor and observation error variance (Section 2.4).

2.2 Theory of the Local Ensemble Transform Kalman Filter (LETKF)

Here we give a review of the LETKF (Hunt 2004, Hunt et al. 2007), the only data assimilation method used in this research. As an introduction, the general concept of data assimilation and some related issues and terms are also briefly described.

2.2.1 Introduction to data assimilation

The notation in this study is based, wherever possible, on Ide et al. 1997. Let $\mathbf{x}_i^t$ be the unknown true atmospheric state at a certain analysis time i. Two available sources of information used to estimate $\mathbf{x}_i^t$ are the background $\mathbf{x}_i^b$, generally provided by a model forecast $\mathbf{x}_i^f$ with errors $\boldsymbol{\varepsilon}_i^f = \mathbf{x}_i^f - \mathbf{x}_i^t$, and an observation vector $\mathbf{y}_i^o$ with errors $\boldsymbol{\varepsilon}_i^o = \mathbf{y}_i^o - H(\mathbf{x}_i^t)$, where H is the non-linear observation operator, mapping the model variables to the observational space. The reason for us here to

utilize superscript f rather than widely used b to denote all background related variable is to reserve symbol b for the model bias (section 2.3). Both $\mathbf{x}_i^f$ and $\mathbf{x}_i^t$ are state vectors containing n elements, while $\mathbf{y}_i^o$ is a vector containing p observations at irregularly spaced points. The error covariance matrices for the forecast and the observations, $\mathbf{P}^f$ and $\mathbf{R}$, are then defined in terms of the true state as

$$\mathbf{P}^f = < \boldsymbol{\varepsilon}^f \boldsymbol{\varepsilon}^{f^T} > \tag{2.1}$$

$$\mathbf{R} = < \boldsymbol{\varepsilon}^o \boldsymbol{\varepsilon}^{o^T} > \tag{2.2}$$

where $\langle \cdot \rangle$ denotes the statistical expected value. Though both $\mathbf{P}^f$ and $\mathbf{R}$ are time-dependent, for simplicity, in this dissertation we drop the subscript i indexing time, when all the variables are valid at the same time, and include it whenever necessary.

With the statistical estimates of $\mathbf{P}^f$ and $\mathbf{R}$, data assimilation process seeks the best combination of the model forecasts $\mathbf{x}^f$ and the observations $\mathbf{y}^o$, to generate improved initial conditions ("analysis" $\mathbf{x}^a$).

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}^f)) \tag{2.3}$$

where $\mathbf{K}$ is the weighting matrix. From linear estimation theory, the estimate is optimal (the analysis error $\boldsymbol{\varepsilon}^a = \mathbf{x}^a - \mathbf{x}^t$ is a minimum) when the weighting matrix is given by the Kalman gain:

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \tag{2.4}$$

where $\mathbf{H}$ is the linear perturbation of the forward observational model H. The validity of the optimal $\mathbf{K}$ is dependent on the accuracy of the statistical estimates of the forecast and observation errors.

2.2.2 Forecast error covariance and the ensemble Kalman filter

The observation and forecast error covariance are the two key components in the standard Kalman gain given by (2.4). However, estimating these two covariances is a major challenge for data assimilation. In practice, the observation error covariance is assumed to be diagonal and stationary. In 3DVAR (e.g. Parrish and Derber 1992), a data assimilation scheme used in many operational centers, the forecast error covariance is assumed to be isotropic and stationary. This assumption is very crude since in reality the forecast error covariance depends on the current atmosphere state. In the traditional Kalman filter, the forecast and analysis error covariance matrices are evolved with time by

$$\mathbf{P}_i^f = \mathbf{M}_i \mathbf{P}_{i-1}^a \mathbf{M}_i^T + \mathbf{Q}_i \tag{2.5}$$

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i \mathbf{H}_i) \mathbf{P}_i^f \tag{2.6}$$

where $\mathbf{P}_{i-1}^a$ is the analysis error covariance at time step $i-1$, defined as $\mathbf{P}_i^a = < (\mathbf{x}_i^a - \mathbf{x}_i^t)(\mathbf{x}_i^a - \mathbf{x}_i^t)^T >$, $\mathbf{M}_i$ is the tangent linear model of the nonlinear dynamics, $\mathbf{H}_i$ is the linear perturbation of the forward observational model H, and $\mathbf{Q}_i$ is the model error covariance matrix, given by

$$\boldsymbol{\eta}_i = \mathbf{x}_i^t - M(\mathbf{x}_{i-1}^t) \tag{2.7}$$

$$\mathbf{Q}_i = \langle \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \rangle \tag{2.8}$$

where $\boldsymbol{\eta}_i$ denotes model errors, i.e. the forecast errors due to the model deficiencies rather than to the initial condition. (2.5) implies that the total forecast errors are the sum of the "internal error" (errors in the initial state and their dynamical growth during the forecast process) and the "external error" due to the model deficiencies. In

the traditional Kalman filter, $\boldsymbol{\eta}_i$ is assumed to be a noise process with zero mean. In reality model errors have significant biases which in principle should be removed from the forecasts before proceeding.

Combining (2.3) (2.4) (2.5) and (2.6) we obtain the traditional Kalman filter which gives the best analysis in the case of unbiased forecasts and observations and a linear forecast model.

$$\begin{cases} \mathbf{P}_i^f = \mathbf{M}_{i-1}\mathbf{P}_{i-1}^a\mathbf{M}_{i-1}^T + \mathbf{Q}_i \\ \mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \\ \mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}^f)) \\ \mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f \end{cases} \tag{2.9}$$

Here again, we have dropped the subscript i when all the terms are at the same time step i.

However, updating forecast-error covariance by (2.5) is unfeasible for real numerical models due to the huge dimensions of $\mathbf{M}$ and $\mathbf{P}^a$. An alternative to the traditional Kalman filter is the ensemble Kalman filter (EnKF), originally proposed by Evensen (1994). In the EnKF, the error evolution of (2.5) is computed by using ensemble integration. We assume that K forecasts $\mathbf{x}_{e(k)}^f$, k=1, 2,…, K have been created and therefore we define the ensemble covariance matrices around the ensemble mean, $\overline{\mathbf{x}}^f$,

$$\mathbf{P}_e^f = \frac{1}{K-1}\sum_{k=1}^{K}(\mathbf{x}_{e(k)}^f - \overline{\mathbf{x}}_e^f)(\mathbf{x}_{e(k)}^f - \overline{\mathbf{x}}_e^f)^T = \frac{1}{K-1}\mathbf{X}_e^f\mathbf{X}_e^{fT} \tag{2.10}$$

where the $n \times K$ matrix X_e^f is the ensemble perturbation matrix, whose k th column is $\mathbf{X}_{e(k)}^f = \mathbf{x}_{e(k)}^f - \overline{\mathbf{x}}_e^f$, and n is the dimension of model state. In principle, if the

ensemble covers all possible realizations of the actual atmospheric state, the ensemble error covariance $\mathbf{P}_e^f$ is a good approximation of the 'true' forecast error covariance. However, in practice, the ensemble members are usually generated by integrating the different initial model states forward in time using the same model. Therefore, the estimated $\mathbf{P}_e^f$ from those ensemble members can only represent the first term on the right-hand side of (2.5) due to the internal error but does not allow for the inclusion of a model error covariance $\mathbf{Q}$ (If the forecasts are made with different model parameters or different models, $\mathbf{P}_e^f$ will contain certain estimates of model error covariance). As a result, the forecast error covariance matrix should be given by

$$\mathbf{P}^f = \mathbf{P}_e^f + \mathbf{Q} \tag{2.11}$$

in EnKF and model error covariance matrix $\mathbf{Q}$ has to be specified. Since little is actually known about model error statistics in complex systems, a crude way to account for model error is to add additional noise with zero mean and a prescribed covariance $\mathbf{Q}$ to each forecast ensemble member.

$$\mathbf{x}_{(k)}^f = \mathbf{x}_{e(k)}^f + \mathbf{q}_k \tag{2.12}$$

$$\mathbf{Q} = \frac{1}{K-1}\sum_{k=1}^{K} \mathbf{q}_k \mathbf{q}_k{}^T \tag{2.13}$$

This will enlarge the dynamically evolved forecast perturbations without changing the ensemble mean. The final forecast error covariance matrix then is,

$$\begin{aligned}\mathbf{P}^f &= \frac{1}{K-1}\sum_{k=1}^{K}(\mathbf{x}_{e(k)}^f + \mathbf{q}_k - \overline{\mathbf{x}}_e^f) * (\mathbf{x}_{e(k)}^f + \mathbf{q}_k - \overline{\mathbf{x}}_e^f)^T \\ &= \frac{1}{K-1}\mathbf{X}^f \mathbf{X}^{fT}\end{aligned} \tag{2.14}$$

where $\mathbf{X}^f$ is the modified ensemble perturbation matrix, whose k th column is $\mathbf{X}^f{}_{(k)} = \mathbf{x}^f_{e(k)} + \mathbf{q}_k - \overline{\mathbf{x}}^f_e$

The following equations constitute the Ensemble Kalman filter,

$$\begin{cases} \mathbf{P}^f = \dfrac{1}{K-1}\mathbf{X}^f\mathbf{X}^{fT} \\ \mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \\ \mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}^f)) \\ \mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f \end{cases} \tag{2.15}$$

Although all EnKF systems share the same basic formulation, the implementations of (2.15) are not unique. Here we do not attempt to review all the ensemble data assimilation schemes but focus on the local ensemble transform Kalman filter (LETKF, Hunt et al. 2007).

2.2.3 Local ensemble transform Kalman filter

Several characteristics of the LETKF are: (1) the analysis is performed *locally* in model grid space and the analysis at the different grid points is obtained independently; (2) the forecast error uncertainty is estimated from the *ensemble* perturbations by using equation (2.14); (3) the analysis ensemble perturbations are obtained from the forecast ensemble perturbations through a *transform* matrix; (4) the matrix inverse is done in the ensemble space when compute the Kalman gain.

Directly computing the Kalman gain $\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}$ requires a matrix inverse in the observation space with a size of order 10^5-10^7. To implement it in an efficient way, LETKF re-writes the gain matrix $\mathbf{K}$ so that the matrix inverse can be done in ensemble space at every grid point.

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}$$

$$= \frac{1}{K-1}\mathbf{X}^f \mathbf{X}^{fT} \mathbf{H}^T (\mathbf{H}\frac{1}{K-1}\mathbf{X}^f \mathbf{X}^{fT} \mathbf{H}^T + \mathbf{R})^{-1}$$

$$= \mathbf{X}^f \widetilde{\mathbf{H}}^T [\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T + (K-1)\mathbf{R}]^{-1}$$

where $\widetilde{\mathbf{H}} = \mathbf{H}\mathbf{X}^f$ and K is the ensemble size. When H is non-linear, $\widetilde{\mathbf{H}}$ can be defined as:

$$\widetilde{\mathbf{H}}_k = H(\mathbf{x}_k^f) - \overline{H(\mathbf{x}^f)}$$

where k represents the k th ensemble member and $\overline{H(\mathbf{x}^f)}$ is the ensemble mean of the forecast ensemble in observation space. This substitution avoids the cost involved in finding the linearized observation operator required by $\mathbf{H}\mathbf{X}^f$.

We claim that

$$\widetilde{\mathbf{H}}^T [\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T + (K-1)\mathbf{R}]^{-1} = [\widetilde{\mathbf{H}}^T \mathbf{R}^{-1}\widetilde{\mathbf{H}} + (K-1)\mathbf{I}]^{-1} \widetilde{\mathbf{H}}^T \mathbf{R}^{-1}$$

This identity is easily verified by multiplying on the left by $\widetilde{\mathbf{H}}^T \mathbf{R}^{-1}\widetilde{\mathbf{H}} + (K-1)\mathbf{I}$ and on the right by $\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T + (K-1)\mathbf{R}$. Thus we have

$$\mathbf{K} = \mathbf{X}^f [\widetilde{\mathbf{H}}^T \mathbf{R}^{-1}\widetilde{\mathbf{H}} + (K-1)\mathbf{I}]^{-1} \widetilde{\mathbf{H}}^T \mathbf{R}^{-1} \tag{2.16}$$

The analysis error covariance in ensemble space is

$$\widetilde{\mathbf{P}}^a = [\widetilde{\mathbf{H}}^T \mathbf{R}^{-1}\widetilde{\mathbf{H}} + (K-1)\mathbf{I}]^{-1} \tag{2.17}$$

Thus

$$\mathbf{K} = \mathbf{X}^f \widetilde{\mathbf{P}}^a \widetilde{\mathbf{H}}^T \mathbf{R}^{-1} \tag{2.18}$$

(2.17) clearly shows that the matrix inverse is done within a K by K matrix where K is the ensemble size, a number much smaller than the observation size. $\mathbf{R}$ is typically diagonal or block diagonal.

With the Kalman gain obtained from (2.18), the LETKF updates the analysis only for the ensemble mean by

$$\overline{\mathbf{x}}^{a} = \overline{\mathbf{x}}^{f} + \mathbf{K}(\mathbf{y}^{o} - H(\overline{\mathbf{x}}^{f})) \tag{2.19}$$

In order to update the individual analysis ensemble member, the analysis ensemble perturbations are computed first then added to the analysis mean. Similar to the forecast ensemble perturbation matrix $\mathbf{X}^{f}$, the analysis ensemble perturbation matrix $\mathbf{X}^{a}$ is defined by $\mathbf{P}^{a} = \frac{1}{K-1}\mathbf{X}^{a}\mathbf{X}^{a^{T}}$, and can be computed by $\mathbf{X}^{a} = \mathbf{X}^{f}\mathbf{T}$ where $\mathbf{T}$ is a $n \times K$ transform matrix we will now determine.

Since $\mathbf{P}^{a} = (\mathbf{I} - \mathbf{KH})\mathbf{P}^{f}$, we have

$$\begin{aligned}
\mathbf{P}^{a} &= (\mathbf{I} - \mathbf{X}^{f}\widetilde{\mathbf{P}}^{a}\widetilde{\mathbf{H}}^{T}\mathbf{R}^{-1}\mathbf{H})\frac{1}{K-1}\mathbf{X}^{f}\mathbf{X}^{f^{T}} \\
&= \mathbf{X}^{f}(\mathbf{I} - \widetilde{\mathbf{P}}^{a}\widetilde{\mathbf{H}}^{T}\mathbf{R}^{-1}\widetilde{\mathbf{H}})\frac{1}{K-1}\mathbf{X}^{f^{T}} \\
&= \mathbf{X}^{f}\widetilde{\mathbf{P}}^{a}(\widetilde{\mathbf{P}}^{a^{-1}} - \widetilde{\mathbf{H}}^{T}\mathbf{R}^{-1}\widetilde{\mathbf{H}})\frac{1}{K-1}\mathbf{X}^{f^{T}} \\
&= \mathbf{X}^{f}\widetilde{\mathbf{P}}^{a}(K-1)\mathbf{I}\frac{1}{K-1}\mathbf{X}^{f^{T}} \\
&= \mathbf{X}^{f}\widetilde{\mathbf{P}}^{a}\mathbf{X}^{f^{T}}
\end{aligned} \tag{2.20}$$

On the other hand, $\mathbf{P}^{a} = \frac{1}{K-1}\mathbf{X}^{a}\mathbf{X}^{a^{T}}$, thus

$$\mathbf{X}^{a} = \mathbf{X}^{f}[(K-1)\widetilde{\mathbf{P}}^{a}]^{1/2} \tag{2.21}$$

The analysis ensemble perturbation matrix $\mathbf{X}^{a}$ can be obtained by transforming the forecast ensemble perturbation $\mathbf{X}^{f}$ through a transform matrix $\mathbf{T} = [(K-1)\widetilde{\mathbf{P}}^{a}]^{1/2}$

In practice, the implementation of the LETKF algorithm requires the following steps:

$$
\begin{cases}
\widetilde{\mathbf{H}}_k = H(\mathbf{x}_k^f) - \overline{H(\mathbf{x}^f)} \\
\widetilde{\mathbf{P}}^a = [\widetilde{\mathbf{H}}^T \mathbf{R}^{-1} \widetilde{\mathbf{H}} + (K-1)\mathbf{I}]^{-1} \\
\mathbf{K} = \mathbf{X}^f \widetilde{\mathbf{P}}^a \widetilde{\mathbf{H}}^T \mathbf{R}^{-1} \\
\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{K}(\mathbf{y}^o - H(\overline{\mathbf{x}}^f)) \\
\mathbf{X}^a = \mathbf{X}^f [(K-1)\widetilde{\mathbf{P}}^a]^{1/2} \\
\mathbf{x}^a = \overline{\mathbf{x}}^a + \mathbf{X}^a
\end{cases}
\tag{2.22}
$$

The algorithm (2.22) is applied to each local patch independently and the local analyses can be processed in parallel.

2.3 Methods to deal with model errors in the EnKF

In the previous section model errors are assumed to be Gaussian noise. However, in reality, they are usually biased. The overall model errors are the combination of model bias and random noise. In Chapter 1, we have introduced several methods for handling model errors. Here we focus on issues related to implementation of these methods.

2.3.1 Multiplicative inflation

Multiplicative inflation simply inflates the ensemble error covariance $\mathbf{P}_e^f$, by a factor 1+Δ to approximate the 'true' error covariance $\mathbf{P}^f$

$$
\mathbf{P}^f \leftarrow (1+\Delta)\mathbf{P}_e^f \tag{2.23}
$$

where Δ is a tunable parameter. (2.23) provides an increase in the ensemble covariance $\mathbf{P}_e^f$ to account for the model errors which are not included in the original $\mathbf{P}_e^f$. Implicitly this method assumes that model errors have the same error

structure as the internal errors so that their error covariance $\mathbf{Q}$ can be represented by dynamically evolved error covariance $\mathbf{P}_e^f$.

2.3.2 Additive inflation

Additive inflation parameterizes model errors by adding random perturbations with a certain covariance structure, to each ensemble forecast member after it has been propagated from the previous analysis ensemble using a dynamical model. In theory, additive inflation should attempt to select perturbations consistent in structure with the model errors. However little is known about the real model error covariance structure, the additional perturbations have been generated in several different ways. Houtekamer et al (2005) chose random perturbations consistent in structure with a 3D-Var background-error covariance; Hamill et al (2005) parameterized additive error as the differences between two forecasts with different resolution; Whitaker et al (2007) generated additive noise by randomly selecting samples from the 6-h tendencies field of NCEP/NCAR reanalysis (NNR) states. Those additive perturbations differ in the space structure and there is no conclusion so far on which of them is optimal. In this study, we chose the NNR 6-hour tendencies for the additive inflation scheme and tuned their amplitude.

The multiplicative inflation and additive inflation are two simple methods for accounting for the effect of model errors on the forecast error statistics, but without correcting the model errors from the ensemble mean. With a biased forecast Dee and da Silva (1998) have shown the resulting analysis is also biased even if the forecast error statistics are perfectly specified. In theory, if the systematic model errors can be

exactly estimated, directly removing them from the ensemble mean should result in a better analysis. Several schemes have been introduced to estimate and correct model errors.

2.3.3 Dee and da Silva bias estimation Method (DdSM)

Let us define the 'true' bias as

$$\mathbf{b}^t = < \mathbf{x}^f - \mathbf{x}^t > \tag{2.24}$$

and use $\mathbf{b}^o$, $\mathbf{b}^f$ to denote the 'observed bias' and 'forecasted bias', respectively, assuming we have a system to observe the bias and a model to forecast the bias.

Dee and da Silva (1998) developed a two-stage bias estimation algorithm, in which the estimation procedures for the bias and the state are carried out successively. At the first step of the analysis process, bias is estimated on every model grid point by assimilating the *forecast-minus-observation* residuals $H(\mathbf{x}^f) - \mathbf{y}^o$ which includes the model bias and errors due to imperfect initial conditions as well, as the observed bias $\mathbf{b}^o$ using Kalman filter equation

$$\begin{aligned} \mathbf{b}^a &= \mathbf{b}^f + \mathbf{K}_b[H(\mathbf{x}^f) - \mathbf{y}^o - H(\mathbf{b}^f)] \\ &= \mathbf{b}^f - \mathbf{K}_b[\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^f)] \end{aligned} \tag{2.25}$$

$$\mathbf{K}_b = \mathbf{P}_{bb}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}_{bb}^f \mathbf{H}^T + \mathbf{H}\mathbf{P}_{xx}^f \mathbf{H}^T + \mathbf{R})^{-1} \tag{2.26}$$

where the matrix $\mathbf{P}_{xx}^f$ is the forecast error covariance for the state variables, defined as in (2.1) and $\mathbf{P}_{bb}^f$ the forecast error covariance for bias, defined as:

$$\mathbf{P}_{bb}^f = < (\mathbf{b}^f - \mathbf{b}^t)(\mathbf{b}^f - \mathbf{b}^t)^T > \tag{2.27}$$

In practice the bias forecast error covariance $\mathbf{P}_{bb}^{f}$ is unknown, so that following DdS we assume that

$$\mathbf{P}_{bb}^{f} = \alpha \mathbf{P}_{xx}^{f} \tag{2.28}$$

Substituting (2.28) into (2.26), we have

$$\mathbf{K}_{b} = \alpha \mathbf{P}_{xx}^{f} \mathbf{H}^{T} [(1+\alpha)\mathbf{H}\mathbf{P}_{xx}^{f}\mathbf{H}^{T} + \mathbf{R}]^{-1} \tag{2.29}$$

The scalar α is a tunable parameter which controls the adaptability of the estimated bias. If α is too small the estimated bias evolves slowly and may take too long to spin-up or may not capture the time-dependent variability of the true bias. If α is too large then too much of the residuals $H(\mathbf{x}^{f}) - \mathbf{y}^{o}$ at a particular analysis time will be attributed to the bias and not enough to the internal errors.

At the second step, the analysis for the state variables is obtained using the standard analysis procedure with the unbiased forecast state $\mathbf{x}^{f} - \mathbf{b}^{a}$

$$\mathbf{x}^{a} = (\mathbf{x}^{f} - \mathbf{b}^{a}) - \mathbf{K}_{x}[\mathbf{y}^{o} - H(\mathbf{x}^{f} - \mathbf{b}^{a})] \tag{2.30}$$

$$\mathbf{K}_{x} = \mathbf{P}_{xx}^{f}\mathbf{H}^{T}(\mathbf{H}\mathbf{P}_{xx}^{f}\mathbf{H}^{T} + \mathbf{R})^{-1} \tag{2.31}$$

As for the bias forecast model, following Carton et al (2000) we will use a damped persistence,

$$\mathbf{b}_{i}^{f} = \mu \mathbf{b}_{i-1}^{a} \tag{2.32}$$

where μ <1 and we tune the scalar μ. (2.32) assumes there are no dynamics for the bias itself and that it is 'forgotten' with time.

The cost of the DdSM is about twice that of no bias estimation, since the updated equations are solved twice, first for the bias estimation and then for the state

variables. However, this double cost problem can be avoided if $\alpha << 1$ in which case (2.29) becomes

$$\mathbf{K}_b \approx \alpha \mathbf{P}_{xx}^f \mathbf{H}^T [(\mathbf{H}\mathbf{P}_{xx}^f \mathbf{H}^T + \mathbf{R}]^{-1} \tag{2.33}$$

$$= \alpha \mathbf{K}_x$$

Reversing the order of the bias estimation step and that of the state analysis step, we obtain a simplified version of Dee and da Silva scheme (Radakovich et al., 2001).

$$\mathbf{x}^a = (\mathbf{x}^f - \mathbf{b}^f) + \mathbf{K}_x [\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^f)] \tag{2.34}$$

$$\mathbf{b}^a = \mathbf{b}^f - \alpha \mathbf{K}_x [\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^f)] \tag{2.35}$$

In this approach the computational of (2.35) is almost cost free after the state analysis $\mathbf{x}^a$ has been updated by (2.34), since $\mathbf{K}_x[\mathbf{y}^o - H(\mathbf{x}^f - b^f)]$ is simply the analysis increment for the state variables.

