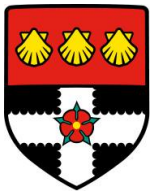


UNIVERSITY OF READING

School of Mathematical and Physical Sciences

Estimating Forecast Error Covariance Matrices with Ensembles

by

Emine Akkuş

August 2014

A dissertation is submitted to the Department of Mathematics in partial fulfilment of the requirements for the degree of Master of Science

Abstract

The problem of variational data assimilation for a nonlinear coupled atmosphere and ocean model is formulated as an optimization problem to find the best initial condition. The input data includes errors of observations and background. Therefore, the optimal solution involves error. The modelling of the background error covariance matrix is important in any data assimilation methods in the sense that it determines the spread of the errors. Therefore, in this study we try to estimate the forecast (background) error covariance matrix calculated by the idea of ensemble Kalman Filter (EnKF). To do this, we use a method which is an ensemble of four dimensional variational (4DVar) methods. We generate ensemble members by perturbing the background and observations with different random numbers. Then, we set up different ensembles and investigate how many ensemble members can make the forecast error covariance matrix convergence. We look at the convergence of each component of the matrix with the ensemble size getting bigger, and want to find a sufficient ensemble size which makes all components convergence. Then, we look at the effects of the error correlations of model variables and, the accuracy and frequency of observations on the convergence of forecast error covariance matrix.

Acknowledgements

I would like to thank my supervisor Dr Amos S. Lawless so much for being helpful and supportive during this dissertation. I would also like to thank all my lecturers in this MSc course especially to Professor Michael Baines and Dr Peter K. Sweby for their guidance and patience.

A special thanks to the God, my family, my fiance and my friends. Their supports provide me high motivation during this course.

Declaration

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

Emine Akkus

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In general, data assimilation (DA) combines all available information to estimate the state of a system. It uses observations and prior information, which is called the background, and combines them with a computer model to find the analysis which is an approximation to the observed reality. Lawless (2013) discusses the progress of data assimilation in more detail. Basically, it started being used in the 1940s and with the development of data assimilation techniques its usage has reached a wide spectrum of areas such as environmental modelling, climate monitoring, traffic modelling and numerical weather prediction (NWP). We are interested in its use in NWP. DA aims to provide the most appropriate initial condition for a forecast. In the assimilation window, we have an initial value problem. Thus to be able to make a good forecast, we need to choose a model which represents the atmosphere or the ocean or both as realistically as possible and to have an initial condition which should be defined as accurately as possible.

1.1 Motivation

For a long time, DA has been applied to the atmosphere and the ocean separately. On the other hand, recent studies show that coupling them gives numerically the most certain outcomes of climate change investigations (Sausen and Voss, 1995). To simulate long-term climatic variations, the coupled atmosphere and ocean models need to be used. In practice, there are some complications. For instance, it needs a lot of computational work because of the big difference between the scales of atmosphere and ocean. The ocean is much slower than the atmosphere because of the thermal inertia of the ocean (Dubois et al., 1999).

Variational methods solve the assimilation by minimizing a function with the ability of using future observations. These methods involve a background error covariance matrix but they do not provide any information about this matrix. Therefore, this matrix remains constant for different assimilations. Ensemble methods, which are sequential so all things in sequence, provide information about background error covariance matrix. It is expected that DA methods should consider 'errors of the day' (Cheng et al., 2010). This can be implemented by using updated background error covariance matrix in every assimilation window. The idea of hybrid methods is to take the best features of variational and ensemble methods. These methods use updated background information to capture the system dynamics and find a variational solution using future observations. Most hybrid methods are based on the combination of ensemble Kalman Filter (EnKF) and incremental 4DVar, for example in the cases of Fairbairn et al. (2014) and Liu et al. (2008).

Estimating the forecast error statistics is crucial to make more accurate

assimilation, because those statistics become the background error statistics for the next assimilation window.

In this project, the behaviour of the Molteni et al. (1993) coupled atmosphere and ocean model will be investigated with the ensembles of non-incremental 4DVar methods. The ensemble members are generated by using the idea of Isaksen et al. (2010). We perturb the background and observations with different random numbers to get a new ensemble member. Then, we try to estimate the forecast error covariance matrix $\mathbf{P}_f$ generated by using the idea of EnKF. We use different ensembles to be able to estimate $\mathbf{P}_f$. In this project, the research questions are:

(1) How many ensemble members are needed to capture the forecast covariances correctly?

(2) What are the effects of error correlations on the estimation of forecast covariances?

(3) How these are affected by observation errors and frequencies?

To investigate these questions, first we try to estimate $\mathbf{P}_f$ with different ensemble sizes then try to understand how many ensemble members make the matrix $\mathbf{P}_f$ convergence. The convergence of each component of $\mathbf{P}_f$ is investigated separately with the ensemble size getting bigger. Then, we look at the effects of error correlations and the effects of the accuracies and the numbers of observations on the convergence of $\mathbf{P}_f$.

1.2 Outline

This project is divided into six main chapters. In Chapter 2, we introduce the 4DVar method in mathematical detail with the use of tangent linear and adjoint model. The 4DVar is the foundation of the experiments. Then, as a sequential method the Kalman Filter algorithm is presented to show the formulation of analysis error covariance matrix $\mathbf{P}_a$ and forecast error covariance matrix $\mathbf{P}_f$. The idea of ensemble Kalman Filter (EnKF) is adopted to calculate the ensemble based flow dependent background error covariance matrix $\mathbf{P}_f$. Then, we describe the process of Isaksen et al. (2010) for perturbing a system to generate different ensemble members in the hybrid methods section.

Chapter 3 gives the system of Molteni et al.(1993) coupled model, which we use as the toy model in our experiments, and the second order Runge-Kutta method, which we use to discretize this coupled model.

In Chapter 4, we describe the methodology we follow step by step. It starts with generating truth state and continues with generating background and observations, which we assimilate with 4DVar to find the best fit initial state. Then, we perturb the system by using the idea of Isaksen et al. (2010) to get different ensemble members. Thus, we have an ensemble of analyses at the initial time. Then, we forecast each of them to get a forecast error covariance matrix $\mathbf{P}_f$.

Chapter 5 demonstrates our experiments and their results. In this chapter, we try to investigate our research questions described in Section (1.1).

In Chapter 6, we give a brief summary with conclusions of this project and propose some ideas for future work.

Chapter 2

Background

There are two types of data assimilation methods which are sequential and variational. Sequential data assimilation only considers observations made in the past until the time of analysis, while the variational one can use observations from the future. It can be seen from Figure (2.1)(b) that variational data assimilation methods produce a continuous analysis trajectory (the black line) by using the background trajectory (the red line with the initial value $\mathbf{x}_0^b$) and the observations $\mathbf{y}_i$, for $i = 1, ..., k$ where k is the number of observations over the assimilation window $[t_0, t_N]$. This gives the best initial analysis state $\mathbf{x}_0^a$ which minimizes the cost function $\mathcal{J}$, Section (2.1.1). In variational assimilation, the correction to the analysed state is smoother than the sequential one in time, which is more realistic in a physical sense (Bouttier et al., 2002). The sequential assimilation process can be seen from Figure (2.1)(a). There are forecast and analysis steps denoted by $\mathbf{x}_i^b$ and $\mathbf{x}_i^a$ respectively and observations $\mathbf{y}_i$ for $i = 1, ..., k$ where k is the number of analysis steps ($\mathbf{x}_0^b$ is the initial background state). For examples of the sequential and variational assimilations, we can refer to the Kalman Filter (Kalman, 1960) and 4DVar (Le Dimet and Talagrand, 1986) respectively.

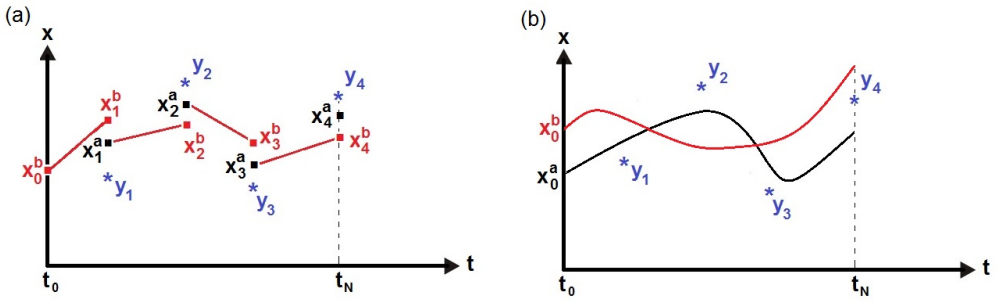


Figure 2.1: The processes of sequential (a) and variational (b) data assimilation methods over the assimilation window with the range $[t_0, t_N]$.