In the application of the DdSM to the LETKF, no additional ensemble members are required for the bias since the bias forecast error covariance $\mathbf{P}_{bb}^f$ is obtained directly from the state forecast error covariance $\mathbf{P}_{xx}^f$. For the second analysis step in which the state variables are updated, because the analysis equations (2.30)-(2.31) for the state variables are exactly the same as those in the traditional Kalman filter except that the original forecast has been replaced by the bias-corrected forecast, we can directly utilize the algorithm (2.22). For the first step, in which the bias analysis is performed, the implementation is different. First we do not need calculate the analysis ensemble perturbations since there is no ensemble for the bias. Second, the formula of Kalman gain $\mathbf{K}_b$ for bias is slightly different from that for state variables. In order to

implement DdSM in the LETKF, we have to re-write (2.31) following a derivation similar to that of (2.16), so that the Kalman gain for bias is written as

$$\mathbf{K}_b = \alpha \mathbf{X}^f [(1+\alpha)\widetilde{\mathbf{H}}^T \mathbf{R}^{-1} \widetilde{\mathbf{H}} + (K-1)\mathbf{I}]^{-1} \widetilde{\mathbf{H}}^T \mathbf{R}^{-1} \tag{2.36}$$

where the variables in the right-hand side are defined the same as before in section 2.2.3.

2.3.4 Baek et al (2006) bias estimation scheme

Based on the *state space augmentation* method (e.g. Jazwinski 1970), Baek et al. (2006) propose a scheme for the EnKF to obtain the optimal estimates of the state and bias variables simultaneously. In this method both the state and bias have ensemble members and for each member the state vector is augmented with the uncertain model bias vector. The analysis updates the augmented vector

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{b} \end{pmatrix}$$

by assimilating observations of the state variables, where $\mathbf{x}$ is the model state and $\mathbf{b}$ model bias. Unlike in DdSM, there is no assumed $\mathbf{b}^o$ in this scheme. The bias is updated by state observations through the cross-correlation between the forecast state and bias, while this cross-correlation is assumed to be zero in DdSM. Therefore, the major difference between DdS and Baek et al. is that the former uses the bias error covariance P_{bb}^f which is assumed to be proportional to state error covariance, while the latter relies on the cross covariance between the bias and state variables. In addition, Baek et al. (2006) consider "correcting" the observation bias with respect to the model, rather than the model with respect to the observation in order to maintain the analysis within the model attraction.

This scheme is not tested in this study due to a parallel work by Baek who is implementing this scheme to the SPEEDY model.

2.3.5 Low-Dimensional Method (LDM)

Danforth et al (2007) separated the 6-hour forecast model errors into a constant forecast bias, the periodic (diurnal) component of the bias, and the state-dependent model errors. If a reference state $\mathbf{x}^r$ is available to approximate the atmospheric truth (in our case we use the NCEP/NCAR reanalysis), the 6-hour forecast error can be defined in terms of the reference state as

$$\mathbf{x}^e = \mathbf{x}^f - \mathbf{x}^r$$

The low-dimensional scheme allows for three types of model errors,

$$\mathbf{b} + \sum_{l=1}^{L} \beta_l(t)\mathbf{e}_l + \sum_{m=1}^{M} \gamma_m(t)\mathbf{f}_m \tag{2.37}$$

in which t denotes the time step. The forecast bias $\mathbf{b}$ is obtained by averaging the errors over a certain time period $\mathbf{b} = < \mathbf{x}^e >$ and the leading EOFs $\mathbf{e}_l$ from the anomalous error field $\mathbf{x}^{e\prime} = \mathbf{x}^e - < \mathbf{x}^e >$ which is not represented in the constant bias, are used to estimate diurnal or other periodic errors. The state-dependent systematic component is given by the leading Singular Vector Decomposition (SVD) modes (SVDs) $\mathbf{f}_m$ of the covariance of the coupled model state anomalies $\mathbf{x}^{f\prime} = \mathbf{x}^f - < \mathbf{x}^f >$ and corresponding error anomalies $\mathbf{x}^{e\prime}$. L and M are the number of leading modes of EOFs and SVDs, respectively. The spatial fields $\mathbf{b}$, $\mathbf{e}_l$ and $\mathbf{f}_m$ are time-independent and pre-computed using the samples in the training period. We call this approach low-dimensional because the shape of the model errors is pre-estimated

separately whereas only the amplitudes (Principal Components, PCs) $\beta_l(t)$ and $\gamma_m(t)$ which have a much lower dimension (L and M) than the full model dimension, are estimated online.

During the training period, the time-series of β_l is calculated by projecting the EOFs $\mathbf{e}_l$ onto $\mathbf{x}^e{}'$. Since β_l is dominated by bias in the diurnal cycle, the time-dependent $\beta_l(t)$ can be estimated by averaging the β_l over the diurnal cycle in the training period. For example, we can use the samples in the training period to calculate the average β_l for 00z, 06z, 12z and 18z separately and apply them in the current time-step.

Rather than calculating $\mathbf{x}^e{}'(t) = \sum_{m=1}^{M} \gamma_m(t)\mathbf{f}_m$ Danforth et al. obtained the error anomalies $\mathbf{x}^e{}'(t)$ at the current time step t based on the statistical correlation of the SVD mode m and current state anomalies $\mathbf{x}^f{}'(t)$. Denote $s_m(n)$, $\mathbf{g}_m$ and $\gamma_m(n)$ and $\mathbf{f}_m$ as the PCs and SVD mode for the state anomalies $\mathbf{x}^f{}'(n)$ and the 6-hour error anomalies $\mathbf{x}^e{}'(n)$ at time step n in the training period, the PCs $s_m(n)$ and $\gamma_m(n)$ are obtained by projecting the coupled signals $\mathbf{g}_m, \mathbf{f}_m$ onto the samples $\mathbf{x}^f{}'(n), \mathbf{x}^e{}'(n)$ as follows

$$s_m(n) = \mathbf{x}^f{}'(n) \cdot \mathbf{g}_m \tag{2.38}$$

$$\gamma_m(n) = \mathbf{x}^e{}'(n) \cdot \mathbf{f}_m \tag{2.39}$$

The correlation coefficient between the error anomalous field $\mathbf{x}^e{}'(n)$ and the m th PC $s_m(n)$ is given by

$$\rho[\mathbf{x}^e{}'(n), s_m(n)] = \frac{\lambda_m \mathbf{f}_m}{\sqrt{< \mathbf{x}_m^{e\,'2}(n) >}\sqrt{< s_m^2(n) >}} \tag{2.40}$$

where λ_m is the m th singular values. Based on the linear-regression theory, $\mathbf{x}_m^e{}'(t)$ at the current time step t can be computed by

$$\mathbf{x}_m^e{}'(t) = \rho[\mathbf{x}^e{}'(n), s_m(n)] \cdot \frac{\sqrt{< \mathbf{x}_m^{e\,'2}(n) >}}{\sqrt{< s_m^2(n) >}} \cdot s_m(t)$$

$$= \rho[\mathbf{x}^e{}'(n), s_m(n)] \cdot \frac{\sqrt{< \mathbf{x}_m^{e\,'2}(n) >}}{\sqrt{< s_m^2(n) >}} \cdot \mathbf{x}^f{}'(t) \cdot \mathbf{g}_m \tag{2.41}$$

where we have utilized the current state anomalies $\mathbf{x}^f{}'(t)$ and the correlation between the state anomalies and the error anomalies to obtain the state-dependent errors. By plugging (2.40) into (2.41), (2.41) can be simplified to

$$\mathbf{x}_m^e{}'(t) = \frac{\lambda_m \mathbf{f}_m}{< s_m^2(n) >} \cdot \mathbf{x}^f{}'(t) \cdot \mathbf{g}_m \tag{2.42}$$

Now what remains to be determined is how to get the samples of 6-hour forecast error $\mathbf{x}^e$. In this study, following Danforth et al.(2007), in the training period we run the SPEEDY initialized with the NCEP/NCAR reanalysis (NNR) fields and take the differences between the SPEEDY 6-hour forecasts and NNR fields valid at the same time to obtain the samples of $\mathbf{x}^e$.

2.4 Adaptive estimation of inflation factor and observation errors

2.4.1 Adaptive estimation of inflation parameter

Denoting $\mathbf{d}_{o-b} = \mathbf{y}^o - H(\mathbf{x}^b)$ as the innovations, i.e. the difference between the observations and their background counterparts $H(\mathbf{x}^b)$. Thus

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

Here $\boldsymbol{\varepsilon}^o = \mathbf{y}^o - H(\mathbf{x}^t)$ and $\boldsymbol{\varepsilon}^b = \mathbf{x}^b - \mathbf{x}^t$ are the observation and background error. Then the covariance of the innovation is

$$< \mathbf{d}_{o-b}(\mathbf{d}_{o-b})^T >=< \boldsymbol{\varepsilon}^o \boldsymbol{\varepsilon}^{oT} > + \mathbf{H} < \boldsymbol{\varepsilon}^b \boldsymbol{\varepsilon}^{bT} > \mathbf{H}^T$$

where the observational errors $\boldsymbol{\varepsilon}^o$ are assumed to be uncorrelated with the background errors $\boldsymbol{\varepsilon}^b$. Using the definitions of background error covariance (2.1) and observation error covariance (2.2), we have

$$< \mathbf{d}_{o-b}(\mathbf{d}_{o-b})^T >= \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R} \tag{2.43}$$

This is a classical and widely-used relationship that provides a global check on the specification of the sum of background error and observation error covariances in observational space. If R is perfectly known, (2.43) can be used to diagnose the accuracy of the specified background error covariance.

Another diagnostic on background errors can be obtained by the combination of innovation $\mathbf{d}_{o-b}$ and *analysis-minus-background* $\mathbf{d}_{a-b}$ (Desroziers et al. 2005). Since $\mathbf{d}_{a-b}$ can be written

$$\mathbf{d}_{a-b} = H(\mathbf{x}^a) - H(\mathbf{x}^b) \approx \mathbf{H}(\mathbf{x}^a - \mathbf{x}^b) = \mathbf{H}\mathbf{K}\mathbf{d}_{o-b}$$

Therefore we have

$$< \mathbf{d}_{a-b}(\mathbf{d}_{o-b})^T >= \mathbf{H}\mathbf{K} < \mathbf{d}_{o-b}(\mathbf{d}_{o-b})^T >$$

Matrix $\mathbf{H}\mathbf{K}$ is given by $\mathbf{H}\mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}$, thus

$$< \mathbf{d}_{a-b}(\mathbf{d}_{o-b})^T >= \mathbf{H}\mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} < \mathbf{d}_{o-b}(\mathbf{d}_{o-b})^T >$$

Combining with (2.43), this whole expression simplifies to

$$< \mathbf{d}_{a-b}(\mathbf{d}_{o-b})^T >= \mathbf{H}\mathbf{P}^f\mathbf{H}^T \tag{2.44}$$

if the matrices specified in $\mathbf{H}\mathbf{K} = \mathbf{H}\mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}$ agree with the true covariances for background and observation error,

(2.43) and (2.44) provide two consistency checks on background error covariance in observation space. In the ensemble filter, because of a variety of errors, the ensemble error covariance $\mathbf{P}_e^f$ estimated from the ensemble of forecast fields tends to underestimate the true background error covariance and multiplicative inflation is an attempt to rectify this, with

$$\mathbf{P}^f = (1+\Delta)\mathbf{P}_e^f \tag{2.45}$$

Plugging (2.45) into (2.43) or (2.44) and considering only the diagonal term, we can estimate on-line the inflation factor Δ by

$$\widetilde{\Delta} = \frac{\mathbf{d}_{o-b}^T\mathbf{d}_{o-b} - Tr(\mathbf{R})}{Tr(\mathbf{H}\mathbf{P}_e^f\mathbf{H})} - 1 \tag{2.46}$$

$$\widetilde{\Delta} = \frac{\mathbf{d}_{a-b}^T\mathbf{d}_{o-b}}{Tr(\mathbf{H}\mathbf{P}_e^f\mathbf{H})} - 1 \tag{2.47}$$

where *Tr* denotes the trace of a matrix.

We denote (2.46) and (2.47) as OMB^2 method and *AMB*OMB* method, respectively. An accurate estimate of Δ from these two methods requires a correct observation error covariance. This is obvious for (2.46) but also implicitly true for (2.47) where $\mathbf{d}_{a-b}$ itself is based on the use of (generally incorrect) specified observation error variances.

2.4.2 Adaptive estimation of observation errors

The $\mathbf{d}_{o-a}$ *observation-minus-analysis* differences are given by

$$\begin{aligned}\mathbf{d}_{o-a} &= \mathbf{y}^o - H(\mathbf{x}^a) \\ &= \mathbf{y}^o - H[\mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - H\mathbf{x}^b)] \\ &\approx \mathbf{y}^o - H(\mathbf{x}^b) + \mathbf{HK}(\mathbf{y}^o - H\mathbf{x}^b) \\ &= (\mathbf{I} - \mathbf{HK})\mathbf{d}_{o-b} \\ &= \mathbf{R}(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d}_{o-b}\end{aligned}$$

Then the statistical expectation of the cross-product between $\mathbf{d}_{o-a}$ and $\mathbf{d}_{o-b}$ is

$$< \mathbf{d}_{o-a}(\mathbf{d}_{o-b})^T >= \mathbf{R}(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1} < \mathbf{d}_{o-b}(\mathbf{d}_{o-b})^T > \tag{2.48}$$

Plugging (2.43) into (2.48), we have

$$< \mathbf{d}_{o-a}(\mathbf{d}_{o-b})^T >= \mathbf{R} \tag{2.49}$$

if the matrices specified in $\mathbf{HK} = \mathbf{HP}^f\mathbf{H}^T(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1}$ agree with the true covariances for background and observation error. This is the diagnostic providing a consistency check on observation error covariance. One application of this diagnostic is to diagnose observation error variance offline or on-line estimate it. For any subset of observations i with p_i observations, it is possible to compute the variance

$$(\tilde{\sigma}_o)_i^2 = (\mathbf{d}_{o-a})^T{}_i(\mathbf{d}_{o-b})_i / p_i = \sum_{j=1}^{p_i}(y_j^o - y_j^a)(y_j^o - y_j^b)/p_i \tag{2.50}$$

where y_j^o is the value of observation j and y_j^a, y_j^b are their analysis and background counterparts.

We denote (2.50) as *OMA*OMB* method. The accuracy of this method relies on $\mathbf{d}_{o-a}$ and $\mathbf{d}_{o-b}$ which depend themselves on the observation and background (therefore on inflation factor in EnKF) errors variances.

2.4.3 Simultaneous estimation of inflation and observation errors

As discussed in the previous two sections, adaptive estimation of inflation requires the perfect observation error variance $(\sigma_o)^2$ while an accurate estimate of $(\sigma_o)^2$ relies on the optimal inflation factor. This becomes a nonlinear problem when neither the optimal inflation factor nor true $(\sigma_o)^2$ is known beforehand and both of them need to be estimated on-line. In this case, we propose to estimate the inflation and observation errors simultaneously within analysis cycle and allow the system itself to take some time to converge to the optimal value (range) for observation error variance (inflation factor).

2.4.4 Smoothing

We estimate the observation error variances and inflation parameter adaptively at each analysis time step. However, the number of samples available at each step may not be enough to avoid large sampling error. To increase the sampling, we use adaptive regression based on a simple scalar KF approach usually used to post-process model output (e.g. Kalnay 2003, Appendix C) to accumulate past information and make the observation error variance and inflation gradually converge to the optimal value while still allowing for time variations. This approach can be thought as a time smoother and has been used by Miyoshi 2005. We regard the estimation obtained directly from OMB^2 or *AMB*OMB* or *OMA*OMB* as an observed estimate

α^o (either $\widetilde{\Delta}^o$ or $(\widetilde{\sigma}_o^o)^2$ in this study) for the current time step. Instead of directly using it as the final estimation for that time step, we use simple scalar KF approach to best combine α^o and α^f, the value derived by persistence from the previous time step, to get a new estimate denoted as the analysis α^a:

$$\alpha^a = \frac{v^o \alpha^f + v^f \alpha^o}{v^o + v^f} \tag{2.51}$$

where v^f (v^o) denotes the forecast (observational) error variance for the adaptive regression. The relative ratio of v^o / v^f determines the degree of smoothing. The bigger this ratio, the more α^o obtained from current step will be smoothed. The error variance of α^a is given by

$$v^a = (1 - \frac{v^f}{v^f + v^o}) v^f \tag{2.52}$$

Assuming persistence as the forecast model for the estimated variable, and allowing for some error in the "persistence forecast" (Kalnay, 2003, Appendix C), we have:

$$\alpha^f{}_{t+1} = \alpha^a{}_t \tag{2.53}$$

$$v^f{}_{t+1} = \kappa v^a{}_t \tag{2.54}$$

here κ is a parameter which allows the slow increase of the forecast error. Although two additional control parameters: the observation error variance v^o and error growth parameter κ have been introduced here, Miyoshi 2005 has shown the final estimate is not sensitive to either of them. Following Miyoshi 2005, we use v^o=1.0 and κ=1.03 in this study.

Chapter 3

The LETKF performance in perfect model experiments

3.1 Introduction

In this Chapter, we implement the LETKF into an AGCM model, the SPEEDY model, under the assumption of a perfect model. Issues related to sampling errors are addressed. The performance of the LETKF is examined by checking the characteristics of the analysis, background errors and their ensemble spread. Those characteristics will be served as a reference to the experiments in next chapter where model is no longer perfect.

3.2 The SPEEDY model

The SPEEDY (Simplified Parameterizations, primitivE-Equation DYnamics) model (Molteni 2003) is a recently developed atmospheric general circulation model (AGCM) with simplified physical parameterizations that are computationally efficient, but that maintain the basic characteristics of a state-of-the-art AGCM with complex physics.

The SPEEDY model solves the primitive equation for prognostic variables of zonal wind (u), meridional wind (v), temperature (T), specific humidity (q), and surface pressure (p_s) at the truncation of wavenumber 30, corresponding to 96*48 grid points and 7 sigma levels (0.950, 0.835, 0.685, 0.510, 0.340, 0.200, 0.080). The inputs are only taken in sigma levels, but the resulting forecasts are in both sigma levels and pressure levels (925, 850, 700, 500, 300, 200, 100 hPa).

3.3 Experimental setup

We applied the LETKF to the SPEEDY model in the perfect model scenario where a true "nature run" is generated by integrating the SPEEDY model from 0000 UTC 1 January, 1987, using the NCEP-NCAR Reanalysis (NNR) as the initial condition, until 1800 UTC 15 February, 1987. The observations are simulated by adding normally-distributed random noise to the nature run, and are available at each model grid points for p_s and every other model grid for u, v, T, q, in both zonal and meridional directions, i.e. 25% of the number of model grid points. The amplitudes of the observation errors are 1 m/s for u, v wind, 1 K for T, 10^{-4} kg/kg for q and 1hPa for p_s. For each forecast-analysis cycle, the SPEEDY model is used to generate the 6-hour forecast and the observations are assimilated by the LETKF.

We start the experiment by running 30 initial ensembles at 0000 UTC 1 January, 1987. The 30 initial ensembles are created by adding the random noise to the "nature run" forecasts at 0600 UTC 1 January, 1987. In this way, the initial ensemble mean is 6 hours apart from the truth. Though the final results after the spin-up period are not sensitive to the initial ensembles, a good initial condition (not too far from the truth) helps to shorten the spin-up time. With our setting of the initial conditions, the spin-up time is about 15 days while it is more than one month in the experiments in Miyoshi 2005. The experiment runs for one and a half month till 1800 UTC 15 February, 2007 and only the last month after the spin-up period is reported to verify the system.

3.4 Sampling errors

In any ensemble data assimilation system, a limited K ensemble members are used to reproduce the N*N forecast error covariance matrix $\mathbf{P}^f$, where N is the number of degrees of freedom of the model. To avoid the filter divergence due to small samples, some additional processing of sample covariance $\mathbf{P}^f$ is usually necessary. The two common methods are multiplicative inflation (Anderson and Anderson 1999) and localization (Hamill et al., 2001; Houtekamer and Mitchell, 2001) techniques. Multiplicative inflation simply inflates the forecast error covariance $\mathbf{P}^f$ by a factor Δ slightly larger than 1, which is equivalent to inflating the ensemble spread by a factor $\delta = \sqrt{1+\Delta} - 1$.

Localization algorithms are meant to correct for errors in the sample covariance between observations and model state variables. In the LETKF, the analysis is performed in a local patch centered by a model grid point for which we try to find the optimal analysis. Only the observations within the local box are used to update the central point. This can be regarded as a cutoff-based localization algorithm. In our experiments, the local patch is simply a two-dimensional square box with length $2l+1$ in both x and y model grid space. The physical shape of the local patch is not a square due to the convergence of the meridians in higher latitudes. The cutoff localization scheme forces the correlations between the analysis point and observations outside the local patch to be zero but without distinguishing the observations inside the local patch. This scheme may not be accurate enough since in our experiments we have only 30 ensemble members, so that the correlations inside the local patch could have sampling errors, especially for those observations that are

on the border of the local patch. In this study, besides the cut-off location, we apply the observation error covariance localization (Miyoshi 2005) to reduce or remove these spurious correlations.

$$\mathbf{R} \leftarrow \rho^{-1} \circ \mathbf{R} \tag{3.1}$$

$$\rho = \exp(\frac{-d^2}{2\sigma^2}) \tag{3.2}$$

where d denotes the distance from the analysis point and observation location and σ the localization scale. Therefore we have,

$$\begin{aligned} \mathbf{K} &= \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \rho^{-1} \circ \mathbf{R})^{-1} \\ &\cong \rho \circ \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\rho \circ \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \end{aligned} \tag{3.3}$$

Equation (3.3) indicates that the observation error covariance localization is an approximation of the localization of $\mathbf{P}^f$ (Houtekamer and Mitchell, 2001). Figure 3.1 shows the localization factor ρ as the function of the ratio d/σ, indicating the correlation is reduced to 60% of its original size for observation at the distance d=σ, while the observation effect is forced to be near zero at the distance larger than 3.5σ.

Both the inflation factor and localization scale need to be tuned. To simplify the task of tuning we first fixed the inflation factor to be 0.04. Table 3.1 compares the analysis RMSE of 500hPa height using cut-off location and observation error covariance location with 30 ensemble members. With cut-off localization, the best result is obtained with the smallest local patch. The use of observation error localization allows the LETKF to use larger patches with more observations. Actually, when the localization scale σ is fixed, the results are not sensitive to the local patch size l. The smallest analysis error 2.33 in Table 3.1 is observed with a

local patch size l=3 and an observational error covariance localization scale σ=1.5. Using this setting of localization scales, we then tuned the inflation factor and found that Δ=0.05 is slightly better than Δ=0.04.

Table 3.1: Analysis RMSE of 500 hPa height field using LETKF with 30 ensemble members in the cases of applying cut-off localization or observation error covariance localization. The RMSE is temporally averaged for a month after the initial 15-days spin-up period and spatially averaged over the globe. A multiplicative inflation factor of Δ=0.04 is applied. Two parameters, the local patch size l and the observational error covariance scale σ, are tuned.

The local patch size l	1	2	3
500Z RMSE (m) (σ=∞)	2.34	2.66	3.19
500Z RMSE (m) (σ=2.0)			2.41
500Z RMSE (m) (σ=1.5)		2.35	2.33

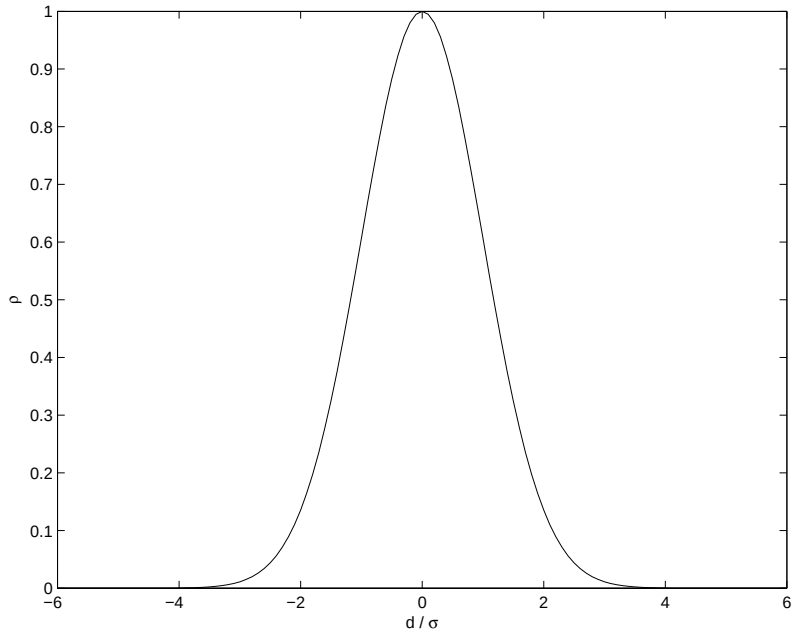

Figure 3.1: The localization factor $\rho = \exp(\frac{-d^2}{2\sigma^2})$ as the function of ration d/σ

3.5 LETKF performance in a perfect model experiment

Figure 3.2 summarizes the assimilation results with our final setting of the configurations (l=3, σ=1.5, Δ=0.05 and 30 ensemble members) for the LETKF under the perfect model assumption. Shown are time-series of root-mean-square (rms) analysis error (defined as analysis minus true state), averaged over the whole globe, for zonal wind (u), geopotential height (Z), temperature (T) and specific humidity (q). By assimilating the observations every 6-hour, the analysis error reduces rapidly from a large initial value and approaches a stable level. It is clear that after the initial spin-up period, the analysis RMSE for all the variables are much smaller than the observational error standard deviations.

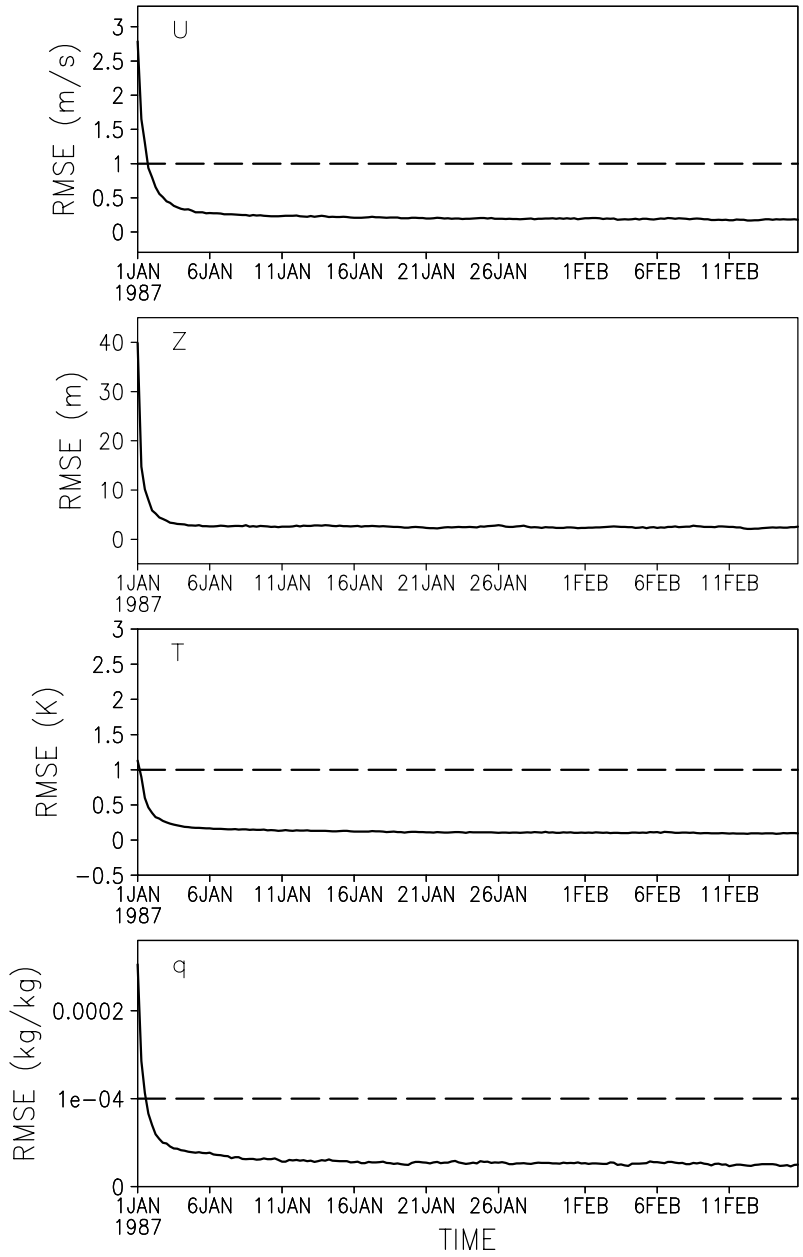

Figure 3.2: Analysis RMSE (solid curve) at 500 hPa for the period between 0000 UTC 1 January 1987 and 1800 UTC 15 February 1987. The observational error standard deviations are shown as dash lines wherever applicable. The four panels from the top to the bottom correspond to zonal wind, geopotential height, temperature and specific humidity, respectively.

To find why the LETKF performs well, we compare the ensemble spread with the ensemble mean error. Carrying out perfect model experiments, where we know the truth, allows us to calculate the ensemble mean error, which is the ensemble mean minus the truth. The ensemble spread is defined as the 6-hour forecast error deviation with respect to the ensemble mean. In the ensemble filter, the background error covariance is approximated using the sample covariance from an ensemble of model forecasts. To obtain a good performance, we require that the ensemble spread be representative of the true background error in both structure and magnitude. Figure 3.3 shows the background error field (shaded) and the ensemble spread (contour) at an arbitrary time. In general these two fields agree with each other very well, indicating the ensemble spread has captured the forecast error structure.

As for the amplitude, we compare the background RMSE and the spread averaged over the whole globe (Figure 3.4). The value of spread is a little smaller in the lower level and higher in the upper levels in terms of the background RMSE. But in general, they are close enough to each other at all levels. Therefore we can conclude that with our setting of configurations (l=3, σ =1.5, Δ =0.05 and 30 ensemble members), the spread among 30 ensemble members has captured well both the true error structure and the true error magnitude.

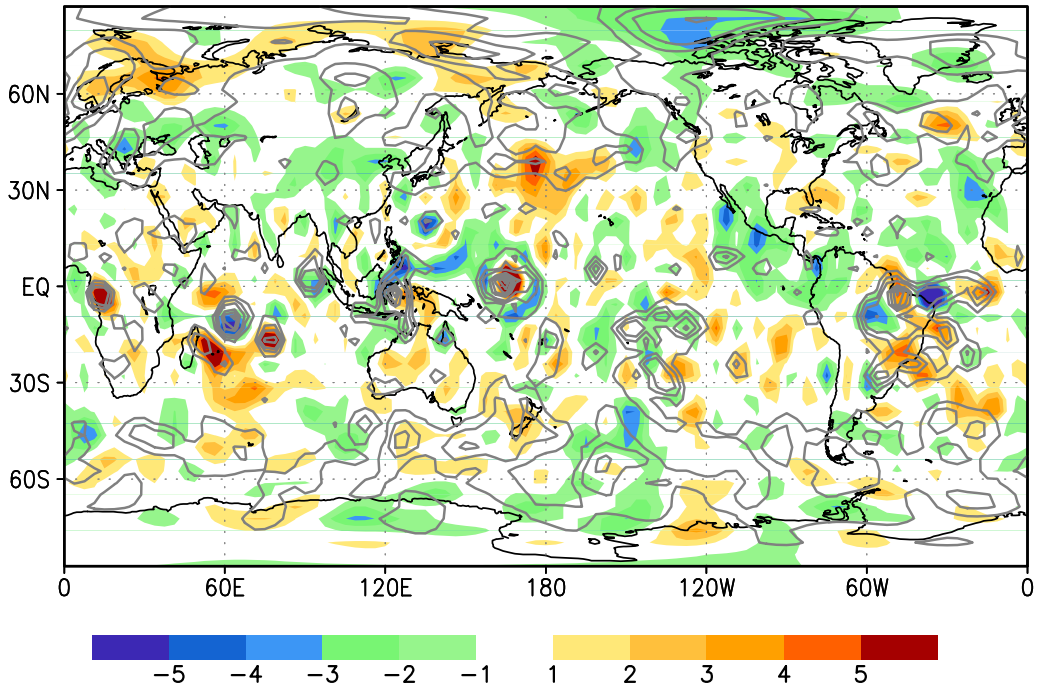

Figure 3.3: The background (6-hour forecast) error field (shaded) and the ensemble spread of 500 hPa height field (contour) at an arbitrary time.

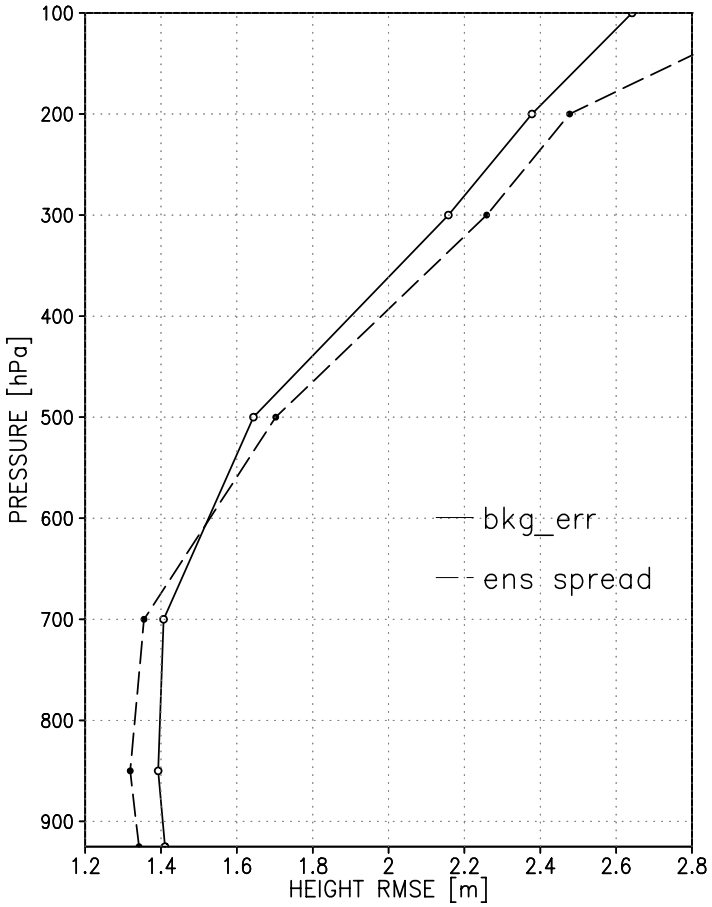

Figure 3.4: Background RMSE at all pressure levels (solid line) and background ensemble spread (dashed line) of height field, temporally averaged for one-month after the spin-up period.

Chapter 4

Accounting for and correcting model errors in the LETKF

4.1 Introduction

In Chapter 3, we have shown the LETKF works very well in the perfect model scenario. In reality, however, the numerical model is never perfect. In this chapter we drop the assumption of a perfect model by assimilating observations simulated from the NCEP/NCAR Reanalysis (NNR) as in Miyoshi (2005). Several methods to account for and/or correct model errors are tested.

4.2 NCEP/NCAR reanalysis (NNR)

4.2.1 NNR observations

To assess the performance of the LETKF in the presence of model errors, we replace the 'nature' run in the perfect model experiments by the NNR fields. Since the NNR assimilated real observations, we assume the NNR fields are an approximate estimate of the unknown 'true atmosphere'. A quantitative validation of this assumption is beyond the scope of this research. We add the random noise with the same standard deviation used in the perfect model experiments to simulate the NNR observations. The density of observations remains the same as that in Chapter 3. Since the NNR data are on pressure levels and we perform the LETKF in SPEEDY sigma coordinates, for simplicity, the NNR observations are converted to the SPEEDY sigma coordinates determined by the SPEEDY forecasted p_s. Horizontal and vertical linear interpolations are required to obtain the interpolated observations

in SPEEDY sigma coordinates. $\log p$ is used to determine the weighting for the vertical interpolation.

4.2.2 Characteristics of model errors

Since the observations are generated from the NNR while the SPEEDY has its own climatology, the model is no longer perfect. To investigate the SPEEDY model errors against the NNR field, the SPEEDY model is run every 6 hours using NNR as the initial conditions and the 6-hour forecast errors are calculated by the differences between the SPEEDY 6-hour forecasts and the NNR verified at the same time. Assuming the NNR is an estimation of the unknown 'truth', these 6-hour forecast errors are an approximation of the 'true' model errors.

Figure 4.1 shows the SPEEDY 6-hour forecast errors against the NNR field of the three prognostic variables (zonal wind, temperature and specific humidity) and one diagnostic variable (geopotential height), temporally averaged over our experimental period between 0000 UTC 1 January, 1987 and 1800 UTC 15 February, 1987 using 184 samples. The time-mean of forecast errors is regarded as the model bias. For zonal wind, the largest bias is found in the polar regions or associated with the position of the jet. For temperature, the bias is significant at lower or upper levels (mostly over land) but small at the 500 hPa level. The bias of specific humidity is large in the region where the value of the humidity itself is large, i.e., in the tropics and at low levels. As for the geopotential height, the largest bias can be seen over elevated surface terrain, which is due to the topographic differences in the SPEEDY mode and the NNR field.

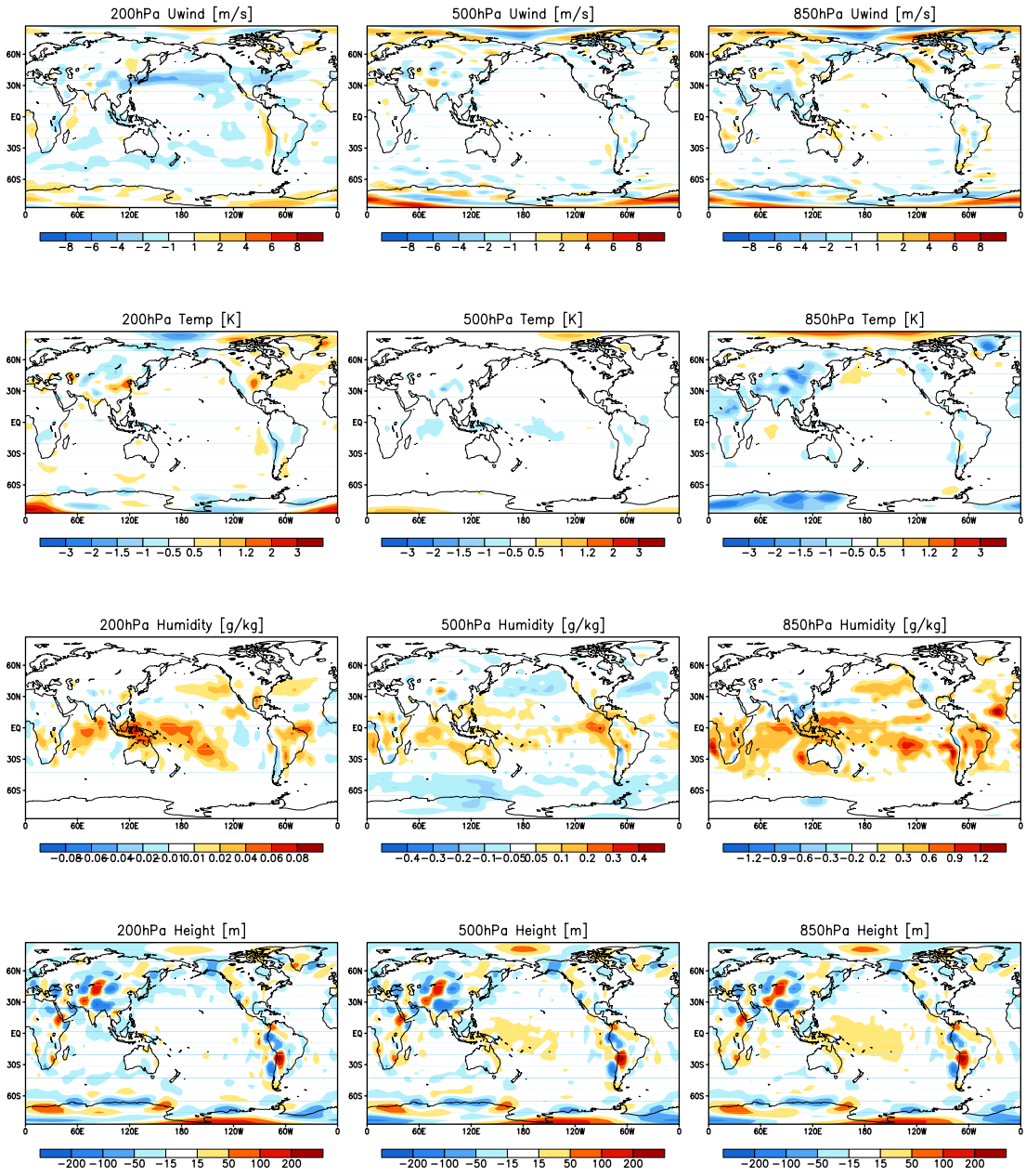

Figure 4.1: Mean of 6-hour SPEEDY forecast errors initialized from the NNR fields (model bias), temporally averaged over the experimental period between 0000 UTC 1 January, 1987 and 1800 UTC 15 February, 1987 using 184 samples.

Besides the time-averaged bias, we also investigate the time-varying component of model error to see if there is another significant signal which is not included in the time-averaged bias (Danforth et al. 2007). To do so, we subtract the time-averaged bias from the time series of model errors, and compute the empirical orthogonal function (EOFs) for the model error anomalous fields. Figure 4.2 shows the first two EOF patterns and their corresponding time amplitudes, Principal Components (PCs) for temperature at the bottom level (925 hPa). It is clear that both of them correspond to a diurnal variation of SPEEDY model bias. Combining the EOFs and PCs, we found the SPEEDY model underestimates (overestimates) the near-surface local daytime (nighttime) temperatures. This is due to the lack of diurnal changes in solar forcing in the SPEEDY model. This detected temperature diurnal bias is mainly over land and largest near the surface (925 hPa) but also significant at 850 hPa. We did not observe a significant diurnal bias signal in other levels and other variables.

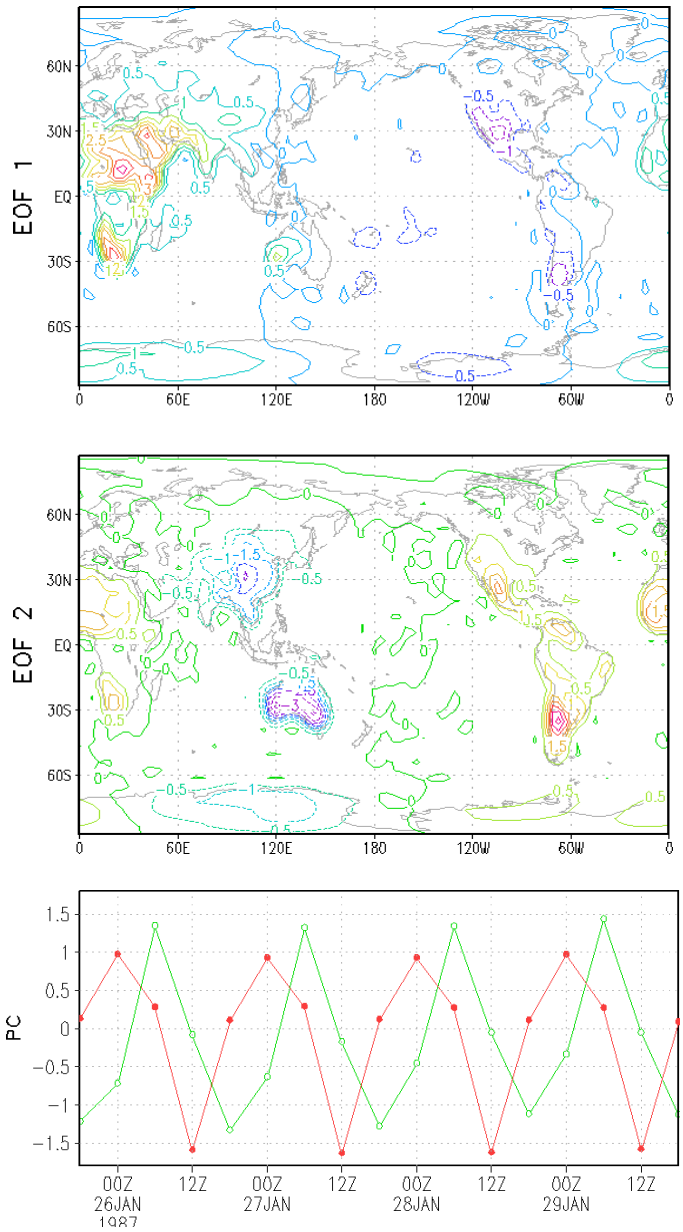

Figure 4.2: Top and middle panels: the first two leading EOF patterns of the 925 hPa temperature model error anomalous fields using samples in the experimental period between 0000 UTC 1 January, 1987 and 1800 UTC 15 February 1987. Their corresponding Principal Components (red line for PC1 and green line for PC2) are shown for an arbitrary period in January of 1987. The explained variances for these two EOF modes are 35.7% and 32.5% respectively.

4.3 Effects of model errors on the LETKF

4.3.1 Experimental setup

To assess the effect of model errors on the performance of the LETKF, we assimilate the NNR observations every 6-hour with the LETKF using the same configurations as those in the perfect model experiments (l=3, σ=1.5, Δ=0.05 and 30 ensemble members). Assimilations were performed for the period between 0000 UTC 1 January 1987 and 1800 UTC 15 February, 1987. The initial ensembles at 0000 UTC 1 January were generated by adding random noise to the NNR field at 0000 UTC 2 January. No extra method is applied to deal with model errors. This experiment is regarded as the '**control run**' for the later experiments in section 4.4. After an initial spin-up period of half-month, the analyses and forecasts are verified against the NNR field which is assumed to be an estimation of the unknown truth.

4.3.2 'Control run' results

Figure 4.3 shows the analysis and background RMSE at all pressure levels of the geopotential height field. For comparison, the analysis RMSE of height in the case of perfect model is also shown. We can see the strong negative influence of the model errors on the performance of the LETKF. In the presence of models errors, the 500 hPa height analysis RMSE has increased from 2.4 m to 50 m due to the model errors and their accumulated effects. With a more sophisticated and high-resolution numerical model, such as those currently used in operation, the negative influence should be much smaller but will still exist as long as the model is not perfect.