Nowadays, the use of ensemble methods has a significant part in NWP. Operational centres have started to seek different methods that consist of an ensemble of data assimilation techniques. Many of these methods are generated from the theory of the Kalman Filter (Kalman, 1960). Isaksen et al. (2010), described in Section (2.1.4), shows a way of producing ensemble data assimilation methods. We can also refer the Ensemble Kalman Filter (EnKF) method as an example of sequential ensemble data assimilation method described in Section (2.1.3).

4DVar has been used for 20 years at the most advanced centre ECMWF, while the EnKF is an upcoming method. They are competitive in skill. Hence, combining them may work best. Therefore, the idea of hybrid techniques comes to the surface. These data assimilation techniques are a mixture of ensemble and variational methods. As a result, operational centres are now exploring hybrid methods. Most of them are produced by a combination of the 4DVar and the EnKF and called four dimensional ensemble variational methods (its abbreviation shows some differences in the DA literature). For example, Liu et al. (2008) combine the incremental 4DVar with the EnKF

by trying to put their advantages together and call the method En4DVAR. It uses flow-dependent background error covariance matrix $\mathbf{P}$ from EnKF and tries to find variational solution likewise 4DVar. This method does not use tangent linear and adjoint models to produce an analysis. In their experiments, the method produces an analysis result which is similar to the analysis produced with tangent linear and adjoint models, which need high computational work. Fairbairn et al. (2014) call the same method the 4DEnVar. From now on we use the abbreviation 4DEnVar for this method not to cause any confusion. By taking inspiration from the idea of 4DEnVar, in this study we use the ensemble of non-incremental 4DVar methods and produce the forecast error covariance matrix.

2.1 Assimilation Methods

2.1.1 The Four-Dimensional Variational Data Assimilation (4DVar)

4DVar method produced by Le Dimet and Talagrand (1986) is an extension of 3DVar with the consideration of observations in time. After the 2000s, the method has the most use in operational centres (Fairbairn et al., 2013). This method helps us get the best trajectory of the system by running the numerical model over a time interval $[t_0, t_N]$ and putting the background and observations into the model. Figure (2.1)(b) demonstrates the process in a general diagram. Lawless (2013) and Nichols (2010) describe the method in more detail. We present here a basic idea of the method. Assume we have a background $\mathbf{x}_0^b$ at time t_0

$$\mathbf{x}_0^b = \mathbf{x}_0 + \boldsymbol{\epsilon} \tag{2.1}$$

where $\mathbf{x}_0$ is the state vector of the system at time t_0 and $\boldsymbol{\epsilon}$ is a vector of random unbiased Gaussian errors with known covariance matrix $\mathbf{B}$ and we have observations $\mathbf{y}_k$ at time t_k that satisfy

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\eta}_k \tag{2.2}$$

where $\mathbf{x}_k$ is the state vector of the system at time t_k for $k = 0, ..., N$, $\mathcal{H}$ is the observation operator which maps the state vector to the observation space and $\boldsymbol{\eta}_k$ are random unbiased Gaussian errors with known covariance matrices $\mathbf{R}_k$.

The purpose of this method is to find the best analysis state at the initial time, which minimizes the cost function $\mathcal{J}$.

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{k=0}^{N}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T \mathbf{R}_k^{-1}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k) \tag{2.3}$$

subject to the nonlinear model $\mathcal{M}$

$$\mathbf{x}_k = \mathcal{M}_{k-1}(\mathbf{x}_{k-1}) \tag{2.4}$$

where $\mathbf{x}_k$ is the state vector of the system at time t_k for $k = 0, ..., N$ and $\mathcal{M}_{k-1}$ gives the evolution of the states from $\mathbf{x}_{k-1}$ to $\mathbf{x}_k$.

4DVar problem is a least square problem whose solution minimizes the sum of the squares of the errors come from background and observations. The method uses the same background error covariance matrix $\mathbf{B}$ for each cycle of the assimilation. The method considers that the model $\mathcal{M}$ is perfect. This means that there is no error in the model (In the case of imperfect model,

the right hand side (RHS) of Equation (2.4) has an error term). Equation (2.4) can also be written as

$$\begin{aligned} \mathbf{x}_k &= \mathcal{M}_{k-1}\mathcal{M}_{k-2}...\mathcal{M}_0(\mathbf{x}_0) && (2.5) \\ &= \mathcal{M}_{0\to k-1}(\mathbf{x}_0) && (2.6) \end{aligned}$$

Therefore, 4DVar is a nonlinear constrained optimization problem so in general it is hard to solve (Bouttier et al., 2002). To deal with this problem, the tangent linear hypothesis is developed. Under the hypothesis, we assume that the observation operator $\mathcal{H}$ can be made approximately linear as follows

$$\mathcal{H}_0(\mathbf{x}_0^b) - \mathcal{H}_0(\mathbf{x}_0) \approx \mathbf{H}_0(\mathbf{x}_0^b)(\mathbf{x}_0^b - \mathbf{x}_0) \quad (2.7)$$

where $\mathbf{H}_0(\mathbf{x}_0^b)$ is the differential of $\mathcal{H}_0$ at $\mathbf{x}_0^b$.

In addition, under this hypothesis it is assumed that the model operator $\mathcal{M}$ can be linearised as

$$\mathbf{y}_k - \mathcal{H}_k\mathcal{M}_{0\to k-1}(\mathbf{x}_0) \approx \mathbf{y}_k - \mathcal{H}_k\mathcal{M}_{0\to k-1}(\mathbf{x}_0^b) - \mathbf{H}_k\mathbf{M}_{0\to k-1}(\mathbf{x}_0 - \mathbf{x}_0^b) \quad (2.8)$$

where $\mathbf{M}_{0\to k-1}$ is the differential of $\mathcal{M}_{0\to k-1}$ and known as the tangent linear model (TLM) in the DA literature. Then, the gradient of $\mathcal{J}$ is

$$\nabla\mathcal{J}(\mathbf{x}_0) = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) - \sum_{k=0}^{N}(\mathbf{M}_{k-1}\mathbf{M}_{k-2}...\mathbf{M}_0)^T\mathbf{H}_k^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k) \quad (2.9)$$

where $(\mathbf{M}_{k-1}\mathbf{M}_{k-2}...\mathbf{M}_0)^T = (\mathbf{M}_0)^T...(\mathbf{M}_{k-2})^T(\mathbf{M}_{k-1})^T$. The transpose of TLM is called the adjoint model. The adjoint models $\mathbf{M}_k^T$ represent the gradient of the cost function with respect to the model variables. Then, the

analysis $\mathbf{x}_0^a$ satisfies $\nabla \mathcal{J}(\mathbf{x}_0^a) = 0$ at the initial time. The aim of 4DVar is to find $\mathbf{x}_0^a$ by minimizing the cost function, $\mathcal{J}$.

2.1.2 The Kalman Filter (KF)

The KF method was produced by Kalman (1960) as a data assimilation (DA) method to estimate the current state. It is a sequential method and its process is explained in Figure (2.1)(a). Basically, the filter works just for linear systems. Jazwinski (1970) discusses the filter in detail. We just present here a brief explanation of how the KF works. In the KF, the system have a forecast step $\mathbf{x}_k^f$ with a forecast error covariance matrix $\mathbf{P}_k^f$ and an analysis step $\mathbf{x}_k^a$ with an analysis error covariance matrix $\mathbf{P}_k^a$ at time t_k. The state forecasts are explained as follows

$$\mathbf{x}_k^f = \mathbf{M}_{k-1}\mathbf{x}_{k-1}^a \tag{2.10}$$

where $\mathbf{M}$ is a linear model. Then, the forecast error covariance matrices

$$\mathbf{P}_k^f = \mathbf{M}_{k-1}\mathbf{P}_{k-1}^a\mathbf{M}_{k-1}^T + \mathbf{Q}_{k-1} \tag{2.11}$$

where $\mathbf{Q}$ represents the model error. In the case of a perfect model, $\mathbf{Q}$ can be neglected. Then, the analysis steps

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^f) \tag{2.12}$$

and the analysis error covariance matrices

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f \tag{2.13}$$

where $\mathbf{K}_k$ are the Kalman gain computations given by

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{2.14}$$

where $\mathbf{H}_k$ are the linear observation operators and $\mathbf{R}_k$ are the observation covariance matrices. As we mention before, this method is for linear systems. The extended Kalman Filter (EKF) applies KF to nonlinear systems by using tangent linear and adjoint models.

2.1.3 The Ensemble Kalman Filter (EnKF)

The EnKF presented by Evensen (1994) is an ensemble method and approximates the EKF for large systems. We do not give the whole algorithm of the method, only we give here the part of producing the ensemble error covariance matrix. Assume, we have m ensemble members which are denoted $\mathbf{x}_i$ for $i = 1, ..., m$, where each $\mathbf{x}_i$ is a state vector of the system, then the mean of ensemble members is

$$\bar{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i \tag{2.15}$$

The ensemble perturbation matrix $\mathbf{X}$ can be derived by

$$\mathbf{X} = \frac{1}{\sqrt{m-1}} (\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_2 - \bar{\mathbf{x}}, ..., \mathbf{x}_m - \bar{\mathbf{x}}) \tag{2.16}$$

and the ensemble covariance matrix $\mathbf{P}$ is given by

$$\mathbf{P} = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \tag{2.17}$$

which can be also written in terms of the ensemble perturbation matrix in Equation(2.16) as follows

$$\mathbf{P} = \mathbf{X}\mathbf{X}^T \tag{2.18}$$

The ensemble covariance matrix $\mathbf{P}$ is flow-dependent and using it as the background error covariance matrix makes the system more realistic. $\mathbf{P}$ is a symmetric and square matrix with the size of $n \times n$ where n is the size of the state vectors $\mathbf{x}_i$. Hence, the form of $\mathbf{P}$ is

$$\mathbf{P} = \begin{pmatrix} Var(e_1) & Cov(e_1, e_2) & \cdots & Cov(e_1, e_n) \\ Cov(e_2, e_1) & Var(e_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & Cov(e_{n-1}, e_n) \\ Cov(e_n, e_1) & \cdots & Cov(e_n, e_{n-1}) & Var(e_n) \end{pmatrix} \tag{2.19}$$

where e_i for $i = 1, ..., n$ denote each variable of the system which have n variables.

In the EnKF, the smoothness of analysis depends highly on the ensemble size m. It should be sufficiently large. If it is not, this leads $\mathbf{P}$ to be low rank. To be able to make the system fully observable, the background error covariance matrix should be full rank. To get rid of the problem caused by low-rankness of $\mathbf{P}$, besides choosing sufficient ensemble size, localization and inflation techniques can be alternative solutions. The localization technique was proposed by Hamill et al. (2001) and, Anderson and Anderson (1999) introduce the idea of inflation technique. However, we do not cover these techniques in this project.

2.1.4 Hybrid Methods

Hybrid methods are combinations of variational and ensemble methods. They try to apply the best features of both methods. These methods solve variational problems with the flow-dependent background error covariance matrix $\mathbf{P}$ instead of fixed background error covariance matrix $\mathbf{B}$. Ideally, the background error covariance matrix should depend on the current flow. For example, consider the pressure in the real atmosphere, sometimes high pressure takes the place of low pressure quickly. This kind of alterations should effect the background error covariance matrix. Additionally, variational solutions use observations more efficiently than sequential ones, because sequential ones only use observations in the past, while variational ones can also use observations from the future. Therefore, hybrid methods are created. Most of them are produced by a combination of the EnKF and the incremental 4DVar and called Four-Dimensional Ensemble Variational DA referred as 4DEnVar in literature. In this project, we use an ensemble of non-incremental 4DVar methods producing ensemble members by perturbing the background and observations. We do these perturbations by choosing different random numbers for each ensemble members similar to the idea of Isaksen et al. (2010). They have perturbed observations by adding random noise from the probability density function (pdf) of observation error, and added further perturbations to account for model error during the forecasts. There is no direct perturbations applied to the background. The Isaksen et al. (2010) system can be described as follows:
Consider the following linear system:

$$\begin{aligned} \mathbf{x}_k^a &= \mathbf{x}_k^b + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^b) \\ \mathbf{x}_{k+1}^b &= \mathbf{M}_k(\mathbf{x}_k^a) \end{aligned} \tag{2.20}$$

where k denotes analysis cycle, $\mathbf{y}_k$ is the vector of observations, $\mathbf{x}_k^a$ is the analysed state, $\mathbf{x}_k^b$, $\mathbf{K}_k$ and $\mathbf{M}_k$ are matrices, and $\mathbf{K}_k$ is a general gain matrix (not specifically the Kalman gain). Then the covariance matrices for this system are

$$\begin{aligned} \mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^b(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\ \mathbf{P}_{k+1}^b &= \mathbf{M}_k\mathbf{P}_a^k\mathbf{M}_k^T + \mathbf{Q}_k \end{aligned} \tag{2.21}$$

where $\mathbf{R}_k$ is the observation error covariance matrix and $\mathbf{Q}_k$ is the model error covariance matrix.

Then we perturb the system (2.20) as follows (denoted by $\sim$)

$$\begin{aligned} \tilde{\mathbf{x}}_k^a &= \tilde{\mathbf{x}}_k^b + \mathbf{K}_k(\mathbf{y}_k + \boldsymbol{\eta}_k - \mathbf{H}_k\tilde{\mathbf{x}}_k^b) \\ \tilde{\mathbf{x}}_{k+1}^b &= \mathbf{M}_k(\tilde{\mathbf{x}}_k^a) + \boldsymbol{\xi}_k \end{aligned} \tag{2.22}$$

where $\boldsymbol{\eta}_k$ and $\boldsymbol{\xi}_k$ are perturbations with covariance matrices $\mathbf{R}_k$ and $\mathbf{Q}_k$ respectively.

Then subtracting the perturbed (2.22) and unperturbed (2.20) systems as follows:

$$\begin{aligned} \boldsymbol{\epsilon}_k^a &= \boldsymbol{\epsilon}_k^b + \mathbf{K}_k(\boldsymbol{\eta}_k - \mathbf{H}_k\boldsymbol{\epsilon}_k^b) \\ \boldsymbol{\epsilon}_{k+1}^b &= \mathbf{M}_k\boldsymbol{\epsilon}_k^a + \boldsymbol{\xi}_k \end{aligned} \tag{2.23}$$

where $\boldsymbol{\epsilon}_k^a = \tilde{\mathbf{x}}_k^a - \mathbf{x}_k^a$ and $\boldsymbol{\epsilon}_k^b = \tilde{\mathbf{x}}_k^b - \mathbf{x}_k^b$. Then, from the system (2.23) the covariance matrices can be formed as

$$\begin{aligned} \overline{\boldsymbol{\epsilon}_a^k(\boldsymbol{\epsilon}_a^k)^T} &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\overline{\boldsymbol{\epsilon}_b^k(\boldsymbol{\epsilon}_b^k)^T}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\ \overline{\boldsymbol{\epsilon}_b^{k+1}(\boldsymbol{\epsilon}_b^{k+1})^T} &= \mathbf{M}_k\overline{\boldsymbol{\epsilon}_a^k(\boldsymbol{\epsilon}_a^k)^T}\mathbf{M}_k^T + \mathbf{Q}_k \end{aligned} \tag{2.24}$$

It can be seen from the comparison of the systems (2.21) and (2.24), if $\overline{\boldsymbol{\epsilon}_b^k(\boldsymbol{\epsilon}_b^k)^T} = \mathbf{P}_k^b$ for some k, then $\overline{\boldsymbol{\epsilon}_a^m(\boldsymbol{\epsilon}_a^m)^T} = \mathbf{P}_m^a$ and $\overline{\boldsymbol{\epsilon}_b^m(\boldsymbol{\epsilon}_b^m)^T} = \mathbf{P}_m^b$ for all $m \geq k$. This means that the analyses and backgrounds perturbations have equal covariances to the corresponding analysis and background error covariances for all subsequent analysis cycle. In this paper, they also provide another perturbed system by applying identical perturbations to observations and model error. They only choose the initial perturbation to be different.