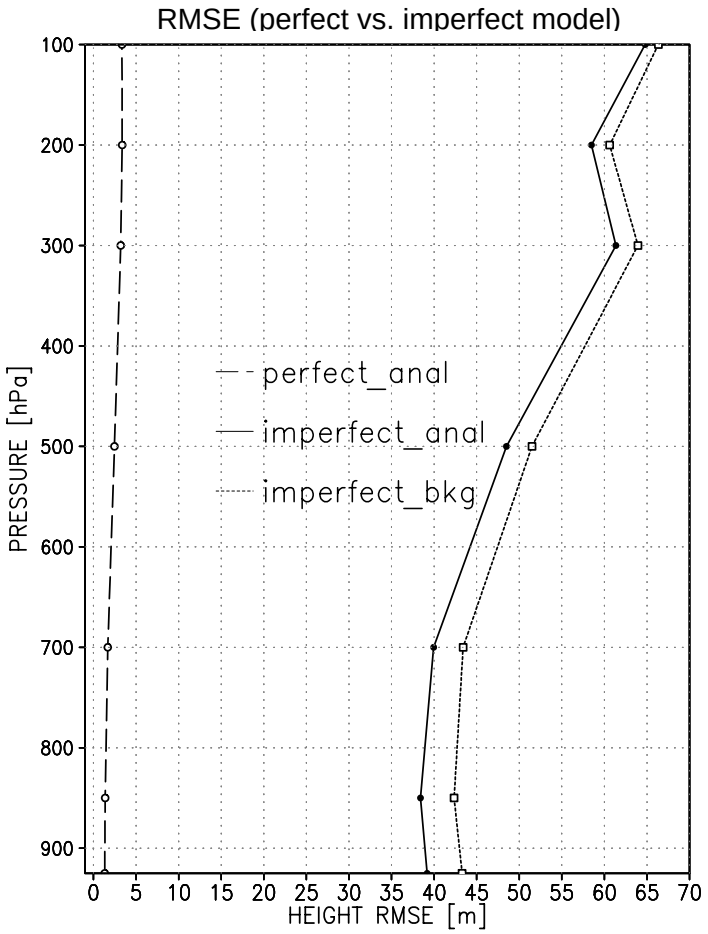

Figure 4.3: Analysis RMSE (solid curve) and background RMSE (dotted curve) at all pressure levels of height field when assimilating the 'realistic' observations from NNR, temporally averaged for a month after the initial spin-up period. For comparison, the analysis RMSE in the perfect model experiment where assimilating the 'simulated' observations from SPEEDY 'nature run' is also shown (dashed curve). (We note that in a more realistic model, with a larger inflation, the imperfect model analysis does not deteriorate as much as in this example, e.g., Figure 2 in Szunyogh et al. 2007)

To investigate why the LETKF performs poorly in the presence of large model errors without additional inflation, we plot the background ensemble spread of the height field (Figure 4.4). It is interesting to find that the spread is similar to that in the perfect model experiment, and much smaller than the actual forecast error shown by the dotted line in Figure 4.3, indicating that the forecast ensembles are 'blind' to model errors. This can be clearly explained by the schematic in Figure 4.5 where the difference between ensembles is small but the whole set of ensembles is far from the 'truth'. In this case, the ensemble spread underestimates the actual forecast error, leads to the wrong confidence to the forecasts and less weight to the observations and, as a result, to large analysis errors as in the 'control run' (Figure 4.3).

4.4 Accounting for and correcting model errors

4.4.1 Experimental design

We have seen the bad performance of the LETKF in the presence of model errors without correcting or accounting for their effects. In this section we test several methods described in Chapter 2 to deal with model errors. As in the 'control run', we run the SPEEDY-LETKF for the period 0000 UTC 1 January, 1987 to 1800 UTC 15 February, 1987 and the verification statistics are computed for analyses and forecasts against the NNR fields after the initial spin-up period of half-month.

4.4.2 Multiplicative inflation

Multiplicative inflation is a method to deal with model errors that is straightforward to implement since it has already been used in the perfect model experiments. Rather than being used to prevent the ensemble from collapsing due to

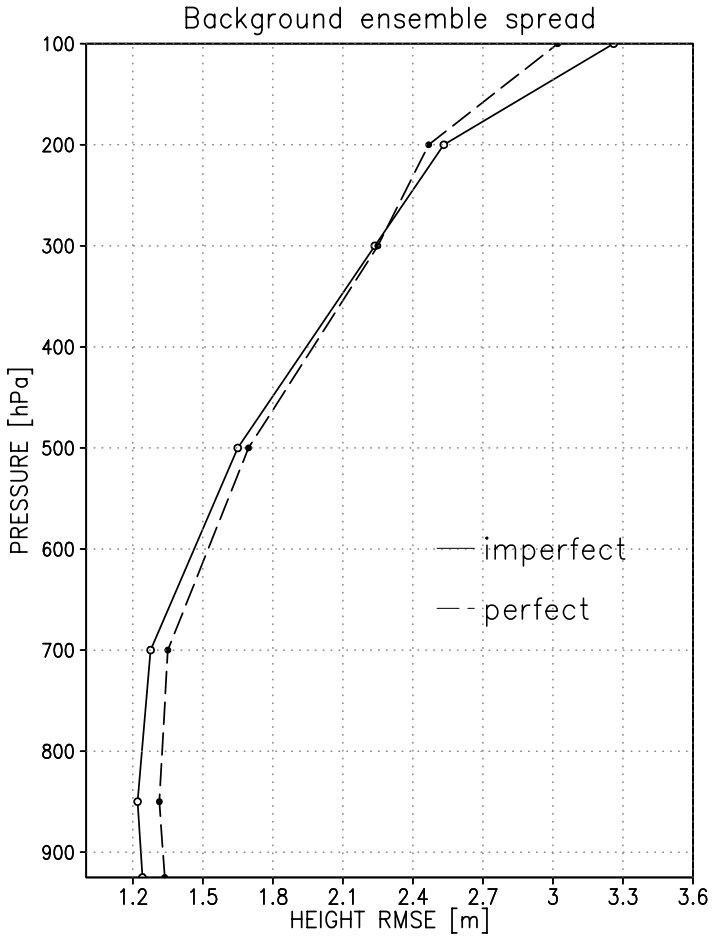

Figure 4.4: Background ensemble spread in height field at all pressure levels temporally averaged over a month after the initial spin-up period in the cases of perfect model assimilating observations generated from the SPEEDY 'nature' run (dashed curve) and imperfect model assimilating observations generated from the NNR field.

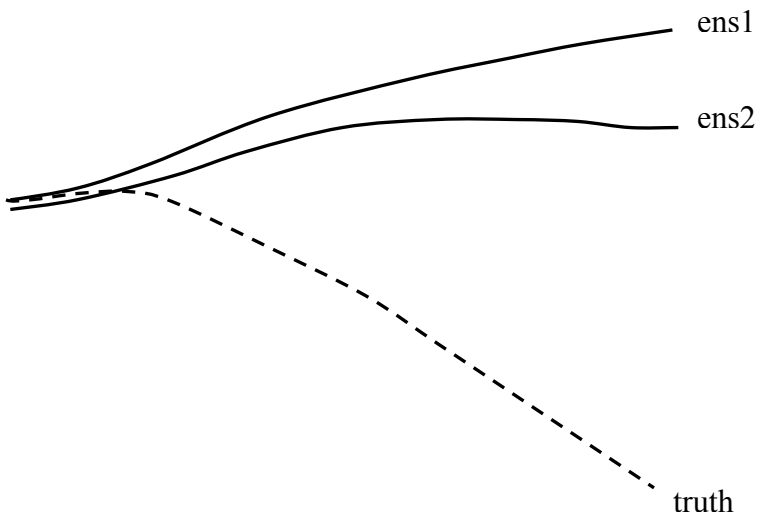

Figure 4.5: Schematic of background ensembles and the 'truth' in the presence of model errors. The solid curves represent the trajectories of two ensembles while the 'truth' (dashed curve) has its own dynamics.

the limited ensemble members in the case of perfect model, here it is intended to enlarge the ensemble forecast error covariance to account for model errors. This method assumes that model error has the same error structure as the dynamically evolved error since the enlargement is done in the space spanned by the ensembles.

We tune the inflation factor in terms of the analysis error. Table 4.1 provides the analysis RMSE of testing different inflation factors. Besides the 500 hPa height field which is often used as a proxy of the performance of a forecast or a data assimilation system, the RMSE are also shown for 200 hPa u, 850 hPa q and 925 hPa T where large constant biases or large diurnal biases are found. By increasing the inflation to be 1.0 or larger, all the results are significantly better than the 'control run' where a very small inflation factor $\Delta = 0.05$ is used. It seems that for different fields, the optimal inflation is different. For the geopotential height field, a large inflation ($\Delta = 2.0$) is required, while a much smaller inflation ($\Delta = 1.0$) is optimal for the 925 hPa humidity field. The RMSE for the 200 hPa u field seems less sensitive to the value of the inflation. These results indicate that we may need different inflation factors for different fields. We see that the SPEEDY model biases are complicated, depending on variables, levels and regions. Ideally, the optimal inflation factor should be a function of variables, regions, and levels. However, in practice, tuning that kind of inflation is prohibitive. In this study we still consider the inflation factor as a single number and choose the value of $\Delta = 1.5$ as its optimal choice, although it is not optimal for all the fields.

Table 4.1: Analysis RMSE of 200 hPa u, 500hPa Z, 850 hPa q and 925 hPa T fields, temporally averaged for a month after the initial spin-up period by applying different amplitudes of multiplicative inflation.

Fields (unit)	200 hPa u (m/s)	500 hPa Z (m)	850 hPa q (g/kg)	925 hPa T (K)
$\Delta = 0.05$ ('control run')	5.45	48.50	0.93	3.19
$\Delta = 1.00$	2.20	26.96	0.56	2.04
$\Delta = 1.50$	2.19	25.84	0.59	2.02
$\Delta = 1.75$	2.21	25.74	0.61	2.05
$\Delta = 2.00$	2.27	25.57	0.63	2.05
$\Delta = 2.25$	2.32	25.65	0.66	2.09

4.4.3 Additive inflation

To implement the additive inflation scheme, we randomly selected samples from a subset of NNR 6-hour tendency fields (Whitaker et al. 2007) in January and February for the years 1982-1986. Unlike random numbers, these randomly selected 6-hour tendency fields are geostrophically balanced. In each analysis cycle, we randomly select 30 tendency fields, scale them, and add the differences between these scaled fields to each background ensemble member, i.e.

$$\mathbf{x}_k^f = \mathbf{x}_{e\,(k)}^f + r\mathbf{q}_k \tag{4.1}$$

where we require that

$$\overline{\mathbf{q}}_k = 0 \tag{4.2}$$

and denote

$$\mathbf{q}_k \mathbf{q}_k^{\ T} = \mathbf{Q} \tag{4.3}$$

here k is the index for each ensemble member, and so $\mathbf{x}^f_{e\,(k)}$ denotes the *kth* ensemble forecast. $\mathbf{q}_k$ is the additive noise added to ensemble member k, and r is its amplitude. In this way, we increase the background ensemble spread without changing the ensemble mean. The modified forecast error covariance is actually given by

$$\begin{aligned}\mathbf{P}^f &= (\mathbf{x}^f - \overline{\mathbf{x}}^f)(\mathbf{x}^f - \overline{\mathbf{x}}^f)^T \\ &= (\mathbf{x}^f_e - \overline{\mathbf{x}}^f_e)(\mathbf{x}^f_e - \overline{\mathbf{x}}^f_e)^T + r^2\mathbf{q}\mathbf{q}^T \\ &= \mathbf{P}^f_e + r^2\mathbf{Q}\end{aligned} \tag{4.4}$$

$\mathbf{P}^f_e$ is the ensemble error covariance before adding the random noise. We tune the scale r for the random samples and found r =1.5 is optimal for our application. This value is much larger than the optimal value 0.33 found by Whitaker et al. (2007) in their application. One reason is that they added the random tendency fields to analysis ensembles rather than background ensembles; another important reason is that their assimilation model (the lower resolution version of the operational NCEP global forecast system) is much more realistic than the SPEEDY model used here, so that the model errors involved are much smaller in their case.

Figure 4.6 shows the analysis RMSE of 200 hPa u, 500 hPa Z, 850 hPa q and 925 hPa T fields, by applying the NNR 6-hour tendencies scaled by 1.5 as the additive noise. For comparison, the results from $\Delta = 1.5$ multiplicative inflation and the 'control run' are also shown. Both inflation schemes result in much better

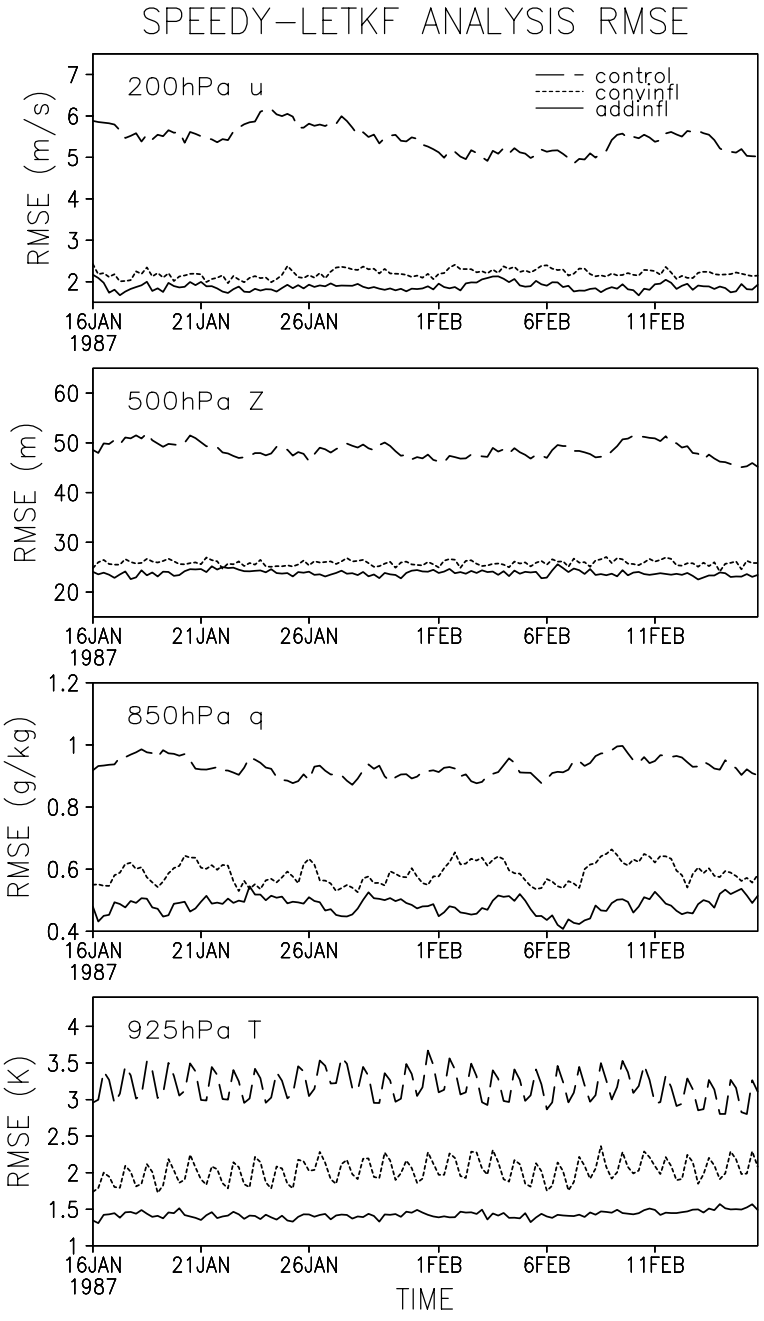

Figure 4.6: Time series of the global-averaged analysis RMSE in the cases of the 'control' run (dashed curve), Δ=1.5 multiplicative inflation (dotted curve) and r=1.5 additive inflation (solid curve). The four panels from the top to the bottom correspond to 200 hPa u, 500 hPa Z, 850 hPa q and 925 hPa T fields, respectively.

analyses than the 'control run' for all the fields. The time-mean of analysis error of 500 hPa height has been reduced by about 50%. Additive inflation outperforms multiplicative inflation in all fields especially in the 850 hPa q and 925 hPa T fields. The diurnal error signal is still clear in the 925 hPa T field with multiplicative inflation while it disappeared with additive inflation.

4.4.4 Dee and da Silva method with inflation (DdSM+)

The DdSM aims to estimate and correct model bias but does not deal with system-noise. In order to have a good performance of the LETKF, we use the additive inflation to account for system-noise. The additive noise is obtained in the same way as that in section 4.4.3. We call the DdSM augmented with additive noise DdSM+.

Recall that in the pure DdSM (section 2.3.3), the forecast model for bias is given by $\mathbf{b}_i^f = \mu \mathbf{b}_{i-1}^a$, and the bias forecast error covariance is assumed to be proportional to the state forecast error covariance, i.e., $\mathbf{P}_{bb}^f = \alpha \mathbf{P}_{xx}^f$. Thus we have two variables (μ and α) to tune. If we use additive inflation to model the system-noise, the amplitude (r) of additive noise is another parameter to be tuned.

To simplify the task of tuning the three parameters, first we fix α =0.5 (following the recommendation of Dee and da Silva 1998) and μ =1.0 (assuming a persistence model for bias prediction) and then tune the amplitude (r) of the additive noises. We start at 0000 UTC 1 January, 1987 by assuming zero bias and run the SPEEDY-LETKF system for one month and a half. We found that no matter how small r is, the filter diverges, especially for temperature fields in the lower levels. The bigger r is, the faster the divergence. One possible reason is that there is a strong

diurnal signal in the low-level temperature bias field (Figure 4.2) while the persistence model (μ=1.0) forecasts the bias to be equal to that 6 hours before, which is incorrect for the current time step. Ideally we may need a better bias prediction to capture this bias in a synoptic timescale. Here, to simplify, we reduce μ to be less than 1 and find μ=0.9 is successful for a wide range of choice of r (Todling, personal communication). μ=0.9 can be regarded as a forgetting factor to reduce the impact of bias from the previous time step.

Now we fix μ=0.9 and tune the pairs of (α, r). The results are summarized in Table 4.2. It is clear that accounting for random system-noise is essential in order to have a good performance of the LETKF. Without the additive noise, the pure DdSM (r=0) is not able to beat the pure additive inflation (α=0) with an optimal amplitude of r=1.5. By adding a small additive noise (r=0.25), the DdSM+ outperforms the pure additive inflation scheme but the optimal choice of α is large (α=0.75). When increasing r to be 0.5, the value of the optimal α reduces to 0.5. These results can be better understood by the expression of $\mathbf{P}_{bb}^{f} = \alpha \mathbf{P}_{xx}^{f}$ where α is an explicit parameter and r is an implicit factor (since $\mathbf{P}_{xx}^{f}$ is affected by r through equation 4.4) to determine the bias forecast error covariance $\mathbf{P}_{bb}^{f}$. When r is small, the system requires a big value of α to obtain an optimal $\mathbf{P}_{bb}^{f}$, while as r increases, the optimal value of α decreases because the forecast error covariance $\mathbf{P}_{xx}^{f}$ for the state variables has already been increased. By increasing r from zero to 0.5, a big improvement is found. Beyond r=0.5, it seems that there is almost no room to improve.

Table 4.2: Analysis RMSE of 500 hPa height using the DdSM+ with different choices of (α , r). When r =0 (i.e. pure DdSM), a small factor (Δ =0.05) of the multiplicative inflation is applied to prevent the filter divergence. For the other choices of r , no multiplicative inflation is used. For comparison, the pure addition inflation application is also shown (α =0 and r =1.5).

α	0.0	0.25	0.50	0.75	1.00
r =0			37.1	35.0	33.9
r =0.25		22.5	22.0	18.9	19.2
r =0.5		19.8	17.6	20.4	20.1
r =0.6			17.3		
r =1.5	23.8				

To compare the performances of the DdSM+ and the additive inflation scheme for the other variables and at the other levels, the analysis RMSE of zonal wind and temperature at all vertical levels are shown in Figure 4.7. With bias correction, a small additive noise (r=0.25) can beat the result from the no bias correction run, whereas a large amount of additive noise (r=1.5) is required to account for both the bias and the system-noise. The optimal results with bias correction (r=0.6) are significantly better than the pure additive inflation application. These results suggest that removing the bias from the ensemble mean is more accurate than accounting for them in the second moment of the ensemble, on condition that the estimated biases are good enough to represent the true forecast biases. We note that the control analysis error cannot be plotted in Figure 4.7 because it is 2-3 times larger than the error obtained with all the approaches to deal with model errors.

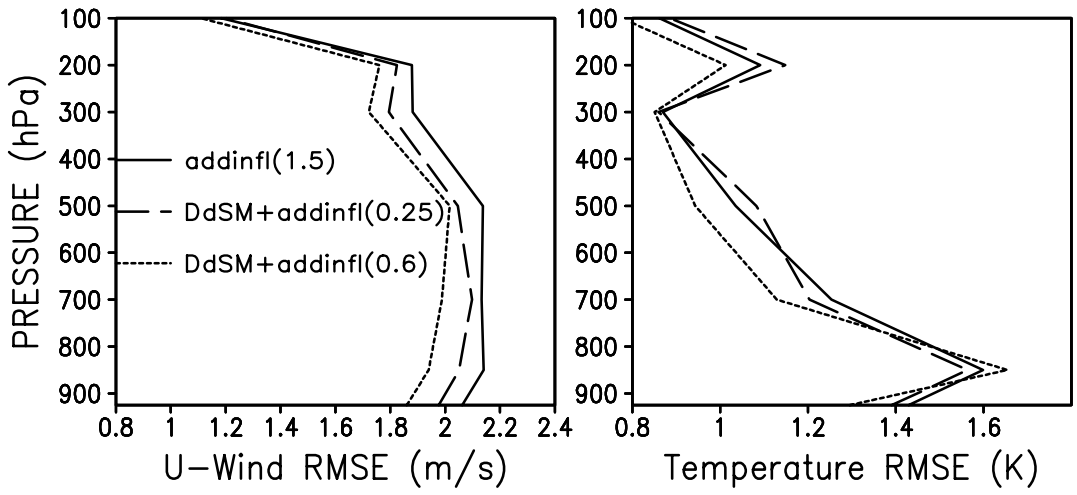

Figure 4.7: Analysis RMSE at all pressure levels temporally averaged for one month after the initial half month spin-up period for zonal wind (left) and temperature (right) in the cases of additive noise with amplitude r=1.5 (solid line), the DdSM together with additive noise with amplitude r=0.25 (dashed line) and the DdSM together with additive noise with amplitude r=0.60 (dotted line).

To assess the estimated bias fields from the DdSM+, we plot the time mean of the estimated bias field, time mean of the biased SPEEDY 6hr forecast, and of the debiased forecast (made by subtracting the estimated bias from the biased forecast) for 200 hPa zonal wind (Figure 4.8). Before the bias correction, the 6-hour SPEEDY forecast field is significantly biased. Large forecast biases can be found in polar regions or associated with the jet. It is remarkable that the estimated bias field captures the structure of the forecast bias well, though the amplitude is in general somewhat smaller. The actual forecast biases should be expected to be larger than the estimated values since the scheme is not perfect. In addition, the DdSM is designed to capture the slowly varying component of forecast errors while the remaining components can be accumulated as the data assimilation system is cycled. After

subtracting the estimated bias from the original biased forecast, the debiased forecast exhibits significantly less bias. With the debiased forecasts, it is not surprising that we obtain a better analysis than with the pure additive inflation application and that the required additive noise is much smaller since most of the model bias has already been removed. We use the term 'debiased' forecast rather than 'unbiased' forecast because the biases are not completely removed. The remaining biases are mostly in the scales close to the model resolution, which suggests that the DdSM is less efficient when the biases are in the smallest scales. This is confirmed in Figure 4.9 showing the DdSM+ does not capture the bias well in 850 hPa zonal wind where we see the small-scale characteristic of the model bias (see Figure 4.1). In our experiments, we have observations at every other grid point therefore the DdSM+ relies on the surrounding observations to estimate bias at those un-observed model grids. If the biases are in the scales close to the model resolution (e.g., the opposite sign of bias between two adjacent points), the estimation for the un-observed grids can be 'blind' or at least underestimate the 'true' bias. As a result, the 'debiased' forecast at 850 hPa is still biased and worse than that at 200 hPa.