Chapter 3

Models

3.1 The Molteni coupled model

As the toy model, we use the coupled atmosphere-ocean model described by Molteni et al. (1993) which couples the chaotic Lorenz system (1963) with the linear oscillatory system. The Lorenz system has three variables X,$Y$and Z which represent the atmosphere, while the linear part has two variables W and V which are the representation of the ocean. The system is coupled by an arbitrary coupling parameter α. It is applied through the X, Y, W and V variables. Molteni et al. (1993) have used the coupled model to examine the interaction of tropical-midlatitute. They have conducted some experiments by setting tropical forcing term W^* variously. The coupled system is also used by Dubois and Yiou (1999). Additionally, they have coupled the chaotic Lorenz system with the chaotic Rössler system. Molteni et al. (1993) system

is the following:

$$\begin{cases} \dot{X} = -\sigma X + \sigma Y + \alpha V \\ \dot{Y} = -XZ + \rho X - Y + \alpha W \\ \dot{Z} = XY - \beta Z \\ \dot{W} = -\Omega V - k(W - W^*) - \alpha Y \\ \dot{V} = \Omega(W - W^*) - kV - \alpha X \end{cases} \tag{3.1}$$

where $\sigma = 10$, $\beta = 8/3$, $\rho = 28$, $\Omega = 2\pi/20$, $k = 0.1$ and $W^* = 2$. The parameters σ, β and ρ are the Lorenz (1963) parameters (see Lorenz (1963) for their meaning) and Ω, k and W^* are the parameters of linear oscillatory system. Molteni et al. (1993) describe these parameters in detail.

In coupled models, experiments can be performed by coupling synchronously or asynchronously. Synchronously coupling means that we substitute all values into the related part as soon as they are calculated. Therefore, at each time step the coupled system always uses updated values. However, asynchronously coupling can be made with different frequencies. For example, if the coupling frequency is 2 then α terms are still applied on every time step but only changing values of other variables in term every 2 steps. Dubois and Yiou (1999) have done synchronous coupling (the coupling frequency is one), and asynchronous coupling with different frequencies on the Molteni et al.(1993) model and Rössler (1976) model. We can refer this paper to compare the results of different coupling frequencies on the related system.

3.2 Second order Runge-Kutta Method (RK2)

In general, Runge-Kutta Methods are used to solve initial value problems (IVPs) which we consider the first order ordinary differential equation (ODE)

system

$$y'(t) = f(t, y(t)) \tag{3.2}$$

with an initial condition $y(0) = y_0$. By doing Taylor expansion, the RK2 method can be derived as

$$y_{n+1} = y_n + \frac{1}{2}h(k_1 + k_2) \tag{3.3}$$

with

$$k_1 = f(t_n, y_n) \tag{3.4}$$

$$k_2 = f(t_n + h, y_n + hk_1) \tag{3.5}$$

where n represent the time step and h is the step size. The toy model is discretized by using RK2. Lawless (2006) discretize the Lorenz (1963) model with the method RK2.

Chapter 4

Experimental Setup

In this project, we use identical twin experiments so we do data assimilation (DA) pretending we do not know the truth state. In our experiments, we investigate our research questions described in Section (1.1). We want to estimate the forecast error covariance matrix $\mathbf{P}_f$ with different ensemble sizes. We look at the convergence of each component of $\mathbf{P}_f$ with the ensemble size getting bigger. Then, we examine the effects of error correlations of model variables on the convergence of $\mathbf{P}_f$. Finally, we examine the effects of the accuracies and the numbers of observations on the convergence of $\mathbf{P}_f$. It is important to estimate the forecast error covariance matrix in any DA system, because it will be used as a background error covariance matrix for the next assimilation. Better estimation we have better analysis for the next assimilation cycle. In order to estimate this matrix, first we generate a method which is an ensemble of 4DVar methods producing ensemble members by using Isaksen et al. (2010) method described in Section (2.1.4). Thus, we produce each ensemble members by perturbing background and observations with different random numbers. Each random numbers for background and observations are generated by the corresponding error covariance matrices $\mathbf{B}$

and $\mathbf{R}$ respectively ($\mathbf{B}$ is the background error covariance matrix while $\mathbf{R}$ is the observation error covariance matrix). This method gives an ensemble of analysis states at the initial time of the assimilation window and we forecast each analysis state and find the forecast error covariance matrix calculated by the idea of ensemble Kalman Filter (EnKF) described in Section (2.1.3).

We choose Molteni et al. (1993) coupled model described in Section (3.1.1) as our toy model. Therefore, the size of state vectors $\mathbf{x}_k$ are 5 consisting of model variables X, Y, Z, W and V. Subscript k shows the time step over the assimilation window. We transfer the continuous model into a discrete model by using the RK2 method. This numerical method which we use to generate the background, the truth and the analysis trajectories through the assimilation window, can be chosen arbitrarily. For example, Dubois and Yiou (1999) use a fourth order Runge-Kutta method (RK4) on coupled atmosphere ocean models. In this project, the Polack-Ribiere flavour of conjugate gradient method is used as the iteration method to minimize the cost function, $\mathcal{J}$. For this method, we set up the tolerance as 0.001 for the stopping criteria as follows

$$\frac{\|\nabla \mathcal{J}^i(\mathbf{x}_0)\|}{\|\nabla \mathcal{J}^0(\mathbf{x}_0)\|} < 0.001 \tag{4.1}$$

where $\nabla \mathcal{J}$ is the gradient of cost function (see Equation (2.9)) and superscript i denotes the iteration step. To avoid overwork, we set up the maximum number of iterations as 200. Even if the tolerance is bigger than 0.001, the iterations will stop at the number of 200.

In all experiments, we do synchronous coupling and so the coupling frequency is one. This means that for each time step we use always updated

values of each variable of the model. We only estimate the state not the coupling parameter α so it is fixed as 1. We set up the number of time steps as 20 and the step length as 0.05. Thus, the length of assimilation window becomes 1.

4.1 Generating the truth

The truth state vectors are denoted as $\mathbf{x}_k$ for $k = 0, ..., 20$. At the beginning, we set up $\mathbf{x}_0$ as the column vector of $[1, 1, 1, 1.1, 1.1]$ to determine the initial truth state where each component of $\mathbf{x}_0$ represents the initial values of model variables X, Y, Z, W and V respectively. This truth state is identical in all experiments. Then, we run the RK2 method from $\mathbf{x}_0$ to generate the truth trajectory over the assimilation window.

4.2 The 4DVar Algorithm

4.2.1 Generating the background

The background state vectors are denoted as $\mathbf{x}_k^b$ for $k = 0, ..., 20$. The initial background state $\mathbf{x}_0^b$ has errors and we assume that errors are uncorrelated so the background error covariance matrix $\mathbf{B}$ becomes diagonal and its form is

$$\mathbf{B} = \begin{pmatrix} (\sigma_b^X)^2 & & & & \\ & (\sigma_b^Y)^2 & & 0 & \\ & & (\sigma_b^Z)^2 & & \\ & 0 & & (\sigma_b^W)^2 & \\ & & & & (\sigma_b^V)^2 \end{pmatrix} \tag{4.2}$$

where $(\sigma_b^X)^2, (\sigma_b^Y)^2, (\sigma_b^Z)^2, (\sigma_b^W)^2$ and $(\sigma_b^V)^2$ are the background variances of the model variables X, Y, Z, W and V respectively.

Then, from Section (2.1.1) the initial background state $\mathbf{x}_0^b$ is generated by adding random noise to the truth state $\mathbf{x}_0$ at time t_0. Different random noises are applied on each element of $\mathbf{x}_0$. Then, Equation (2.1) becomes

$$\mathbf{x}_0^b = \mathbf{x}_0 + \boldsymbol{\sigma}_b \boldsymbol{\epsilon} \tag{4.3}$$

where $\boldsymbol{\epsilon}$ is a vector of random unbiased Gaussian errors and $\boldsymbol{\sigma}_b$ is the 5×5 diagonal matrix whose diagonal terms are the standard deviations which are the square roots of background variances of X, Y, Z, W and V. The matrix $\boldsymbol{\sigma}_b$ allows us to specify how accurate the background is. If its diagonal elements are too small, the additional term in the RHS of Equation (4.3) will become too small and so the background will be quite close to the truth. Finally, we run the RK2 method from $\mathbf{x}_0^b$ to generate the background trajectory over the assimilation window.

4.2.2 Generating observations

The vectors of observations are denoted as $\mathbf{y}_k$ for $k = 1, ..., 20$ (no observation at time t_0). We assume all observations have errors and the errors between observations at each time are uncorrelated. Thus, the observation error covariance matrix $\mathbf{R}$ becomes diagonal and its form is

$$\mathbf{R} = \begin{pmatrix} (\sigma_o^X)^2 & & & & \\ & (\sigma_o^Y)^2 & & 0 & \\ & & (\sigma_o^Z)^2 & & \\ & 0 & & (\sigma_o^W)^2 & \\ & & & & (\sigma_o^V)^2 \end{pmatrix} \tag{4.4}$$

where $(\sigma_o^X)^2, (\sigma_o^Y)^2, (\sigma_o^Z)^2, (\sigma_o^W)^2$ and $(\sigma_o^V)^2$ are the observation variances of the model variables X, Y, Z, W and V respectively.

Then, from Section (2.1.1) the observations $\mathbf{y}_k$ at time t_k are generated by adding random noises to the truth state $\mathbf{x}_k$ at the same time. Then, Equation (2.2) becomes

$$\mathbf{y}_k = \mathbf{x_k} + \boldsymbol{\sigma_o}\boldsymbol{\eta_k} \tag{4.5}$$

where $\boldsymbol{\eta}_k$ is the vector of random unbiased Gaussian errors at time t_k and $\boldsymbol{\sigma}_o$ is the 5×5 diagonal matrix whose diagonal terms are the standard deviations which are the square roots of observation variances of X, Y, Z, W and V. The matrix $\boldsymbol{\sigma}_o$ is identical for each time step.

The states of observations depend on the observation frequency and accordingly the number of time step over the assimilation window. For example, assume the observation frequency is a vector of $[2, 2, 2, 4, 4]$ where each element represents observation frequencies of the model variables X, Y, Z, W and V respectively, and assume there are 20 time steps on the assimilation window. This means that we have observations every two time steps for each variable of the atmosphere (X, Y, Z) and every four time steps for each variable of the ocean (W, V). Hence, this example determines X, Y and Z

have 10 observations separately, in total 30 observations in the atmosphere, and W and V have 5 observations separately, in total 10 observations in the ocean. In our case, we assume that there is no observation at the initial time.

4.2.3 Determining the initial analysis state

We perform the assimilation by using the method 4DVar to find the best analysis trajectory whose initial state $\mathbf{x}_0^a$ minimizes the cost function $\mathcal{J}$ (see Equation (2.3)). We use Polack-Ribiere flavour of conjugate gradients as the iteration method to minimize the cost function $\mathcal{J}$. We use Equation (4.1) as the stopping criteria with the maximum number of iterations 200. Finally, we run the RK2 method from the generated initial analysis state $\mathbf{x}_0^a$ to make a forecast. The effectiveness of the assimilation can be measured by how the analysis is far from the truth. It gives the analysis error from the norm of their difference, $\|\mathbf{x}_0 - \mathbf{x}_0^a\|$.

4.3 Generating ensemble of 4DVar

The 4DVar assimilation process is being made several times by perturbing the background and observations with different random numbers. The idea of Isaksen et al. (2010) in Section (2.1.4) is used to produce each ensemble members. In this project, m denotes the number of ensemble members.

4.3.1 Ensemble of backgrounds

The initial background states of each ensemble members are generated from Equation (4.3). Each background state uses the same diagonal background

error covariance matrix $\mathbf{B}$. We use different vectors of errors denoted $\boldsymbol{\epsilon}^i$ to generate each background ensemble member $\mathbf{x}_0^{b,i}$ where i denotes the ensemble members from 1 to m. Thus, we get certain number of background states at the initial time. It is important to note that each analysis of ensemble members uses the corresponding background state $\mathbf{x}_0^{b,i}$.

4.3.2 Perturbed observations

The observations are generated from Equation (4.5) by choosing different $\boldsymbol{\eta}_k^i$ for each observation $\mathbf{y}_k^i$ for $i = 1, ..., m$. Isaksen et al. (2010) have done this process by adding perturbations to the previous observation. Then, we have certain observations for each background state. Note that each ensemble member has the same number of observations with identical observation frequency.

4.3.3 Generating ensemble of analysis states

We have the initial backgrounds with corresponding observations. Finally, we run the 4DVar for each ensemble member to generate their initial analysis states. Thus, we have m analysis states at the initial time. Then, we generate their trajectories with the RK2 method over the assimilation window. Figure (4.1) is just an example of the evolutions of 900 analysis ensemble members with the truth state on the time length 4. This figure is plotted by our program used in the experiments.

4.3.4 The ensemble based error covariance matrix

The analysis error covariance matrix $\mathbf{P}_a$ at the initial time and the forecast error covariance matrix $\mathbf{P}_f$ at the final time of the assimilation window can

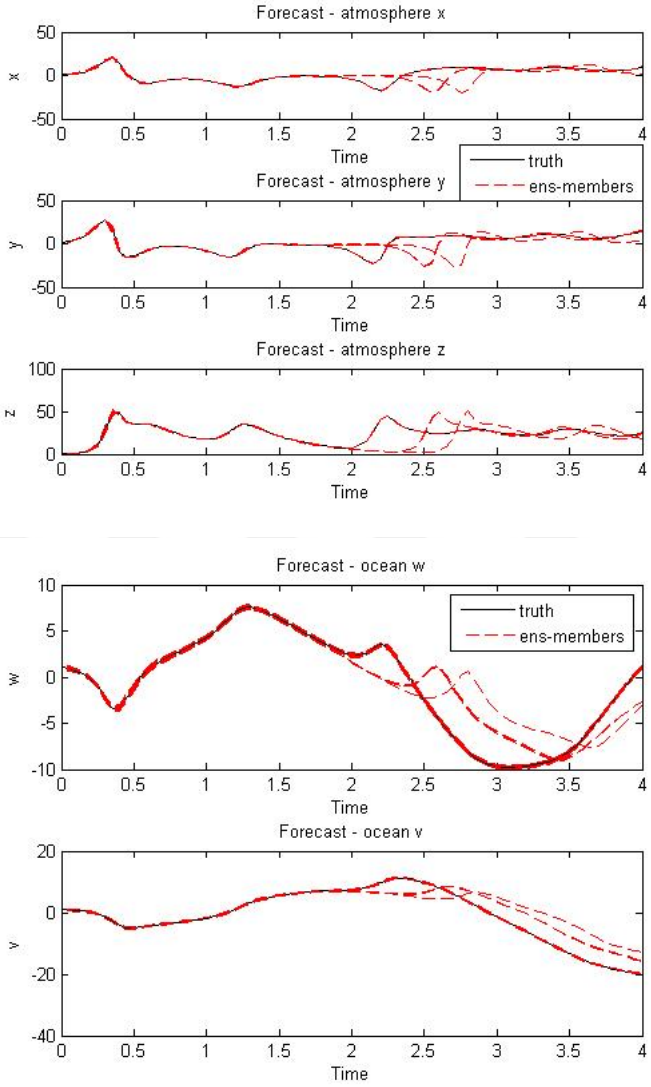


Figure 4.1: The evolutions of 900 ensemble members for each model variable

be derived from Section (2.1.3). The matrix $\mathbf{P}_a$ helps us to understand the errors spread of model variables before doing forecast. Then, the matrix $\mathbf{P}_f$ shows the effects of the model on the errors during forecast (consider Equation (2.11) without the error term $\mathbf{Q}_{k-1}$, because we assume the model is perfect). In our experiments, we try to estimate only the matrix $\mathbf{P}_f$. In the model, the error covariance matrix form in Equation (2.19) becomes

$$\mathbf{P} = \begin{pmatrix} var(X) & cov(X,Y) & cov(X,Z) & cov(X,W) & cov(X,V) \\ cov(Y,X) & var(Y) & cov(Y,Z) & cov(Y,W) & cov(Y,V) \\ cov(Z,X) & cov(Z,Y) & var(Z) & cov(Z,W) & cov(Z,V) \\ cov(W,X) & cov(W,Y) & cov(W,Z) & var(W) & cov(W,V) \\ cov(V,X) & cov(V,Y) & cov(V,Z) & cov(V,W) & var(V) \end{pmatrix} \tag{4.6}$$