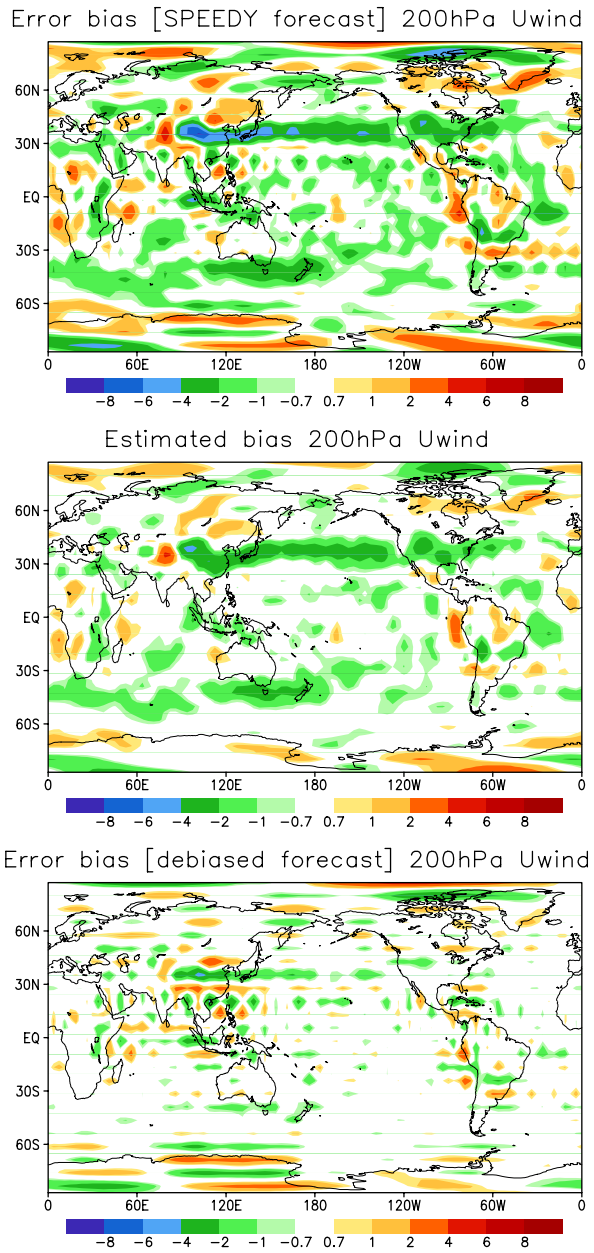

Figure 4.8: Top panel: Time-mean of SPEEDY 6-hour forecast (bias); Middle panel: Time-mean of the estimated bias field; Bottom panel: Time-mean of debiased forecast after subtracting the estimated bias from the SPEEDY forecast (bias of the debiased forecast) in the case of applying the DdSM+. The results are shown for 200 hPa zonal wind and temporally averaged for a month after the initial spin-up period.

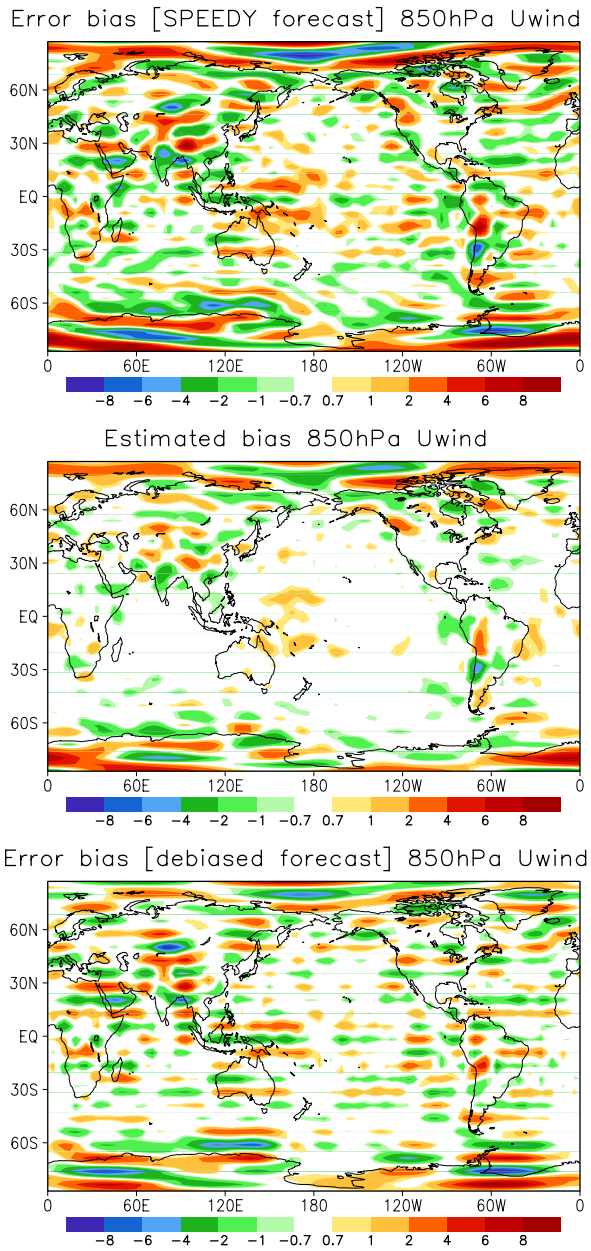

Figure 4.9: Same as Figure 4.8, except for 850 hPa

We have seen the large temperature diurnal biases near the surface over land (Figure 4.2), while the DdSM in its original design aims to correct the slowly varying bias. To see if it is able to handle the fast-varying bias, we plot the estimated bias of the 925 hPa temperature field at a specified location (120°E, 30°S) in Australia where a large diurnal bias is presented in Figure 4.2. As seen in Figure 4.10, the estimated bias follows the forecast error but the amplitude is much smaller. This is because the analyzed biases have been contaminated by the biases at the previous time, which are different from the bias at the current time. Nevertheless, the time series of estimated bias is in phase with that of the forecast error, so that at least the direction of bias correction is correct. With such small amplitude, the estimated bias is able to correct only a small portion of the diurnal bias (the debiased forecast exhibits a little smaller error than the biased forecast).

The DdSM+ works generally well, and is better than the additive inflation method. However, this bias correction scheme is relatively expensive since it doubles the analysis time. To address this problem, we implement a simplified version of the DdSM (Radakovich et al. 2001). The simplified algorithm modified the Kalman gain for bias $\mathbf{K}_b = \alpha \mathbf{P}_{xx}^f \mathbf{H}^T [(1+\alpha)\mathbf{H}\mathbf{P}_{xx}^f \mathbf{H}^T + \mathbf{R}]^{-1}$, to be

$$\mathbf{K}_b \approx \alpha \mathbf{P}_{xx}^f \mathbf{H}^T [\mathbf{H}\mathbf{P}_{xx}^f \mathbf{H}^T + \mathbf{R}]^{-1} = \alpha \mathbf{K}_{xx} \tag{4.5}$$

on condition that α is small enough. The expression of (4.5) implies that the analysis increment for the bias is proportional to the analysis increment for the state variable. Thus, the bias analysis process is computationally free if it is done after the state analysis has been updated (see section 2.3.3 in Chapter 2 for details).

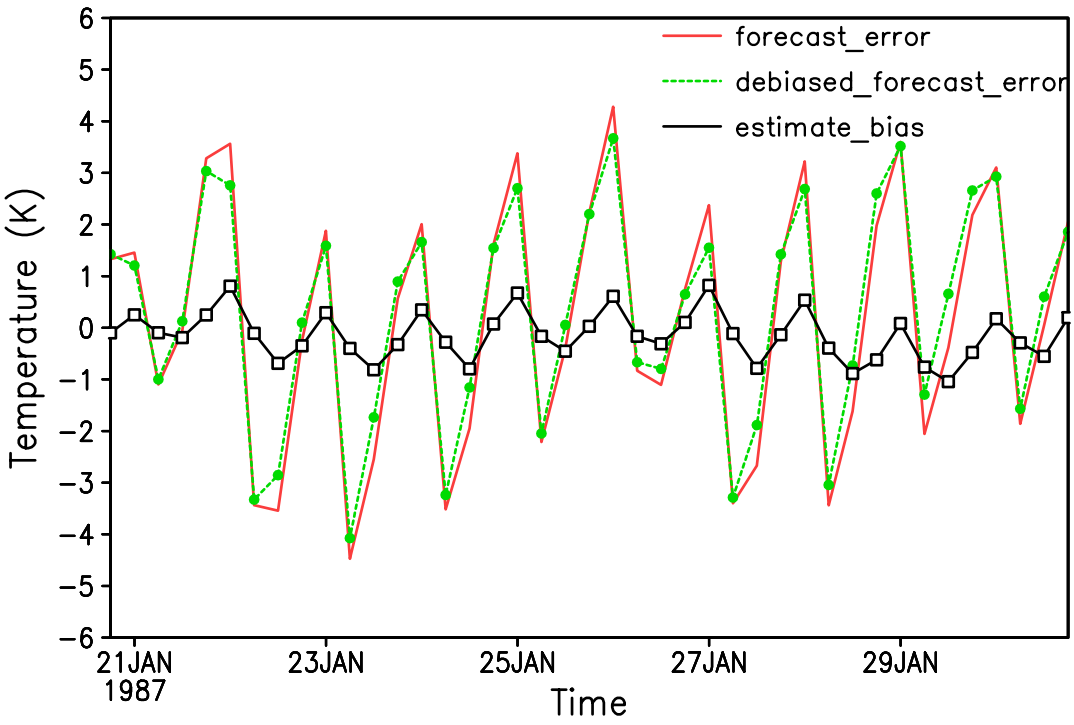

Figure 4.10: Time series of the biased SPEEDY forecast error (red line), the estimated bias by the DdSM+ (black line), and the error of the debiased forecast by subtracting the bias from the biased SPEEDY forecast (green dotted line) of the 925 hPa temperature field at (120°E, 30°S) where a large diurnal bias is presented.

We take the same setting of the parameters as the optimal choice used in the DdSM+ ($\mu = 0.9$, α =0.5, r =0.6) and run the SPEEDY-LETKF system with the simplified DdSM+. The resulting analysis is worse than that from the DdSM+ for all the variables but better than that from the pure additive inflation in the 500 hPa height field and comparable results are found in the other fields, except for specific humidity fields (Table 4.3).

This result is not unexpected. First, in the simplified approach, we use α =0.5 so that the assumption of a small α is not met. Second, the DdSM+ updates the bias

field first then uses the current 'analysis' bias to obtain the debiased forecast while with the simplified version we have to use the 'forecast' bias (usually taken from the previous time and, as a result, less accurate than the 'analysis' bias) since the bias analysis is done after the state variables have been updated.

Table 4.3: Comparison of analysis RMSE between the applications of additive inflation, the DdSM+ and the simplified DdSM+. Results are shown for 200 hPa zonal wind (u), 500 hPa height (Z), 850 hPa specific humidity (q) and 925 hPa temperature (T) fields, temporally averaged for a month after the initial spin-up period.

Fields (unit)	200 hPa u (m/s)	500 hPa Z (m)	850 hPa q (g/kg)	925 hPa T (K)
Additive inflation	1.88	23.82	0.48	1.43
Simplified DdSM+	1.87	19.54	0.61	1.37
DdSM+	1.76	17.32	0.47	1.30

4.4.5 Low-dimensional model error correction

The low-dimensional method (LDM) attempts to correct three types of model errors: 1) Time-averaged bias correction; 2) Diurnal bias correction; and 3) State-dependent error correction. Danforth et al. (2007) succeeded in correcting model errors during the forecast period while neglecting errors in the initial condition to be perfect. Here we test this method in a more realistic situation. The forecast-analysis is cycled so that we have both the errors in the initial conditions and the model errors.

We use the LETKF as the data assimilation scheme and focus on the impact of low-dimensional error correction on the performance of the LETKF.

A. Pure LDM

First we test the pure LDM, i.e. we do not use any additional method (e.g. additive inflation) to account for system-noise. A small value (Δ =0.05) of multiplicative inflation is applied to prevent the ensemble spread from collapsing.

a. Time-averaged bias correction

In this section we compare the impact of two different training periods to generate the time-averaged bias. Model errors are sampled by the difference between the NNR fields and the 6-hour SPEEDY forecasts started from the NNR field. To obtain a time-averaged bias, we need a subset of model error samples to do the average. The bias fields shown in Figure 4.1 are obtained by averaging the 184 samples over our experimental period between 0000 UTC 1 January 1987 to 1800 UTC 15 February 1987. However these samples would not be available beforehand in a real case. Here we test two feasible training periods over which we average the model error samples. One is one month prior to the experiment, e.g., if the experimental time is in January 1987 we use December 1986 as the training period; the other is based on the 5-year climatology for the years 1982-1986, following Danforth et al. (2007). The time-averaged bias is separated for January and February.

Figure 4.11 compares the analysis RMSE of 200 hPa zonal wind, 500 hPa temperature and 850 hPa specific humidity using time-averaged biases obtained from

two different training periods. It clearly shows that the climatological bias gives better results than the prior one-month bias estimate. The results for the other variables and at the other levels are similar. Thus, in the rest of our experiments for the low-dimensional bias correction, the samples of model error are from the 5-year climatological dataset.

b. Diurnal bias correction

We saw the strong signal of diurnal bias near the surface in temperature fields that is projected onto the first two leading EOFs (Figure 4.2). In order to correct these diurnal biases, we compute the first two leading EOFs of the error anomalous fields for temperature at the bottom two sigma levels (σ =0.95 and σ =0.835) using samples from the 5-year climatological dataset. The Principal Components (PCs) of these two modes are estimated by projecting the leading eigenvectors onto the error anomalous fields. The diurnal magnitude of the two modes is estimated by averaging the time-series of the PCs over the daily cycle for the years 1982-1986. A diurnal bias correction is computed as a function of the time of day. Figure 4.12 compares the temperature analysis RMSE at 925 hPa and 850 hPa with and without the diurnal bias correction. With only the time-averaged bias correction, the analysis RMSE is higher and a strong diurnal variability can be found. Correction including the diurnal bias reduces the global-averaged analysis RMSE by about 10% at 925 hPa and 5% at 850 hPa and this improvement can be mostly attributed to the error reduction over land (Figure 4.13).

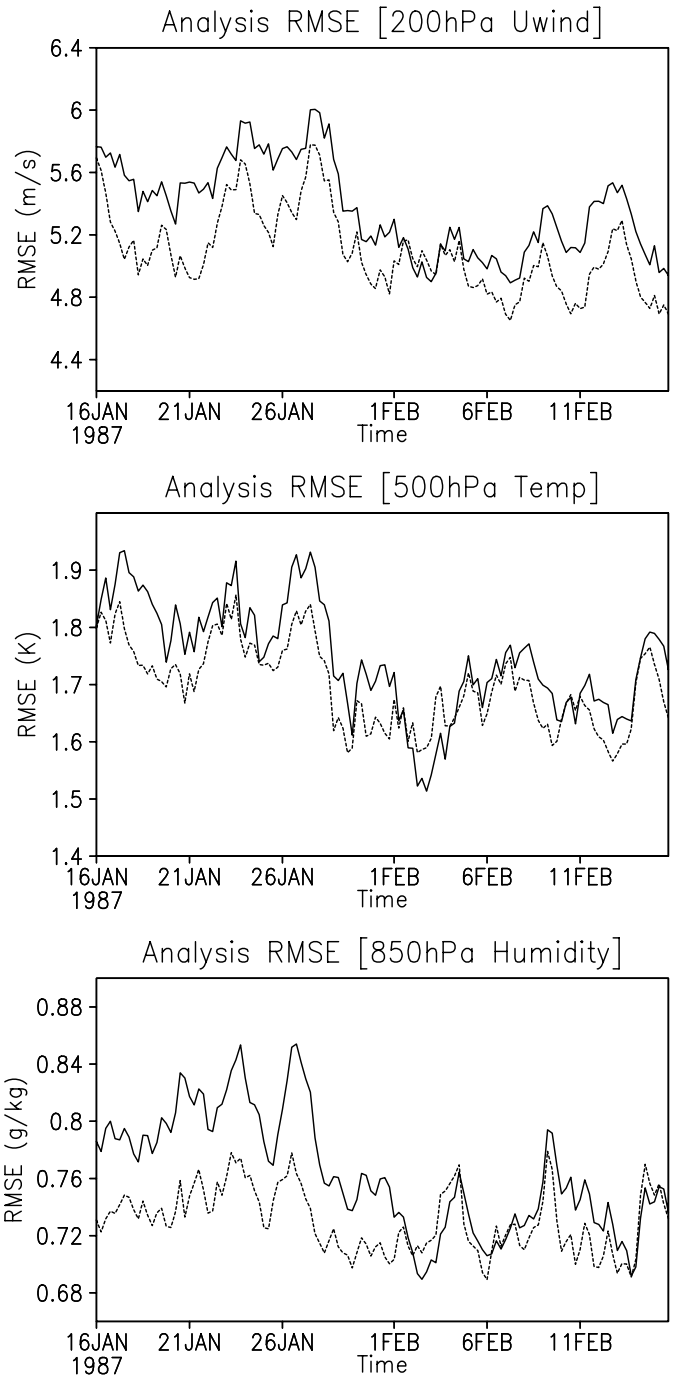

Figure 4.11: Comparison of analysis RMSE with time-averaged bias estimated from one-month prior samples (solid curve) and estimated from 5-year climatology (dotted curve). Time series are shown for 200 hPa zonal wind, 500 hPa temperature and 850 hPa specific humidity.

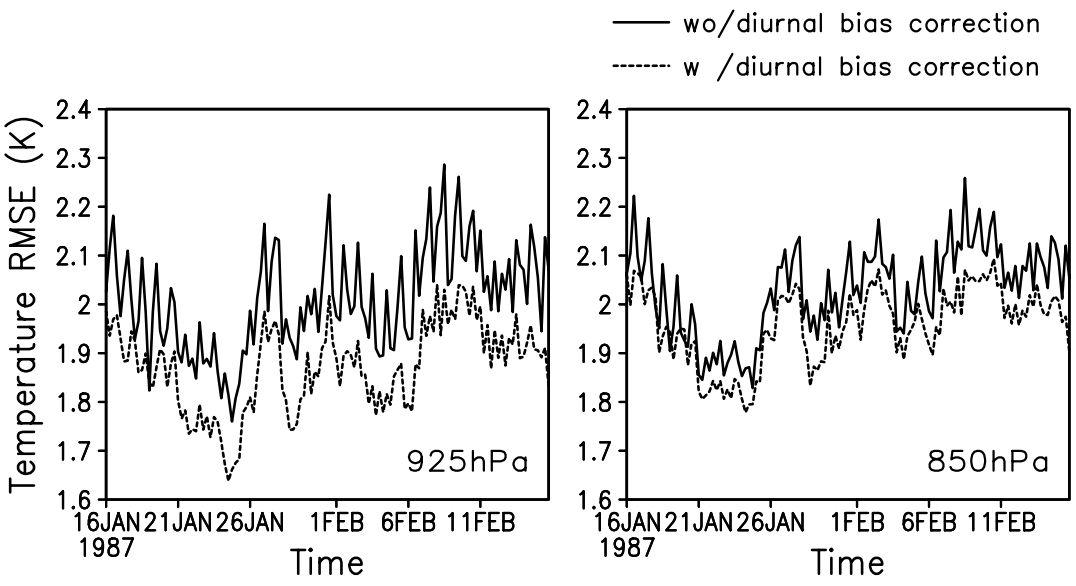

Figure 4.12: Time series of global-averaged analysis RMSE corrected for the constant bias with (dotted line) and without (solid line) diurnal bias correction for temperature at 925 hPa (left) and 850 hPa (right).

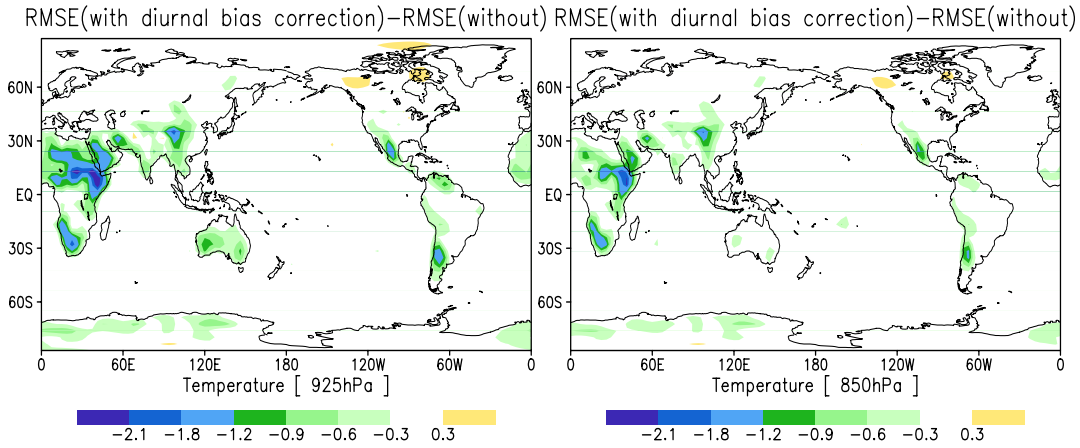

Figure 4.13: Analysis RMSE difference between the runs with and without diurnal bias correction, averaged for a month after the initial spin-up period, for temperature at 925 hPa (left) and 850 hPa (right). The green color indicates the RMSE reduction by including the diurnal bias correction.

c. State-dependent correction

In addition to the time-averaged and diurnal bias, we also compute the leading mode of the Singular Value Decomposition (SVD) of the covariance matrices between the state anomalies and the model error anomalies to allow a residual correction of the state-dependent errors (see section 2.3.5 in Chapter 2 for details). The SVD modes are generated separately for different variables and levels using model error samples from the years 1982-1986. Ten SVD modes are used to calculate the state-dependent model error anomalies. To avoid spurious long-distance correlations, following Danforth et al. (2007), the correlations between two grid points that are more than 5 grid points away from each other are forced to be zero.

The impact of the state-dependent error correction on the global averaged analysis RMSE is generally small but positive. For some of the variables at certain levels (for example, 500hPa temperature field in Figure 4.14), the positive and negative impact is mixed with time while for the others we observed the consistent positive impact (for example, 925hPa meridional wind field) though the impact is relatively small.

To better understand the results for the 925 hPa meridional wind field, we plot the three leading coupled SVD modes between the state anomalies and the model error anomalies using the climatological samples in January from the years 1982-86, and the time-averaged analysis RMSE reduction by including the state-dependent error correction (Figure 4.15). First we see the strongly coupled modes indicating a significant relationship between the model error anomalies and state anomalies. Then we find most of the analysis RMSE reduction actually occurs in the regions where the

coupled signal is the strongest. Since we assume a local structure in the covariance matrices, the positive impact is also restricted to some certain local regions. For those areas where there is no coupled signal, for example in the southern Hemisphere and tropics, the impact of the state-dependent error correction is negligible. This explains why the improvement for the globally averaged analysis RMSE is small, though locally the improvement can be as large as 30%.

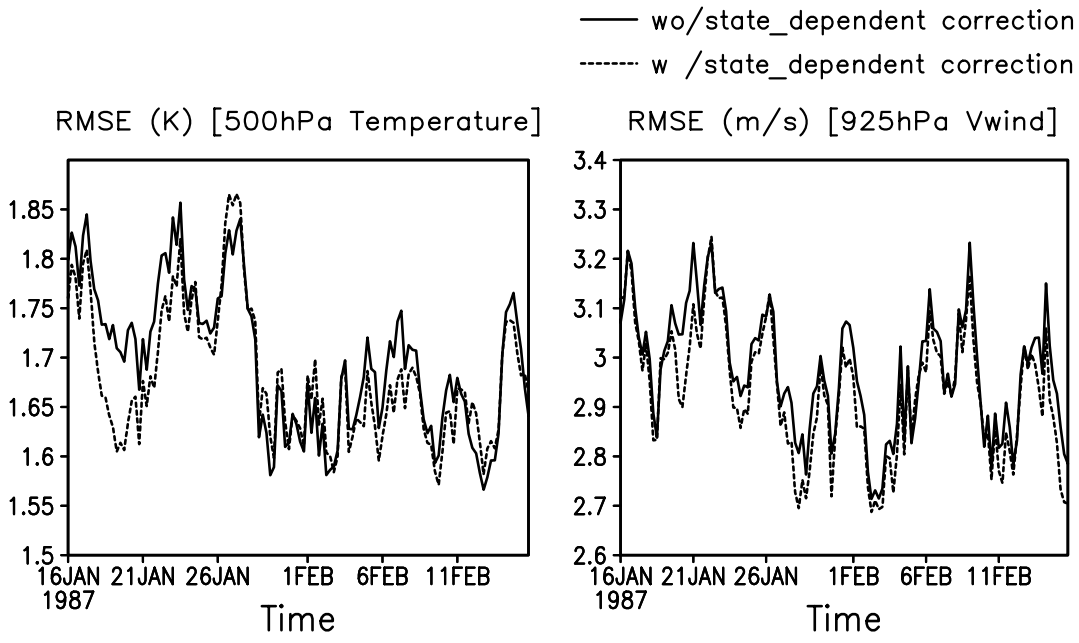

Figure 4.14: Time series of globally averaged analysis RMSE corrected for the constant and diurnal bias with (dotted line) and without (solid line) the state-dependent error correction for temperature at 500 hPa (left) and meridional wind at 925 hPa (right).