To understand the relationship between errors of the model variables, we get the correlation matrix of $\mathbf{P}$. The off-diagonal terms of the matrix $\mathbf{P}$ can be transform into error correlations:

$$\rho(e_i, e_j) = \frac{cov(e_i, e_j)}{\sqrt{var(e_i)var(e_j)}} \tag{4.7}$$

where $\rho(e_i, e_j)$ is the general form of a correlation matrix. In this study, we get the correlation matrix from the following process:
Assume $\mathbf{P}$ is a covariance matrix and $\mathbf{D}$ is the diagonal part of $\mathbf{P}$, then the correlation matrix $\mathbf{C}$ satisfies

$$\mathbf{P} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2} \tag{4.8}$$

and $\mathbf{C}$ is calculated directly from

$$\mathbf{C} = \mathbf{A}^{-1}\mathbf{P}\mathbf{A}^{-1} \tag{4.9}$$

where $\mathbf{A}$ is the square root of $\mathbf{D}$. The value of each component of $\mathbf{C}$ is in the range of $[-1, 1]$. The value 1 shows variables are fully correlated (while one variable is increasing the other also increases, and vice versa), while in the case of -1 the values are fully anti-correlated (while one variable is increasing, the other decreases, and vice versa). The closer the coefficient is to either -1 or 1, the stronger the correlation between the variables. When the value is getting close to 0, the variables are starting to become uncorrelated. The value 0 shows that the variables are independent and so they are totally uncorrelated.

Chapter 5

Assimilation experiments

In the following three experiments, we try to investigate our research questions which are described in Section (1.1). Therefore, we try to estimate the ensemble based forecast error covariance matrix $\mathbf{P}_f$ with different ensemble sizes. To be able to do this, we look at the convergence of $\mathbf{P}_f$ components separately as the ensemble size gets larger. Then, we investigate how they are affected by the error correlations of model variables and, the accuracy and the number of observations.

The stationary parameters are defined at the beginning of Chapter 4. In addition, the background error variances of the model variables X, Y, Z, W and V are also fixed as 10^{-2}. Thus, the backgrounds of model variables have same accuracy (the distance between background and truth state can show small differences because of the random noise, see Equation (4.2)). We generate our ensembles from size 6 to 1800. We define the minimum ensemble size experimentally. We get the $\mathbf{P}_f$ full rank ($\text{rank}(\mathbf{P}_f) = 5$) with the minimum ensemble size 6. Hence, all experiments start with 6 ensemble members. Then, the next ensemble size is set up as 100 and it carries on as

consecutive multiples of 100 up to 1800 ensemble members. Thus, we use 19 different ensemble sizes in each experiment. Then, we compare the values of $\mathbf{P}_f$ components after each ensemble. To be able to estimate the matrix $\mathbf{P}_f$, it is expected that each component of the matrix converges to a number. Therefore, we try to get understanding of how many ensemble members make the $\mathbf{P}_f$ matrix convergence. This can give approximate relationships between error spreads in the model. Choosing a sufficient and appropriate convergence condition provides a basis for our experiments.

Assume M and N are subsequent ensemble sizes where $M < N$. We define the relative convergence condition as

$$\frac{|e_M - e_N|}{|e_M|} < \lambda \tag{5.1}$$

where e denotes each component of $\mathbf{P}_f$ and λ called convergence condition limit is an arbitrary number discussed in the experiments. This relative condition seems appropriate for a convergence condition because some components of $\mathbf{P}_f$ can have values with different orders of magnitude. Their approximate values can be seen in Tables (5.2) and (5.5). In these tables, each presented value is the mean of its values from ensemble size 6 to 1800.

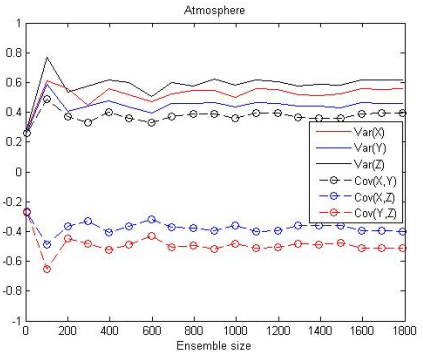
In this project, there are three experiments. In the first one, we set up the parameters arbitrarily. Then, with respect to these parameters we have an idea about the convergence of $\mathbf{P}_f$ as the ensemble size gets larger. Its results provide a basis for the parameters used in the other two experiments. We only change the accuracy and the frequency of observations in Experiment-2 and Experiment-3 respectively.

From now on when we say convergence behaviour of any component, it represents a broken line which combines each value calculated by Equation (5.1) from the ensemble size 6 to 1800. The closeness to the convergence condition limit λ and the smoothness of this line will be compared in our experiments.

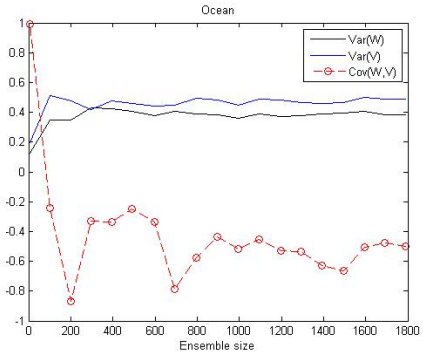
5.1 Experiment-1

In this experiment, we try to investigate our first and second research questions (see Section (1.1)). We set up the observation error variances of X, Y, Z as 10^{-4} and W, V as 10^{-2}. In other words, we use more accurate observations in the atmosphere than in the ocean according to the ranges of their error spreads. Additionally, the observation frequency is $[5, 5, 5, 5, 5]$ so we have observations in every five time steps for each model variable. We set up the number of time steps as 20. Hence, we have 4 observations for each variable X, Y, Z, W and V at times t_5, t_{10}, t_{15} and t_{20} (we assume there is no observation at the initial time). This means that there are total 20 observations in each assimilation period that 12 of them are in the atmosphere while 8 of them are in the ocean.

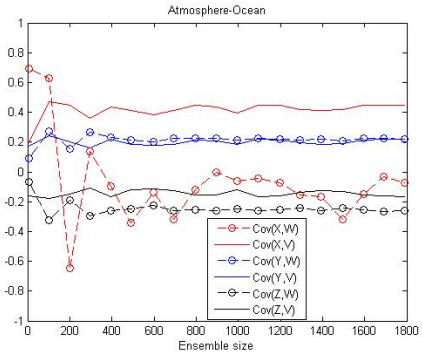
Figure (5.1) shows the relative values of $\mathbf{P}_f$ components with the ensemble sizes from 6 to 1800. The values of each component are plotted by multiplying with an appropriate one of the multiples of 10 to make its evolution in the range of $[-1, 1]$. For example, the values of $Var(X)$ in each ensemble are between 10^{-5} and 10^{-4}. We plot its values by multiplying 10^4. If the value range is between 10^{-5} and 10^{-3}, it will be plotted by multiplying



(a) Atmosphere part



(b) Ocean part



(c) Atmosphere-Ocean part

Figure 5.1: The relative values of each component of $\mathbf{P_f}$ for different ensemble sizes

10^3. This way is one of the clearest ways to see their behaviour in the same scale. Otherwise, it is hard to see their relative changes in different scales. Their approximate values can be seen in the Experiment-1 part of Table (5.2). In this table, the represented values are the average of the values of $\mathbf{P}_f$ components from 6 to 1800 ensemble members.