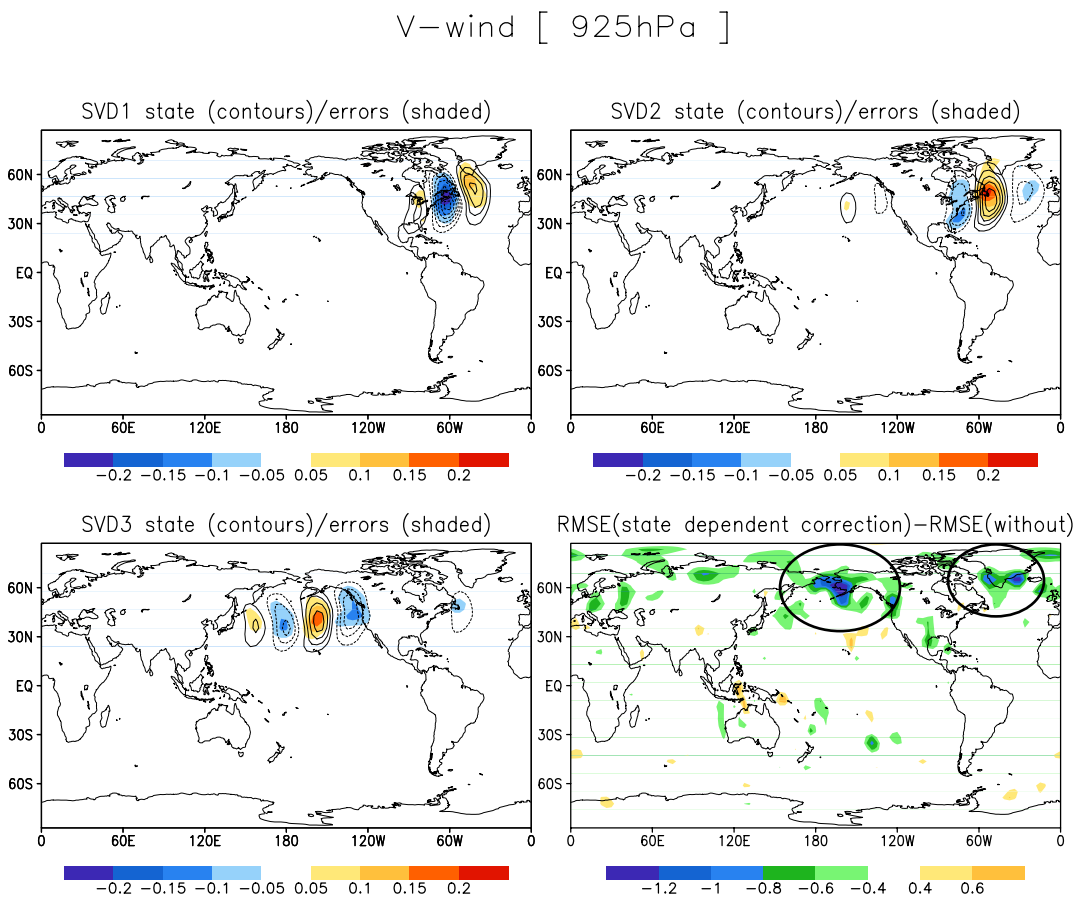

Figure 4.15: The first three leading SVD structures between the model state anomalies (contours) and the model error anomalies (shades) in meridional wind at 925 hPa in January for 1982-1986, and the analysis RMSE difference between the runs with and without the state-dependent error correction, averaged over the period from 11 January 1987 to 31 January 1987 (bottom right panel). The green color indicates the RMSE reduction by including the state-dependent error correction. Most of the reduction is found in the circled regions where the leading coupled signals are the strongest.

B. LDM with inflation (LDM+)

So far we have not dealt with system-noise. As in the case of the DdSM method, without handling this random noise, the LDM is not able to beat additive inflation with an optimal amplitude r=1.5 (Figure 4.16). To parameterize system-noise, we add randomly selected NNR 6-hour tendency field to each forecast ensemble member and tune their amplitude (additive inflation). The optimal value of r=0.4 in this case is much smaller than 1.5 which is optimal in the pure additive inflation application. The LDM, plus a small amount of additive noise (LDM+ hereafter), outperforms the pure additive inflation scheme everywhere. These results indicate that in the presence of complicated model errors we have to deal with both model bias and system-noise. For the biases, estimating and correcting them for the ensemble mean give better results than accounting for their effects in the second moment of the ensemble.

We have shown that the DdSM+ is less accurate when the biases are small-scale (Figure 4.9). To see if the LDM+ is able to handle this problem, we plot the same figure as Figure 4.9, but using the LDM+ (Figure 4.17). The estimated bias field captures the structure of the forecast bias quite well even in small scales. The resulting 'debiased' forecast is much less biased compared to that from the DdSM+. Unlike the DdSM+ estimating bias in observation space, the bias estimation in the LDM+ is done in model space. As a result, the LDM+ is less affected by the observation density and can capture the bias structure well at all model grid points as long as the bias structure in the training period is similar to that in the experimental period.

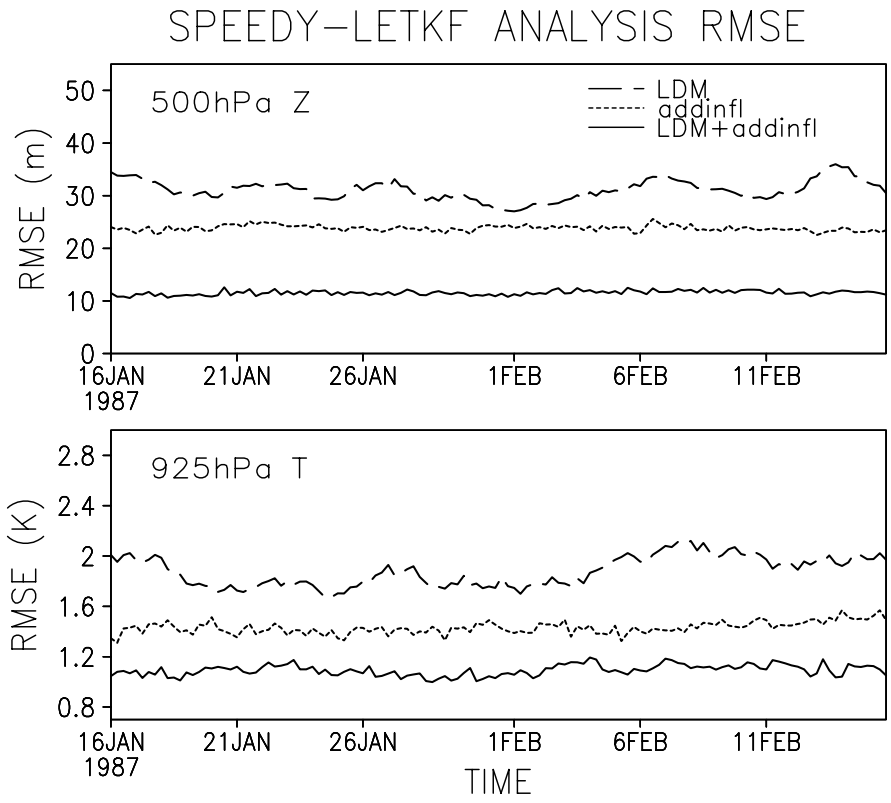

Figure 4.16: Time series of the global-averaged analysis RMSE of the 500 hPa Z and 925 hPa T fields, in the cases of the LDM alone (dashed curve), r =1.5 additive inflation (dotted curve) and the LDM together with additive inflation with an amplitude r =0.4 (solid curve).

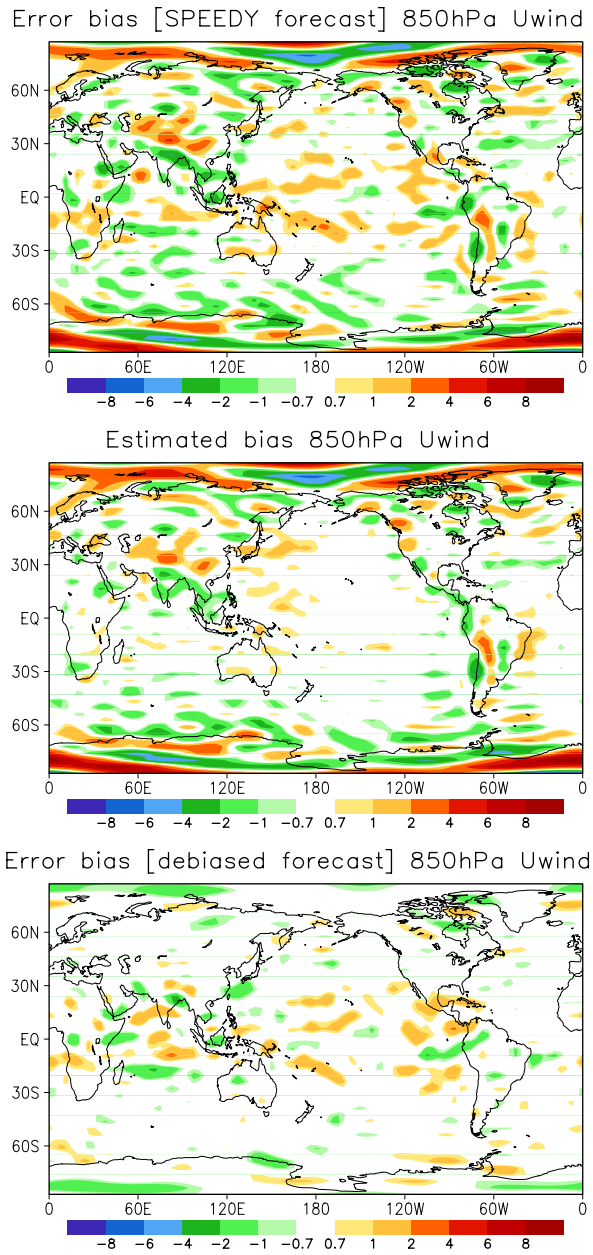

Figure 4.17: Same as Figure 4.9, except in the case of applying the LDM+.

4.4.6 Overall comparison

We have assessed the performance of each method separately. Finally we compare them with each other. As before, we verify the results against the NNR fields, and we choose optimal parameters for each method. For the LDM, the DdSM and its simplified version, we add additive noise and their optimal amplitude is chosen accordingly.

a. Analysis verification

As seen in Figure 4.18, the LDM+ provides much better analyses than the other methods for all the variables and at all levels. The DdSM+ generally outperforms both inflation schemes. Its simplified version is worse than the original version but is comparable to the additive inflation (except for specific humidity) and better than the multiplicative inflation scheme that has the worst results in all five methods. However, it should be noted that all the methods have made a huge analysis improvements compared to the 'control run' (red line in Figure 4.19). Thus, we can conclude that the LETKF would not work well without at least accounting for the effects of model errors, and that correcting model biases is, in general, better than only accounting for their effects in the second moment of the ensemble *on the condition that* we have a good method to estimate model biases. The estimated biases are not good enough for the simplified DdSM+ to outperform the additive inflation scheme for most of the variables. However, the DdSM+ generally beats the additive inflation. The performance of the LDM+ is remarkable. It outperforms the DdSM+ in

all the fields throughout all pressure levels especially at lower levels. For the zonal wind field, we have shown this is due to bias in the small scale, which the DdSM+ handles less efficiently. For temperature, the DdSM+ can only correct a small part of the diurnal bias.

To further understand what causes the analysis RMSE difference between different methods, we examine the analysis bias for each method. Since bias is a three-dimensional field, in a spatial average, a large positive bias in one region could cancel out a large negative bias in another region. To avoid this problem, we take the absolute value of bias before the spatial average (Figure 4.20). As the DdSM+ is better than its simplified version and additive inflation is better than the multiplicative inflation, we concentrate on comparing only three schemes: the LDM+, the DdSM+ and additive inflation alone. Figure 4.20 shows that the analysis bias of the DdSM+ is generally smaller than that of additive inflation, while the LDM+ analyses have much smaller bias than the other two methods. The bias difference between methods is nearly as large as the RMSE difference shown in Figure 4.18 (note that the plotting scales in Figure 4.18 and Figure 4.20 are the same). This indicates that the analysis RMSE difference between different methods is to a large extent attributable to their ability to handle model bias.

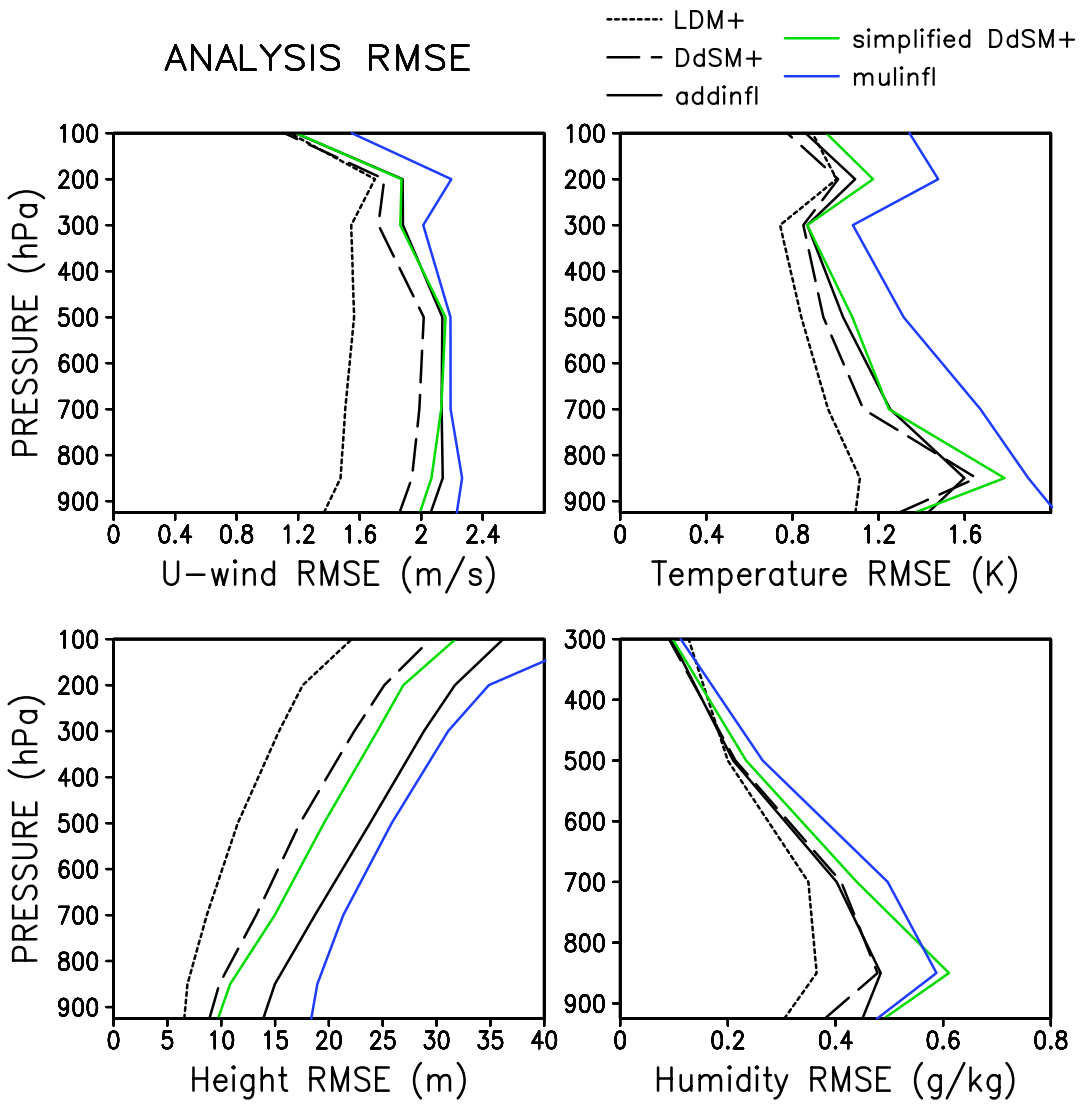

Figure 4.18: Analysis RMSE at all pressure levels in the cases of the LDM+ (black dotted line), the DdSM+ (black dashed line), additive inflation (black solid line), simplified DdSM+ (green line) and multiplicative inflation (blue line). The four panels correspond to u-wind field, temperature field, height field and specific humidity field, respectively. The averages are taken for a month after the initial half-month spin-up period.

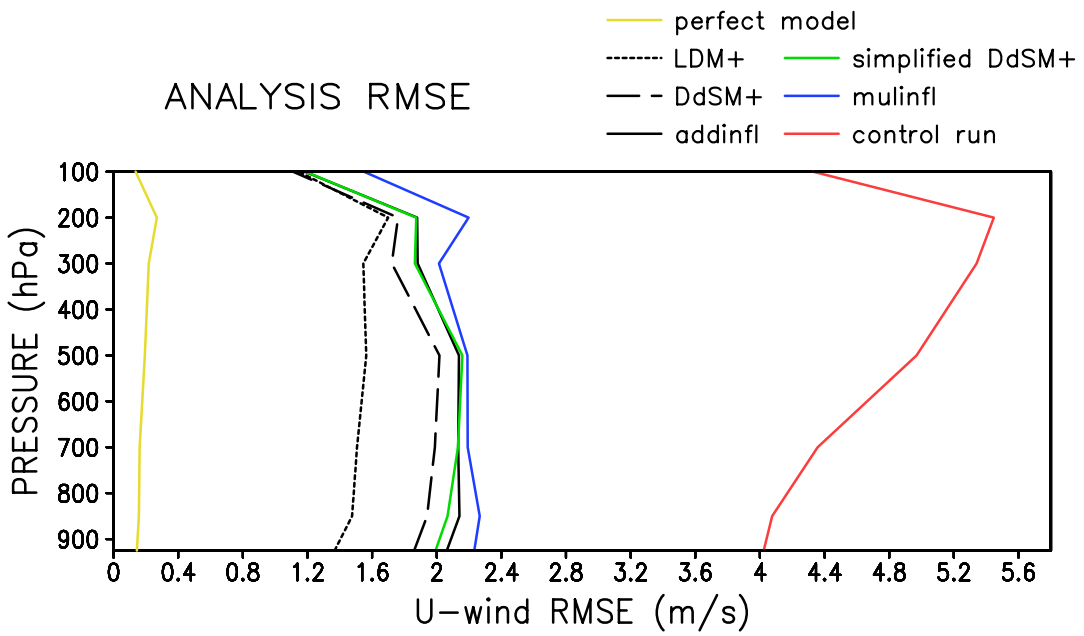

Figure 4.19: Same as the top-left panel in Figure 4.18, but also shows the result for the 'control run' (red line). We note again that in a more realistic operational model, the negative effect of model errors is not as large as in this 'control run' and a significant amount of inflation leads to better results even in the absence of bias correction (e.g., Szunyogh et al. 2007).

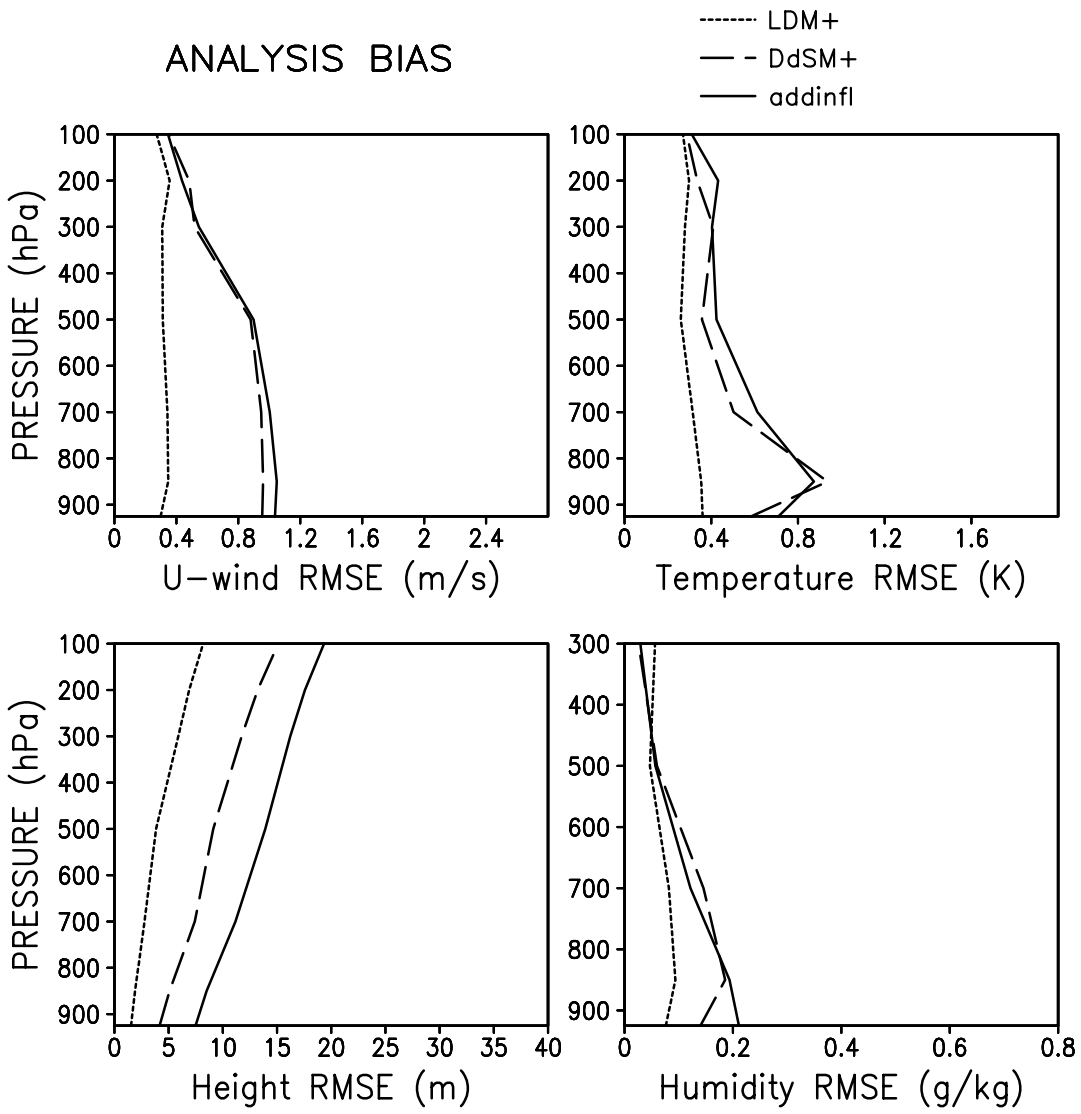

Figure 4.20: Global-averaged absolute value of the difference between analysis and the NNR 'truth' (analysis bias) at all pressure levels in the cases of the LDM+ (black dotted line), the DdSM+ (black dashed line) and additive inflation (black solid line). The four panels correspond to u-wind field, temperature field, height field and specific humidity field, respectively. The global-averaged absolute bias is calculated by averaging the difference between analysis and NNR over one month (bias) at each model grid point, then finding the absolute value for the bias fields, and finally taking a spatial average of these absolute values over the whole globe.

b. 48-hour forecast verification

So far we have focused on the comparisons in terms of the analysis accuracy. However, the goal of developing more accurate analyses is to improve the short-term forecasts. Within an imperfect model, the short-term forecast errors come from both growing errors in the initial condition and model deficiencies. Without correcting the model errors during the forecast process, we would like to see if the advantage of one method can be retained over the forecast period. Otherwise, there would be no benefit in improving the initial analysis on the short-term forecasts. Figure 4.21 shows the global-averaged 48-hour forecast RMSE at all pressure levels. The advantage of DdSM+ over additive inflation becomes less obvious for most of fields, but remains significant for geopotential height fields at all levels. The big advantage of the LDM+ over the other two methods also decreases due to the contamination of the model errors. However, it is still quite obvious and significant, except for the zonal wind above 200 hPa and the humidity above 700hPa.