We can see from Figure (5.1), after 1000 ensemble members the most of $\mathbf{P}_f$ components are likely to converge. The significant fluctuations are seen on $Cov(W, V)$ (Figure (5.1)(b)) and $Cov(X, W)$ (Figure (5.1)(c)) with the ensemble size getting bigger. However, these figures are unsatisfactory in

the sense of the convergence condition in Equation (5.1). To be able to get a more clear idea of their convergences, we plot their relative changes with respect to Equation (5.1) with the ensemble size getting bigger. The results are seen in Figures (5.2), (5.3) and (5.4). Some peak values cannot be seen in these figures because we fixed the scale of figures between -0.5 and 1 to clarify their changes. On these figures, there are two dotted lines: red and green. We will discuss them later as a limit of convergence condition λ in Equation (5.1). As we expect from Figure (5.1), the remarkable jumps are seen on the relative changes of $Cov(W, V)$ (Figure (5.4)(d)) and $Cov(X, W)$ (Figure (5.3)(a)) with the ensemble size getting larger.

To have a general idea about how the error in a component is correlated with the error in another one, we calculate the correlation matrix of $\mathbf{P}_f$ represented in Section (4.3.4). Table (5.1) shows the average values of each different component of correlation matrix $\mathbf{C}$. We calculate their arithmetic means from the results of each ensemble size from 6 to 1800. We do not think plotting their values at each ensemble member necessary, because the values of each component is quite close to the its mean value. From this table, we can see there is almost no correlation between the errors of X and W, and similarly between the errors of W and V. On the other hand, the atmosphere itself has highly correlated errors.

If we consider the convergence behaviour of one component with its corresponding correlation mean value, we can see that they seem to be related. For example, consider the case of $Cov(X, W)$ in Figure (5.3)(a), it has the most spurious behaviour with the ensemble sizes and then we see in Table (5.1) that the errors of W and V have the lowest correlation (-0.0151). On

the other hand, the convergence behaviour of $Cov(X, V)$ in Figure (5.3)(b) has much smoother behaviour than $Cov(X, W)$. Then, we look at the error correlations of X and V in Table (5.1), they are almost fully correlated with the correlation value 0.8558. Therefore, we can conclude that the less correlation in errors the less convergence tendency in its convergence behaviour, and vice versa. From this view, we can expect to see the best convergence behaviour on the variances of X, Y, Z, W and V because they are fully correlated in their own right (each diagonal term of correlation matrix $\mathbf{C}$ is 1). This can determine the limit of convergence condition λ.

Now, we can focus on the limit of convergence condition λ. In the figures showing relative conditions of each model variable, there are two specified λs. The red dotted line is the line of 0.05 whereas the green one is that of 0.2. From Figure (5.2), to be able to make the variances of each component always under the green dotted line after a certain ensemble size (i.e the size of 300), we can define the limit of convergence condition as the line 0.2 (the green dotted line). Otherwise, on the condition of 0.05 (the red dotted line) we cannot generalize that these variances have always satisfied the convergence condition after a certain ensemble size. Thus, from now on we define the limit of convergence condition λ as 0.2.

The reason for significant jumps on the convergence behaviour of $Cov(X, W)$ can be because of the model system (see Equation (3.1)). There is no direct relationship between X and W variables. This may indicate that more ensemble members are needed to capture the information between their errors. From this view, we can expect similar jumps on the convergence behaviours of $Cov(Y, V)$, $Cov(Z, W)$ and $Cov(Z, V)$, because there are also no direct

Atmosphere	
Cor(X,Y)	0.7737
Cor(X,Z)	-0.6748
Cor(Y,Z)	-0.9596
Ocean	
Cor(W,V)	-0.1113

Atmosphere-Ocean	
Cor(X,W)	-0.0151
Cor(X,V)	0.8558
Cor(Y,W)	0.5304
Cor(Y,V)	0.4471
Cor(Z,W)	-0.5340
Cor(Z,V)	-0.2847

Table 5.1: Averages of each error correlation from the ensemble size 6 to 1800 in Experiment-1

relationships between Y and V, Z and W, and Z and V. However, their convergence behaviours are much smoother than that of $Cov(X, W)$ with the ensemble size getting bigger. This can be because Y and Z have non-linear formulas in the model system that can help the variables capture the information more quickly than linear ones.

In Table (5.1), the error correlations of (Y, V) , (Z, W) and (Z, V) are 0.4471, -0.5340 and -0.2847 respectively. From our hypothesis (the less correlation in errors the less convergence tendency in its convergence behaviour, and vice versa), we expect that $Cov(Z, W)$ have the smoothest convergence behaviour than $Cov(Y, V)$ and $Cov(Z, V)$ with respect to their correlation magnitudes. Figures (5.3)(d), (e) and (f) show that their behaviour agrees with our hypothesis. However, when we look at the ensemble size where the components satisfy the convergence condition first, $Cov(Z, W)$ converges with the highest ensemble size 300 among $Cov(Y, V)$, $Cov(Z, W)$ and $Cov(Z, V)$. It is important to note that interpreting the results by just considering a specific point in figures may not be realistic, because we do only one experiment with each ensemble size. The next experiment with exactly same parameters can show some differences because of the random

numbers used in generating ensemble members. Therefore, considering the convergence behaviours of the components of $\mathbf{P}_f$ from the ensemble size 6 to 1800 in figures can be more logical to compare the convergences of components.

As the ensemble size getting bigger from 6 to 1800, the convergence behaviour of $Cov(W, V)$ shows significant fluctuations especially up to the size 800 in Figure (5.1)(b). In addition, it can be seen in Table (5.1) that there is a small correlation in the errors of W and V (Cor(W,V)=-0.1113). These results can be from the parameters we set up. In this experiment, we have less accurate observations in the ocean (each variance of W and V is 10^{-2}) than in the atmosphere (each variance of X, Y and Z is 10^{-4}), and the number of observations in the ocean (8 observations) less than in the atmosphere (12 observations). In the following two experiments, Experiment-2 and Experiment-3 we will examine the effects of the accuracy of ocean observations and the number of ocean observations on the convergence of $\mathbf{P}_f$ respectively. The setup parameters used in this experiment will provide a basis for the parameters used in other two. In the figures, the results of each following experiment will be plotted with the results of this experiment to notice their differences clearly.

5.2 Experiment-2

In this experiment, we examine the effects of the accuracy of ocean observations on the convergence of $\mathbf{P}_f$, which is a part of our third research question (see Section (1.1)). Thus, we do similar experiments as in Experiment-1 with more accurate ocean observations. The only difference in the parameters set

up in the previous experiment is that the observation error variance of W and V becomes 10^{-4} from 10^{-2}. By doing so, we make the observation variances of all variables (X, Y, Z, W and V) the same. Table (5.2) shows the approximate value of each component of $\mathbf{P}_f$. This table is created by the mean of the values of each $\mathbf{P}_f$ component as the ensemble size gets bigger from 6 to 1800. We compare the results of Experiment-1 and 2 and see that more accurate ocean observations lead the errors to spread less. Remarkable decreases are seen in the ocean covariances and the cross-covariances between the atmosphere and ocean. However, the convergence property becomes more erratic in general that we can also expect from Table (5.3) according to our hypothesis. In other words, when we get more accurate ocean observations there are more fluctuations on the convergence behaviour of each component of $\mathbf{P}_f$. In this table, except the atmosphere part, the approximate error correlations in the ocean and in the atmosphere-ocean part are almost zero so we can expect significant fluctuations on their convergence behaviours. Then to see whether or not our expectations are satisfied, the relative changes of $\mathbf{P}_f$ components are plotted in Figures (5.5),(5.6) and (5.7). The figures also include the results of the first experiment to see the differences between them clearly. More accurate ocean observations cause significant jumps in convergence behaviours of most $\mathbf{P}_f$ components.

In Figure (5.6), except the case of $Cov(X, W)$ the convergence behaviour shows more fluctuations in overall when we compare the Experiment-1 result. For example, $Cov(X, V)$ does not satisfy the convergence condition after 500 ensemble members although it always satisfies the convergence condition after 300 ensemble members in Experiment-1. However, in the case of $Cov(X, W)$ the overall convergence behaviour is better than in the first

experiment. When we compare the error correlations of both experiments in Tables (5.1) and (5.3), it can be seen that while $Cor(X, W)$ increases from -0.0151 to -0.1188 (consider the magnitudes of numbers), others in the atmosphere-ocean part decrease remarkably.