Here we focused on the impact of initial analysis on the short-term forecast and did not attempt to correct the model errors during the forecast process. In reality, the low-dimensional method can be used to estimate and correct the short-term model errors in the forecast phase, as Danforth et al. (2007) have done with the SPEEDY model. This correction, which is easy to implement, would lead to much better short-term forecast.

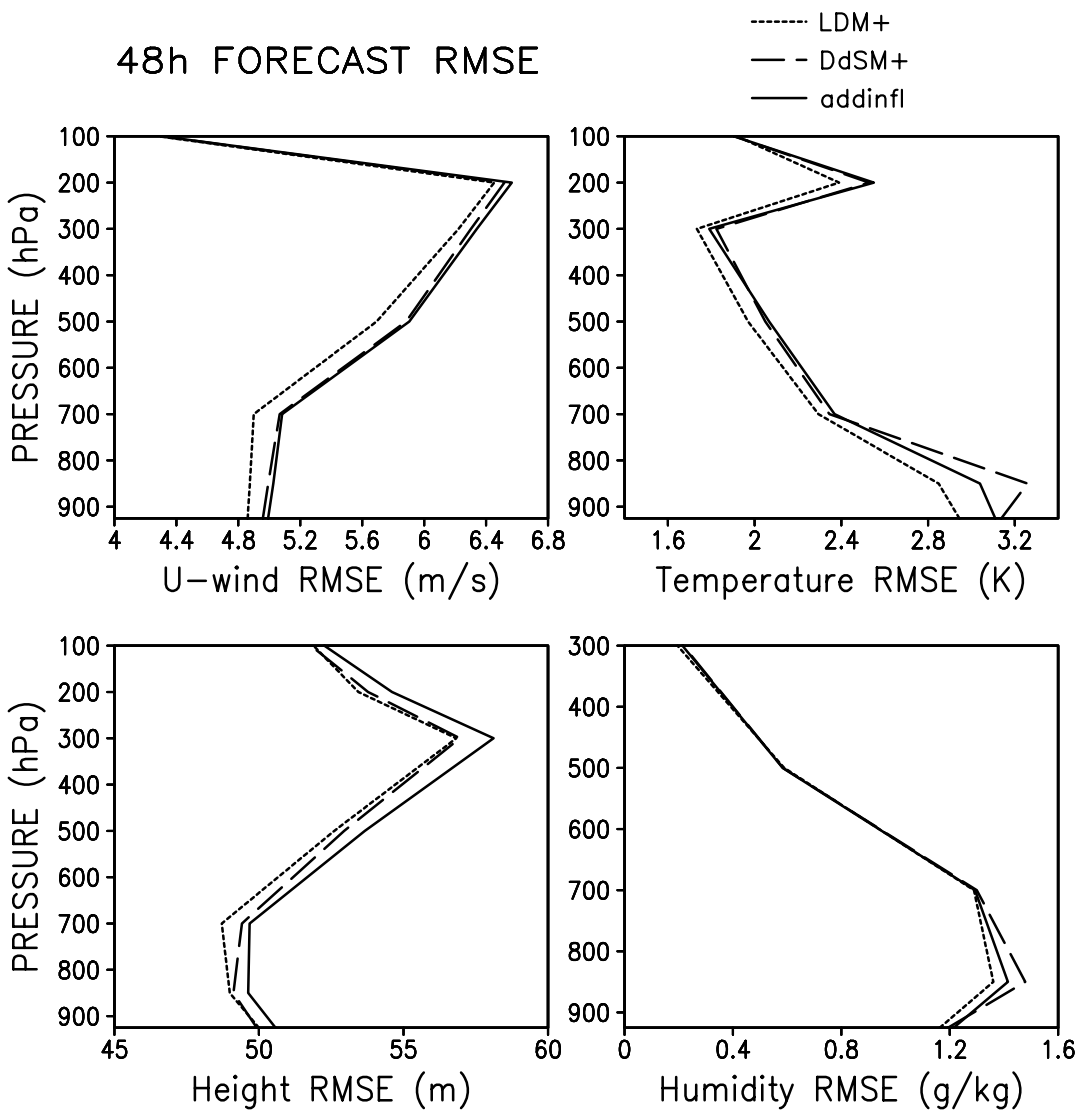

Figure 4.21: 48-hour forecast RMSE at all pressure levels in the cases of the LDM+ (black dotted line), the DdSM+ (black dashed line), and additive inflation (black solid line). The four panels correspond to u-wind field, temperature field, height field and specific humidity field, respectively. The averages are taken over all forecasts stated between 0000 UTC 1 February 1987 and 1800 UTC 15 February 1987.

4.5 Summary and discussion

In this chapter we dropped the assumption of a perfect model by assimilating observations simulated from NCEP/NCAR Reanalysis. Without correcting model errors or at least accounting for their effects, the performance of the LETKF on the SPEEDY model is poor. The ensemble spread is 'blind' to model errors as shown by the fact that we found ensemble spread of an imperfect model is similar in amplitude to that of a perfect model. As a result, in the presence of model errors, the background error estimated from the ensemble spread underestimates the actual forecast errors caused by both the dynamically evolved errors and the model deficiencies.

Methods implemented in this chapter to handle model errors include both multiplicative inflation and additive inflation which account for model errors by enlarging the background error variance, and two bias correction methods (Dee and da Silva method and low-dimensional method) which estimate the model biases and remove them from the forecast prior to the analysis. The results are summarized as follows:

i) Multiplicative inflation with a single inflation factor is worse than additive inflation. We may expect a better result for multiplicative inflation if applying different inflation factors for different variables and different areas.

ii) Model errors include model biases and system-noise. Inflation methods account for the total model errors while bias correction methods only correct model biases. The pure bias correction schemes (LDM and DdSM) give worse results than the inflation schemes.

iii) After accounting for system-noise, the bias correction methods with inflation (LDM+ and DdSM+) are generally superior to any of the inflation methods. The analyses are more accurate and less biased.

iv) The DdSM+ doubles the cost of the analysis, but its simplified version is almost cost free for updating the bias estimate. Although the results from the simplified version are worse than the original version, they are comparable to the additive inflation.

v) The DdSM+ relies on observations. As a result, it is less efficient for those unobserved points, especially when biases are very small-scale. The LDM+ uses NNR (or other reanalysis) and SPEEDY forecast, both in model space, to estimate model errors. Therefore, it is less affected by the observation density.

vi) In the absence of a good bias forecast model, the DdSM+ aims to correct the slowly varying bias. For the fast-varying bias (for example the diurnal bias in temperature fields at the lower levels), it is able to capture the time-varying signal but the correction is too small. By computing the EOFs of the bias anomalies, the LDM+ is effective in removing the diurnal bias.

vii) The LDM+ results in a dramatically better analysis than the DdSM+ especially at the lower levels. This may be explained by the reasons in v) and vi). The advantage is still significant in the 48-hour forecast.

The SPEEDY model has much larger errors than more advanced operational models. With operational GCMs, the analysis differences between different methods

would be much smaller. Our results of low-dimensional errors estimation may be too optimistic, since in our applications we assume the NNR field is an approximation of the unknown truth and use it to generate the samples of model errors. In practice, the NNR field could be biased and generating good samples of model errors is a challenge to the LDM+. The method we have used, deriving the bias from 6-hour forecast errors initialized from a reanalysis, can be used, and a more advanced reanalysis (e.g., ERA-40 or JRA-25) should yield a better sample than the NNR. Another possible way is to use the analysis increments obtained from the forecast-analysis cycles using the same model. In this case, iterations in the training period are required in order to obtain reasonable model error samples since the accuracy of the analysis and the goodness of model error samples are dependent on each other. We intend to investigate this idea and to see whether the final model error samples after convergence are good enough to represent the true model errors.

As for the DdSM+ we have seen it strongly relies on the observations. In our experiments, observations are pretty dense, while in practice the bias correction is expected to be less efficient in the global analysis system if rawinsonde soundings only are used. One may argue that satellite observations can be used to avoid this problem since their density is very high. However, satellite radiance observations themselves are subject to biases due to errors in the instrument calibration and in the observation operator (e.g. Harris and Kelly 2001). Correcting the satellite observation biases is another challenging area for data assimilation.

In summary, the performance of the LETKF can be seriously degraded when using real observations but without accounting for model errors. However, there are

several methods that can be implemented to account for these model errors. Of all these methods, the low-dimensional method with inflation (LDM+) where the time-averaged model bias, diurnal bias and state-dependent errors are estimated from a large number of 6-hour forecast errors, gives the best results.

Chapter 5

Simultaneous estimation of inflation factor and observation errors within the LETKF

5.1 Introduction

In chapter 4, we have seen the important role of multiplicative or additive inflation in the LETKF, but also experienced the inconvenience of manually tuning these parameters. Here we test the methods to estimate on-line the inflation factor. However, the estimation of inflation relies on the information of observation errors which itself is a key component in the accuracy of the resulting analysis. In this chapter we adaptively estimate observational errors (for each type of instrument) and the inflation factor for the background error simultaneously within the LETKF using the method we proposed in section 2.4, considering that these two estimates are dependent on each other. We first test our method in a low-order model then implement it into a more realistic model.

5.2 Low-order model results

5.2.1 The Lorenz-96 model

The Lorenz-96 model (Lorenz 1996; Lorenz and Emanuel, 1998) is given by

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F \tag{5.1}$$

where, $i = 1, \cdots N$, and the boundary is cyclic. As in Lorenz (1996), we choose N=40 and F=8.0 in which case this model behaves chaotically. Equation (5.1) is solved

with a 4th-order Runge-Kutta scheme using a time step of 0.01 non-dimensional units for which the Lorenz system is computationally stable. Lorenz and Emanuel (1998) have shown that a time increment of 0.05 roughly corresponds to 6 hours of the real atmospheric evolution so that we chose as our analysis time step $\Delta t = 0.05$.

5.2.2 Perfect model experiments

First we test our method in the perfect model scenario in which the multiplicative inflation is used to prevent filter divergence due to small ensemble size. We generate the 'true' state by integrating the Lorenz-96 model for 2000 analysis steps. Normally distributed random noise with standard deviation σ_o^2=1 is then added to the 'truth' to generate the observations. We assimilate these observations every analysis cycle using the LETKF with 10 ensemble members. Following the recommendation of Ott et al 2004, we use a cutoff based localization with a local patch l=6 which covers 13 model grids. 40 observations are assumed available and located at model grid points so that no interpolation is required. Since the normally distributed noise is uncorrelated and the error variance is 1, the true observation error matrix is diagonal, i.e. $\mathbf{R}_t = \sigma_{o(t)}^2 \mathbf{I} = \mathbf{I}$.

The Lorenz-96 model and the LETKF are then used to assimilate for 2000 analysis time steps, but results are only reported for the last 1000 steps.

a. Correctly specified observation variance

We first assume that the observation error variance is perfectly known, i.e., the specified value is $\sigma_{o(s)}^2 = \sigma_{o(t)}^2$=1. In this case we do not estimate the observation

errors, but attempt to estimate on-line the inflation parameter using this correctly specified observation error variance. We found that the "observed" inflation $\tilde{\Delta}^o$ directly obtained from OMB^2 or *AMB*OMB* (see section 2.4.1 for details) has large oscillations at each analysis time due to sampling of too few observations in this low-order model. In order to avoid an unrealistically large sampling error that may occur and abruptly ruin the estimation, we impose reasonable upper and lower limits in the "observed" inflation $\tilde{\Delta}^o$, e.g., $-0.1 \le \widetilde{\Delta}^o \le 0.2$ before applying the simple scalar KF smoothing procedure described in section 2.4.4. The final estimate of Δ after smoothing is then used to inflate the background ensemble spread. In a more realistic data assimilation system with a large number of available observations, Wang and Bishop (2003) have shown the "observed" inflation $\tilde{\Delta}^o$ calculated directly from OMB^2 remained within a reasonable range. In that situation, there is no need to prescribe a range for $\tilde{\Delta}^o$ but smoothing of the estimates might still be desirable.

Table 5.1 shows that OMB^2 and *AMB*OMB* methods produce similar results with estimated Δ around 0.04 and an analysis error of about 0.20. These results are quite similar to the best tuned constant inflation obtained from many tuning trials. The experiments in Table 5.1 will serve as a benchmark for the latter experiments where σ_o^2 is not perfectly specified.

Table 5.1: Time mean of adaptive inflation Δ and the corresponding analysis error, averaged over the last 1000 steps of a 2000-step assimilation when the observational error variance (specified) is perfectly known. For comparison, the value of best tuned constant inflation and its resulting analysis error are also shown.

Δ method	$\sigma^2_{o(s)}$	Δ	RMSE
OMB^2	1	0.044	0.202
*AMB*OMB*	1	0.042	0.202
(tuned) constant	1	0.046	0.201

b. Incorrectly specified observation error variance

In reality we do not exactly know the true value of the observation error variance, and the specified value used in the analysis is only an estimate. Our second experiment with the Lorenz-96 model is to use an erroneously specified $\sigma^2_{o(s)}$ which is either one quarter or 4 times the size of the true $\sigma^2_{o(t)}$, equivalent to one-half or twice the true observational error standard deviation. With large $\sigma^2_{o(s)}$=4.0, even if estimated Δ is similar to its optimal value the analysis RMS error is large (Table 5.2), because the LETKF gives much weight to the background and less weight to the observations, making the quality of the resulting analysis extremely degraded.

As for the case of $\sigma^2_{o(s)}$=0.25 $\sigma^2_{o(t)}$ we have noticed the estimated Δ has bumped into the upper-limit 0.2 of the prescribed range, $-0.1 \le \tilde{\Delta}^o \le 0.2$. We checked in detail and found the "observed" inflation $\tilde{\Delta}^o$ at each single analysis time step is always larger than 0.2, but we forced it to be 0.2. Therefore $\tilde{\Delta}^o$ indeed did not

represent the value estimated from equation OMB^2 or *AMB*OMB*. Our experience indicates that $-0.1 \le \widetilde{\Delta}^o \le 0.2$ is a reasonable range of Δ when σ_o^2 is correctly specified. There is no reason to assume it will still be within this normal range in this abnormal experiment. Without this constraint, we obtain Δ to be 7.67 (6.83) with the estimation method OMB^2 (*AMB*OMB*) and the resulting analysis rms error of 0.80 (0.79) much large than the optimal value of 0.2.

Table 5.2: Time mean of adaptive inflation parameter Δ and the resulting analysis error, averaged over the last 1000 steps of a 2000 step assimilation in the case that the specified observation variance $\sigma_{o(s)}^2$ is either 1/4 or 4 times the true $\sigma_{o(t)}^2$

Δ method	$\sigma_{o(s)}^2$	Δ	RMSE
OMB^2	0.25	0.2	0.264
*AMB*OMB*		0.2	0.263
OMB^2	4.0	0.040	1.790
*AMB*OMB*		0.016	1.439

c. Adaptive estimation of both the inflation and the observation error variance

We have seen that neither OMB^2 nor *AMB*OMB* work when estimating the inflation parameter if the specified observation error information is wrong. In the third experiment, we estimate the observation variance and inflation simultaneously by using *OMA*OMB* and OMB^2 (or *AMB*OMB*) followed by the simple KF method.

We start our experiment with an initial miss-specification of the observation error variance. Table 5.3 shows that no matter how poorly the initial specified $\sigma^2_{o(ini)}$ is (one-quarter or four times the true σ^2_o), the *OMA*OMB* method has the ability to correct it. It is remarkable that the time mean of estimated σ^2_o over the last 1000 analysis step is essentially the same as the true σ^2_o. With the corrected R matrix, we obtain a reasonable adaptive inflation Δ which is about 0.04 for all the cases in Table 5.3. The resulting analysis rms errors are also similar to that of the benchmark. The results are not sensitive to the initial incorrect value of $\sigma^2_{o(ini)}$, since σ^2_o is gradually corrected and reaches its 'true' value after a transition period no matter what initial value is specified.

We have shown that the estimation of the adaptive inflation alone does not work with incorrectly specified observation errors. By estimating the inflation and observation errors simultaneously, our method has the ability to retrieve both their 'true' values. We now check whether *OMA*OMB* can retrieve a correct observation error variance if the inflation is wrongly specified. From the previous experiments we know the optimal inflation factor is about 0.04. If we fix it and under-specify it to be 0.01, we get the estimated σ^2_o =10.33, suggesting the estimations of inflation factor and observation errors depend on each other. Unless one of them is perfectly known or at least approximates the 'truth', we have to estimate both of them simultaneously.

Table 5.3: As in Table 5.2, but adaptively estimating both the inflation factor and observation error variance

R method	Δ method	$\sigma^2_{o\,(ini)}$	σ^2_o	Δ	RMSE
*OMA*OMB*	OMB^2	0.25	0.999	0.044	0.204
	*AMB*OMB*		1.000	0.042	0.202
	OMB^2	4.0	0.998	0.042	0.202
	*AMB*OMB*		0.999	0.043	0.203

5.2.3 Imperfect model experiments

We have tested our method in the LETKF with the simulated observations and shown its ability to retrieve the true observation error variance and the optimal inflation parameter in such an idealized system. In this section we focus on a more realistic situation by introducing model errors. Recall all criteria upon which our method is based on the assumption that matrix $\mathbf{HK} = \mathbf{HP}^f\mathbf{H}^T(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1}$ is optimal, which is the case if and only if the specified matrix $\mathbf{P}^f$ and $\mathbf{R}$ in our system agree with the true covariances for background and observation. In the perfect model scenario, the required inflation is a small number and the inflated forecast error covariance $\mathbf{(1+\Delta)P}_e^f$ with a reasonable number of ensemble members usually can approximate well the true background error covariance, but this is not the case for an imperfect model. Without an additional method for model error correction,

covariance inflation has been used to provide an increase in the ensemble variance to account for the effect of model errors. In this case the forecast error covariance $\mathbf{P}^f = \mathbf{(1+\Delta)P}_e^f$ may not be good enough to represent the true background error covariance. Our goal here is to test whether our on-line estimation algorithm will still work well in a more realistic situation.

a. Random model errors

First, we investigate our scheme in the presence of random model errors in which the actual atmosphere is assumed to behave like a noisy version of the numerical forecast model. The evolution of the 'true' atmosphere is simulated by adding the zero-mean random noise to the Lorenz-96 model at each model time step:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F + \alpha \times \varepsilon_i \tag{5.2}$$

where $\varepsilon_i \sim N(0,1)$ and α is a constant factor. Our forecast model is the standard Lorenz-96 model shown in (5.1). In this way, we have introduced the random model errors.

In the perfect model experiments we have seen our method is not sensitive to the initially specified observational error variance and the method to calculate the "observed" inflation parameter. For brevity, we only test our method with $\sigma^2_{o\,(ini)}$=0.25 and use OMB^2 method to estimate the inflation parameter. Since more uncertainties are involved in the imperfect model experiments, we increase the ensemble size from 10 to 20.

Table 5.4 shows the estimated values of observation error, adaptive inflation and their resulting analysis errors by using *OMA*OMB* and OMB^2 simultaneously (case C) in the situation of random model errors with different amplitudes. For comparison, we also estimate on-line the inflation using the 'true' observation error variance (case B) or manually tune the system to find the optimal time-constant inflation (case A). It is remarkable that all three cases give similar results. The bigger the model error, the bigger the analysis error and the required inflation are. When the observation error is perfectly known, adaptive inflation (case B) reaches an analysis error similar to that from constant inflation. Without the correct observation error information initially, we estimate it on-line together with the estimation of inflation, and the 'true' σ_o^2 is also approximated (case C). The estimated σ_o^2 is essentially the same as the 'true' value of 1.0 when the amplitude of the random errors is α=4 and α =20, but still close enough to its 'true' value even with α=100. The resulting analyses are as good as those from the best tuned inflation. All of these indicate that the simultaneously adaptive algorithm is able to produce successful assimilations over a wide range of random model errors. Manually searching for the optimal time-constant inflation factor (case A) requires a considerable number of iterations for each value of α.

Table 5.4: Case A: the best tuned constant inflation and the resulting analysis RMSE; Case B: time mean of adaptive inflation (with perfect σ_o^2 =1) and the resulting analysis RMSE; Case C: time mean of adaptive inflation and observation error, estimated simultaneously on-line, and the resulting analysis RMSE. Each case is tested for different α, amplitude of random model errors. Results are averaged over the last 1000 analysis steps.

Case	A: fixed σ_o^2=1.0 (tuned) constant Δ		B: fixed σ_o^2=1.0 adaptive Δ		C: adaptive σ_o^2 adaptive Δ		
α	Δ	RMSE	Δ	RMSE	σ_o^2	Δ	RMSE
4	0.11	0.260	0.108	0.260	1.003	0.103	0.261
20	0.20	0.348	0.205	0.350	0.998	0.204	0.350
100	0.40	0.471	0.398	0.474	1.028	0.370	0.475

b. Systematic model bias

For our final experiment with the Lorenz-96 model, we introduce a systematic model bias. In the linear estimation theory, basis of most data assimilation schemes, both background and observation error vectors are assumed to be unbiased. Since this is not the case in reality, it would be ideal to estimate and subtract the bias before using the model forecast (Dee and da Silva 1998, Baek et al 2006, Danforth et al. 2007). Here we violate the assumption that background is unbiased in order to check the behavior of our method in a more realistic situation with model bias.

We generate the model bias as in Baek et al (2006) by adding a constant sine function to the forcing term in the Lorenz-96 model.

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F + \alpha \times \beta_i \tag{5.3}$$

where $\beta_i = 1.6\sin(2\pi\frac{i-1}{N})$ describes the spatial structure of the model bias and α determines its size. In Baek et al. (2006) α=1, corresponding to a '6-hour' forecast bias of $b_i = 1.6\sin(2\pi\frac{i-1}{N})\Delta t = 1.6\sin(2\pi\frac{i-1}{N}) \times 0.05 = 0.08\sin(2\pi\frac{i-1}{N})$. This bias is relatively small compared with the observation noise (1.0 in our experiments). Here we examine a wider range of model bias by applying different coefficients α. As in experiments for random model errors we also test our method with 20 ensemble members and with $\sigma^2_{o\,(ini)}$=0.25 and use OMB^2 to estimate the inflation parameter.

Table 5.5 shows the analysis results obtained from the best tuned inflation (case A), adaptive inflation with the 'true' observation error variance (case B), adaptive inflation and adaptive observation error variance (case C), in the presence of model bias. It is clear that model bias has a more negative effect on the assimilation system than random model error (Table 5.4). When α=1, cases A, B and C give a similar analysis error. As α increases, the best tuned inflation gets the best results. The mean values of adaptive inflation in case B are always smaller than the best tuned inflation, resulting in worse analysis than the constant inflation case. As for case C, the mean of adaptive σ^2_o is smaller than its true value. However it is interesting to see the

resulting analysis is actually slightly better than that from adaptive inflation with the true σ_o^2 in case B.

Table 5.5: As in Table 5.4, but in the presence of a constant model bias with different amplitudes (α).

Case	A: fixed σ_o^2=1.0 (tuned) constant Δ		B: fixed σ_o^2=1.0 adaptive Δ		C: adaptive σ_o^2 adaptive Δ		
α	Δ	RMSE	Δ	RMSE	σ_o^2	Δ	RMSE
1	0.35	0.405	0.332	0.408	0.952	0.379	0.408
3	0.80	0.554	0.658	0.570	0.948	0.718	0.564
5	1.20	0.627	0.946	0.638	0.900	1.075	0.633

To better understand the results of Table 5.5, we compare the forecast ensemble spread with the 'true' forecast mean error (ensemble mean minus the true state) averaged over all 40 variables for all three cases when model bias size α=3. Table 5.6 shows that the spatially averaged spread agrees well with the forecast error in case B but is larger than the forecast error in case A and case C. Recall that the inflation scheme inflates the ensemble covariance to approximate the 'true' forecast error covariance, i.e., $\mathbf{P}^f = (1+\Delta)\mathbf{P}_e^f$. In the presence of model bias, this implicitly assumes that the structure of model bias is the same as the dynamical growing error and therefore can also be represented by the ensemble spread. If this assumption is incorrect, i.e., the inflated ensemble covariance is not good enough to capture the true

forecast error structure, these errors lead to suboptimal analysis. In our experiments, the model bias is a sine-function in space and constant with time. This special structure is hard to be represented by the ensemble covariance $\mathbf{P}_e^f$. However the adaptive inflation estimation scheme *OMB*[2] knows nothing about these spatial representativeness errors since it is only concerned with the trace of covariance. Thus $\widetilde{\Delta} = \frac{\mathbf{d}_{o-b}^T \mathbf{d}_{o-b} - Tr(\mathbf{R})}{Tr(\mathbf{H}\mathbf{P}_e^f \mathbf{H})} - 1$ produces a single value of inflation which is optimal in the sense of spatial average but not for the individual observation. Thus, the spatially-averaged spread in Table 5.6 for case B is consistent with the forecast mean error but the resulting analysis is not optimal. The tuned inflation result is expected to be the best because the inflation factor is repeatedly tuned in terms of the resulting analysis error. The best tuned result overcomes the spatial representativeness errors by over-inflating the ensemble covariance to give more weights to the observations. These results are consistent with Anderson 2007 where an adaptive inflation from a hierarchical Bayesian was compared with the best tuned time-constant inflation. Our interest here is in case C where we simultaneously estimate both the inflation and the observational error variance. This approach tends to overcome the spatial representativeness errors in the forecast error covariance by both under-estimating the observation error variance and over-inflating the ensemble covariance. Thus, the estimated observation error variance is not optimal but the analysis is improved compared with case B. In order to get the best estimation of both σ_o^2 and the inflation factor, an additional method is required to remove the model bias. The reader is referred to Chapter 4.

Table 5.6: Time mean of observation error variance (σ_o^2), adaptive inflation (Δ), the ensemble forecast mean rms error and the ensemble forecast spread in the cases of A: best tuned constant inflation; B: adaptive inflation estimated with true observation error variance; C: simultaneous estimation of both σ_o^2 and Δ. Results are reported as an average over the last 1000 steps of a 2000-step assimilation.

Case	A: fixed σ_o^2=1.0 (tuned) constant Δ	B: fixed σ_o^2=1.0 adaptive Δ	C: adaptive σ_o^2 adaptive Δ
σ_o^2	1.0	1.0	0.948
Δ	0.80	0.658	0.718
Error	0.669	0.693	0.682
Spread	0.813	0.727	0.744

5.3 SPEEDY model results

In the previous section we have tested our algorithm in a low-order model where we have only one set of observations with one 'true' observation error variance. In this section we apply our approach to a more realistic model, assimilating a number of sets of observations. The size and unit of different sets of observations can be different. The SPEEDY model, an atmospheric general circulation model (AGCM) with simplified physical parameterizations is used for this purpose. For the time being, we are concerned with a perfect model context. Model errors are discussed in Chapter 4.

The observations are obtained by adding zero mean normally distributed noise to the 'true state', the two-month integration of the SPEEDY model from Jan 1 to Feb 28 in 1982. The observations are available on the model grid at every 4 grid points. The observed variables are zonal wind (u), meridional wind (v), temperature (T), specific humidity (q) and surface pressure (p_s) with error standard deviations of 1 m/s, 1 m/s, 1 K, 10^{-4} kg/kg and 100 pa, respectively.

We double the true observational errors to get our first guess of the observational errors. Within the LETKF, we estimate and correct these initially incorrect observation errors every analysis time step (6-hour). Since the value and unit for different observed variables are all different, we estimate the observational error variance for each observed variable separately.

From the experiments of Lorenz-96 model, we have seen that as long as the observational error is recovered, we can get similar results whatever we use equation OMB^2 or $AMB*OMB$ to estimate the inflation parameter. Therefore here we will only test equation OMB^2 which is more widely used.

Figure 5.1 shows the on-line estimated observational errors for each observed variable. The experiment starts from incorrectly specified observational errors with 2 m/s for u and v, 2K for T, $2*10^{-4}$ kg/kg for q and 200 Pa for p_s. After 30 analysis steps, i.e. about one week, the estimated observational errors are already very close to their corresponding true values. Of all sets of observation errors, temperature error converges fastest (in about 2 days). Since the estimation of the inflation factor by OMB^2 depends on the accuracy of the specified observation error covariance $\mathbf{R}$, there is a delay in the time needed for the inflation factor to reach a stably optimal value

(solid curve in Figure 5.2) compared to the case in which **R** is specified correctly (dashed-dotted curve in Figure 5.2). This dashed-dotted curve can be considered as the optimal choice of the adaptive inflation at each time step, and after the spin-up period, the solid curve follows the dashed-dotted curve very well. This indicates that no matter how poorly the observation error statistics are known initially, as long as we estimate and correct this information we can obtain a quite good adaptive inflation factor and, as a result, good analyses as well. A comparison of the analysis errors from the experiment estimating **R** with the one in which **R** is perfectly known, shows that our approach for estimating **R** works very well. Although the RMS errors are slightly larger than those from the perfect **R** case, they are already quite good (Figure 5.3).

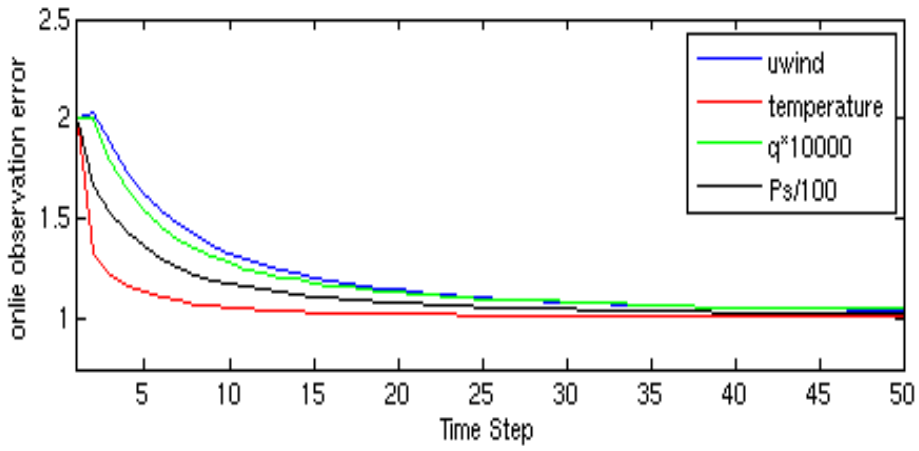

Figure 5.1: Time series of on-line estimated observation errors of u, T, q, and p_s for the first 50 analysis time steps (corresponding to 00z Jan 1 through 06z Jan 13, 1982)

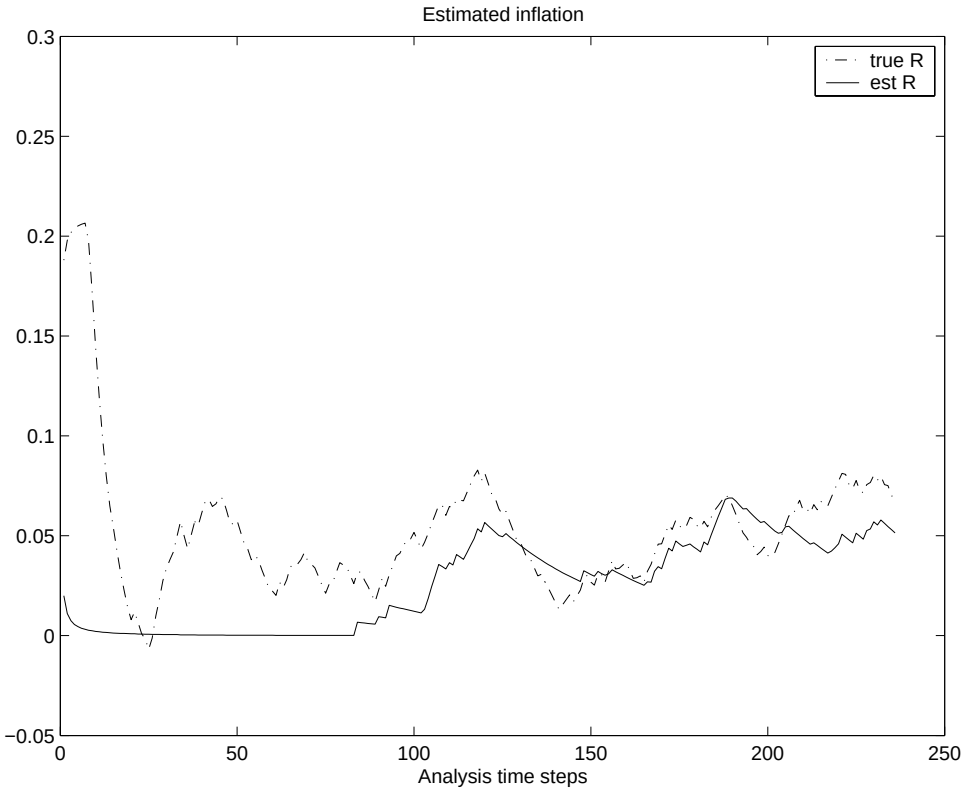

Figure 5.2: Time series of estimated inflation factor, in the cases of using a perfectly specified observation error variance (dashed-dotted line) and using an initially erroneous observation error variance but estimating it adaptively (solid line).

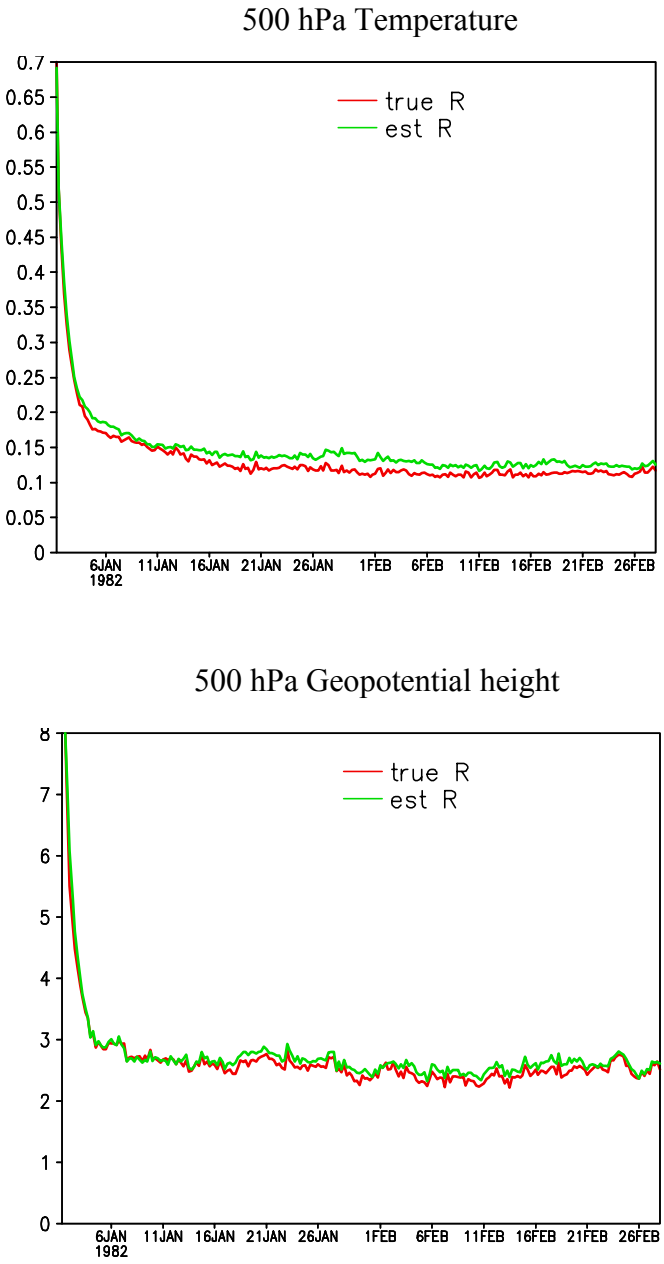

Figure 5.3: Time series of global averaged analysis RMS error of 500 hPa temperature and geopotential height for January and February 1982 with the adaptive inflation, in the cases of using a perfectly specified observation error variance (red line) and using an initially wrong observation error variance but estimating it adaptively (green line).

5.4 Summary

The accuracy of a data assimilation system depends on appropriate statistics for observation and background errors. For the ensemble based Kalman filter, tuning the inflation parameter is expensive. The on-line estimation method can objectively estimate the inflation parameter but requires the accurate information of observational errors. In this study, we estimate observational errors and the inflation coefficient simultaneously within the LETKF. The results are summarized as follows:

i) The adaptive estimation of the inflation parameter with perfect observation error variance produces results that are as good as the best tuned inflation values. Two adaptive methods, OMB^2 and *AMB*OMB*, give similar results. By contrast, tuning (i.e., searching for the best time-constant inflation) requires repeating many assimilations.

ii) Adaptively estimating the inflation parameter alone does not work without estimating the observational errors if the specified observational error variance is not correct.

iii) By estimating the inflation factor and observation error variance simultaneously, our method works perfectly in the cases of perfect model and random model errors. The estimated observation error variances are close enough to its true value, and the resulting analyses are as good as those from the best tuned inflation.

iv) When the forecast model has systematic bias, the best tuned inflation gives the best analysis by over-inflating the ensemble covariance. Adaptive inflation with a perfect observation error variance reports a worse analysis result. If we

also estimate on-line the observation error variance, the estimated value under-estimates the true observation error variance but helps to improve the analysis, since in that case more weight goes to the observations, which acts to reduce the impact of systematic model bias.

v) The SPEEDY model experiments show that the estimation of observation error variance is successful if applied to a more realistic high-dimension model to retrieve the true error variance for different types of instrument separately.

Chapter 6

Conclusions and future directions

This dissertation has addressed several issues relating to EnKF for assimilating real data, 1) model errors, 2) inconvenience or infeasibility of manually tuning the inflation factor when it is regional and/or variable dependent and 3) erroneously specified observation error statistics. Our results should help to accelerate the development of EnKF systems towards operational applications.

First, we addressed the issue of the model errors. In Chapter 3, we performed data assimilation experiments with the LETKF, an efficient approach within the EnKF family, with the SPEEDY model under a perfect model scenario. Our results show that the background ensemble spread captures the true forecast error very well, both in the error structure and the error amplitude. We removed the perfect model assumption in Chapter 4 by assimilating observations generated from the NCEP/NCAR reanalysis fields. Without any additional effort to handle model errors, the performance of the LETKF is seriously degraded. The background ensemble spread in this case is still similar to that in the perfect model and therefore much smaller than the true forecast error that also includes the model errors. The "blindness" of the LETKF to model error is due to the fact that each ensemble member is integrated with the same model. If more forecasts from different systems are available, a multisystem ensemble (Krishnamuti et al. 2000) where we have an ensemble from different models, may be expected to at least partially represent model errors.

In Chapter 4 we investigated two simple ways to represent the effect of model errors and two sophisticated methods to estimate and remove model bias. Our results suggest that multiplicative inflation with a single inflation factor is worse than additive inflation. The pure bias removal methods (DdSM and LDM) remove model bias, but cannot handle system noise; as a result, they are not able to beat inflation schemes that account for the total model errors. Supplemented by additive noise for representing the system noise, bias removal methods generally outperform the inflation schemes. Of all these methods, the low-dimensional method with additive inflation (LDM+) where the time-averaged model bias, diurnal bias and state-dependent errors are estimated from a large number of 6-hour forecast errors, gives the most accurate analyses and 48-hour forecasts. Although the DdSM+ produces worse results than the LDM+, it is generally superior to both inflation schemes. The main disadvantage of this method is the doubled computational cost and its exclusive reliance on observations. When the observations are sparse, the impact of the bias correction in a global analysis system is limited. In the worst case, where the observations themselves are biased, it is not at all obvious that this algorithm can work correctly.

Generating good samples of model errors is a challenge for implementing the LDM+. It is not clear in the real world whether the model error samples generated from the NNR fields are good enough to represent the true model errors. In practice, we could use a more advanced reanalysis, like ERA-40 or JRA-25. Another possible way is to use the analysis increments as model error samples. For the training of error samples, iterations are required since at first the analyses are also biased, so that the

analysis increments cannot sample model error well. We intend to explore this idea with the SPEEDY model to see whether the final model error samples after convergence are good enough to represent the true model errors.

In Chapter 5 we addressed the issues of adaptive estimation of the inflation factor and observation errors. It was found that the estimation of inflation alone does not work without accurate observation error statistics, and vice versa. Therefore we proposed to simultaneously estimate both inflation and observation error variance on-line. Our method was then investigated with a low-order model, the Lorenz-96 model. The results showed that our approach works impeccably in the cases of perfect model and random model errors. The estimated observation error variances are very close to their true value, and the resulting analyses are as good as those from the best tuned inflation value. When the forecast model has systematic bias, our algorithm tends to account partially for model bias by underestimating the observation error variance to give more weight to the observations. As a result, the estimated observation error variance is smaller than the true value, but the resulting analysis errors are comparable with the best tuned inflation value. Finally, we apply our approach to a more realistic high-dimension model, assimilating a number of sets of observations that have errors of different size and units. The SPEEDY model experiments show that the estimation of observation error variance is successful in retrieving the true error variance for different types of instruments separately.

In this study we have addressed the issue of observation error variance but the observational error correlation is another big concern, especially for satellite retrievals. We may extend our approach to estimate off-diagonal terms in the

observation error covariance. Investigations are needed to see whether our approach will be able to estimate adaptively the observation error correlations as well.

Based on our entire research in this dissertation, we suggest the development of a more advanced LETKF with both bias correction and adaptive estimation of inflation within the system. We have seen that the pure bias removal scheme does not work well. To account for system noise, we may still need multiplicative or additive inflation, for which manually tuning the amplitude is expensive. By applying our simultaneous on-line estimation method, we can estimate the regional and/or variable dependent inflation factors. In this case we could expect to obtain a good observation error statistics as well since the model bias has been removed.

In my future job, I plan to implement the LETKF in the regional WRF model for forecasting in the southeast region of China. The lateral boundaries and the scales of weather phenomena are expected to bring more challenges and perhaps require further development of new techniques.

Bibliography

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884-2903.

— 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, **59A**, 210-224.

Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the non-linear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.,* **127,** 2741–2758.

Anderson, 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634-642.

Baek, S.J., B. R., Hunt, E. Kalnay, E. Ott, and I. Szunyogh 2006: Local ensemble Kalman filtering in the presence of model bias, *Tellus*, **58A**, 293-306.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420-436.

Cardinali, C, S. Pezzuli, and E. Andersson, 2004: Influence matrix diagnostic of a data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **130**, 2767-2786.

Carton, J. A, G. Chepurin, and X. Cao, 2000: A simple ocean data assimilation analysis of the global upper ocean 1950-95. Part I: Methodology. *J. Phys. Oceanogr.*, **30**, 294-309.

Chapnik B., G. Desroziers, F. Rabier and O. Talagrand, 2006: Diagnosis and tuning observational error in a quasi-operational data assimilation setting. *Quart. J. Roy. Meteor. Soc.*, **132**, 543-565.

Chepurin, G., J. A. Carton, and D. Dee, 2005: Forecast model bias correction in ocean data assimilation. *Mon. Wea. Re*v., **133**, 1328–1342.

Corazza, M., E. Kalnay, D. J. Patil, S.-C. Yang, R. Morss, M. Cai, I. Szunyogh, B. R. Hunt, and J. A. Yorke, 2002: Use of the breeding technique to estimate the structure of the analysis "error of the day". *Nonlinear Processes in Geophysics*, **10**, 233-243.

Corazza, M., E. Kalnay, and S.-C. Yang, 2007: An implementation of the Local Ensemble Kalman Filter for a simple quasi-geostrophic model: Results and comparison with a 3D-Var data assimilation system. *Nonlinear Processes in Geophysics*, **14**, 89-101.

Danforth, C., E. Kalnay, and T. Miyoshi, 2007: Estimating and correcting global weather model error. *Mon. Wea. Rev.*, **135**, 281-299.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Re*v., **123**, 1128-1196.

Dee, D. P. and A. M. Da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269-295.

Dee, D.P and T. Ricardo, 2000: Data assimilation in the presence of forecast bias. The GEOS moisture analysis. *Mon. Wea. Re*v., **128**, 3268–3282.

DelSole T. and A. Y. Hou, 1999: Empirical correction of a dynamical model. Part I: Fundamental issues, *Mon. Wea. Rev.*, **127**, 2533–2545.

Desroziers G. and S. Ivanov, 2001: Diagnosis and adaptive tuning of osservation-error parameters in a variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **127,** 1433-1452.

Desroziers G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385-3396.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10143-10162.

— 2003: The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343-367.

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

Hamill, T.H. and J.S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.*, **133**, 3132-3147.

Harris, B. and Kelly, G. 2001: A satellite-bias correction scheme for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1453–1468.

Houtekamer, P. L and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

— 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604-620.

Hunt, B. R., E. Kalnay, E. J. Kostelich, E. Ott, D .J. Patil, T. Saucer, I. Szunyogh, J. A. Yorke, and A. V. Zimin, 2004: Four-dimensional ensemble Kalman filtering. *Tellus,* **56A**, 273–277.

Hunt, B. R., E. Kostelich, I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, **230**, 112-126.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: operational, sequential, and variational. *J. Met. Soc. Japna*, **75(1B)**, 181-189.

Kalman, R. E., 1960: Anew approach to linear filtering and prediction problem. *J. Basic Eng., Trans. ASME*, 35-45.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press*,* 341 pp.

Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera, 2007: 4D-Var or ensemble Kalman filter? *Tellus,* **59A,** to appear

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, Roy Jenne, and Dennis Joseph, 1996: The NMC/NCAR 40-Year reanalysis project. *Bull. Amer. Meteorol. Soc,* **77**, 437–471.

Keppenne, C. L., M. M. Rienecker, N. P. Kurkowski1, and D. A. Adamec, 2005: Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction. *Nonlinear Processes in Geophysics,* **12,** 491–503.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E, Williford, S, Gadgil, and S. Surendran, 2000: Moltimodel ensemble forecasts for weather and seasonal climate. *J. Climate,* **13**, 4196-4216.

Lorenz E. N., 1996: Predictability – A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1-18.

Lorenz E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.

Leith, C. E., 1978: Objective methods for weather prediction, *Annual Review of Fluid Mechanics,* **10**, 107–128.

Liu, J., E. Fertig, H. Li, I. Szunyogh, B. R. Hunt, E. Kalnay, E. Kostelich, and R. Todling, 2006: Application of local ensemble transform Kalman filter: perfect model experiments with NASA fvGCM model. *Preprint of 86th AMS annual meeting.*

Martin, M. J., Bell, M. J. and Nichols, N. K. 2002: Estimation of systematic error in an equatorial ocean model using data assimilation. *Int. J. Numer. Meth. Fluids,* **40**, 435-444.

Miyoshi, T. 2005: Ensemble Kalman filter experiments with a primitive-equation global model, PhD Thesis at the University of Maryland.

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Climate Dyn.*, **20**, 175-191.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415-428.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

Radakovich, J. D., P.R. Houser, A. M. da Silva, and M. G. Bosilovich, 2001: Results from global land-surface data assimilation methods. *Proceeding of the fifth symposium on integrated observing systems,* 14-19 January 2001, Albuquerque, NM. 132-134.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the NCEP global model. *Tellus*, **57A**, 528-545.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Scatterfield and J. A. Yorke, 2007: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, accepted.

Talagrand, O., 1999: a posteriori verification of analysis and assimilation algorithms. Proceeding of workshop on *Diagnosis of Data Assimilation Syst*em, November 1998, Reading, United Kingdom: ECMWF, 17-28

Tippett, M., J. Anderson, C.Bishop, T. Hamill, and J. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485-1490.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130,** 1913-1924.

Whitaker, J. S., G. P. Compo, X. Wei, and T.M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190-1200.

Whitaker, J. S., T. M. Hamill, X. Wei, Yucheng Song and Z. Toth, 2007: Ensemble data assimilation with the NCEP global forecasting system. *Mon. Wea. Rev.,* to appear.