In the variances figure (Figure (5.5)), all of them satisfy the converge condition first with less ensemble size than that of Experiment-1. For example, $Var(Z)$ (in Figure (5.5)(c)) satisfies the convergence condition first with the ensemble size 200 in Experiment-1 whereas it needs 100 ensemble members in Experiment-2.

As we can see in Figure (5.7), more accurate ocean observations make the atmosphere covariances $Cov(X, Y)$, $Cov(X, Z)$, $Cov(Y, Z)$ convergence with all ensemble sizes. If we look at their error correlations in Table (5.3), they are almost fully correlated with the values 0.9915, -0.9758 and -0.9570 respectively. However, in the case of $Cov(W, V)$ there is an almost opposite situation. While $Cov(W, V)$ converges with the 300 ensemble members in Experiment-1, here it needs much more ensemble members than in the first experiment to converge. A decrease from -0.1113 to 0.0233 in their error correlations can also prove the convergence behaviour of $Cov(W, V)$. The reason for this can be because we are less observations in the ocean than in the atmosphere. The following experiment shows the results of the case with more ocean observations than in Experiment-1.

	Atmosphere	
	Experiment-1	Experiment-2
Var (X)	0.00005324	0.00001426
Var (Y)	0.00004628	0.00002096
Var (Z)	0.00006124	0.00003629
Cov (X,Y)	0.00003841	0.00001714
Cov (X,Z)	-0.00003864	-0.00002220
Cov (Y,Z)	-0.00005113	-0.00002639

	Ocean	
	Experiment-1	Experiment-2
Var (W)	0.00389633	0.00002497
Var (V)	0.00466764	0.00002583
Cov (W,V)	-0.00046063	0.00000019

	Atmosphere-Ocean	
	Experiment-1	Experiment-2
Cov(X,W)	-0.00000321	-0.00000260
Cov(X,V)	0.00042373	-0.00000053
Cov(Y,W)	0.00022932	-0.00000174
Cov(Y,V)	0.00020299	-0.00000221
Cov(Z,W)	-0.00026479	0.00000298
Cov(Z,V)	-0.00014751	-0.00007084

Table 5.2: Average values of $\mathbf{P}_f$ components from the ensemble size 6 to 1800 in Experiment-1 and Experiment-2

5.3 Experiment-3

In this experiment, we examine the effects of the number of ocean observations on the convergence of $\mathbf{P}_f$, which is another part of our third research question (see Section (1.1)). Thus, we do similar experiments as in Experiment-1 with more ocean observations. The only difference in the parameters is that we set up the number of ocean observations as 20 in total (in Experiment-1 it was 8). In this case, we have more observations in the ocean than in the atmosphere, which has 12 observations in total. We plot

Atmosphere	
Cor(X,Y)	0.9915
Cor(X,Z)	-0.9758
Cor(Y,Z)	-0.9570
Ocean	
Cor(W,V)	0.0233

Atmosphere-Ocean	
Cor(X,W)	-0.1188
Cor(X,V)	-0.0328
Cor(Y,W)	-0.0621
Cor(Y,V)	-0.0998
Cor(Z,W)	0.0758
Cor(Z,V)	0.1166

Table 5.3: Averages of each error correlation from the ensemble size 6 to 1800 in Experiment-2

their relative changes in Figures (5.8),(5.9) and (5.10) by comparing with the first experiment. Table (5.5) shows the approximate values of each component of $\mathbf{P}_f$. This table is created by the mean of values of $\mathbf{P}_f$ components from ensemble size 6 to 1800. We compare the results of Experiment-1 and 3 and see that more ocean observations lead the errors to spread slightly less.

In both experiments (Experiment-1 and 3), the convergence behaviours of variances are quite similar (see Figure (5.8)). We can say more observations in the ocean make their convergence behaviours a bit better than in Experiment-1 especially in the cases of $Var(Z)$. It seems always convergence with all ensemble sizes from 6 to 1800 although it needs minimum 200 ensemble members to converge in the Experiment-1.

In Figure (5.9), we can see again there is no big differences on convergence behaviours of $\mathbf{P}_f$ components between two experiments. However, $Cov(X, W)$ converge with less ensemble size than in Experiment-1. Here, it seems convergence with 1000 ensemble members although in the first experiment it needs 1200 ensemble members to converge. It would be important to remember that we need to focus on general convergence behaviours from

the ensemble size 6 to 1800 in the figures because at each ensemble size the values can show some differences if we do same experiment with exactly same parameters. Random numbers used generating ensemble members can cause small differences in their values at the certain ensemble size. For the case of $Cov(X, W)$, we can compare their correlation changes between Table (5.1) and (5.4). The error correlation between X and W increase slightly in the case of more ocean observations (in Experiment-1). It changes from -0.0151 to -0.0351. Even if an increase in error correlation of variables is too slight, this can lead the corresponding covariance component to need less ensemble members to converge. In general, when we compare Experiment-1 and Experiment-3, we can see that the correlations of variables almost same in both case (see Table (5.3) and (5.4)). Therefore, we should not expect big differences in their convergence behaviours. Interestingly, more observations in the ocean do not make noticeable changes on the convergence of $\mathbf{P}_f$.

Atmosphere	
Cor(X,Y)	0.8062
Cor(X,Z)	-0.7193
Cor(Y,Z)	-0.9539
Ocean	
Cor(W,V)	-0.1279

Atmosphere-Ocean	
Cor(X,W)	-0.0351
Cor(X,V)	0.7987
Cor(Y,W)	0.4394
Cor(Y,V)	0.3991
Cor(Z,W)	-0.4194
Cor(Z,V)	-0.2369

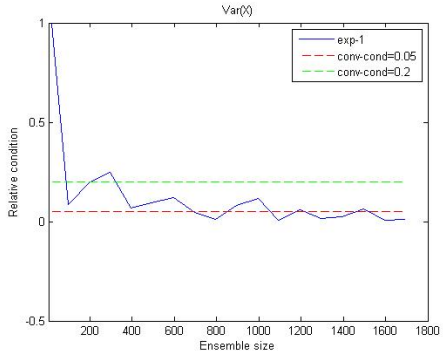
Table 5.4: Averages of each error correlation from the ensemble size 6 to 1800 in Experiment-3

	Atmosphere	
	Experiment-1	Experiment-3
Var (X)	0.00005324	0.00003756
Var (Y)	0.00004628	0.00003586
Var (Z)	0.00006124	0.00005154
Cov (X,Y)	0.00003841	0.00002915
Cov (X,Z)	-0.00003864	-0.00003054
Cov (Y,Z)	-0.00005113	-0.00004090

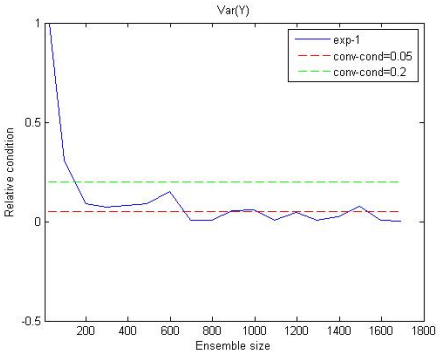
	Ocean	
	Experiment-1	Experiment-3
Var (W)	0.00389633	0.00241492
Var (V)	0.00466764	0.00313906
Cov (W,V)	-0.00046063	-0.00030320

	Atmosphere-Ocean	
	Experiment-1	Experiment-3
Cov(X,W)	-0.00000321	0.00000035
Cov(X,V)	0.00042373	0.00026814
Cov(Y,W)	0.00022932	0.00014535
Cov(Y,V)	0.00020299	0.00011811
Cov(Z,W)	-0.00026479	-0.00016945
Cov(Z,V)	-0.00014751	-0.00007084

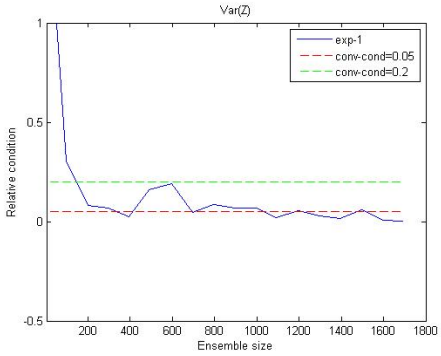
Table 5.5: Average values of $\mathbf{P}_f$ components from the ensemble size 6 to 1800 in Experiment-1 and Experiment-3



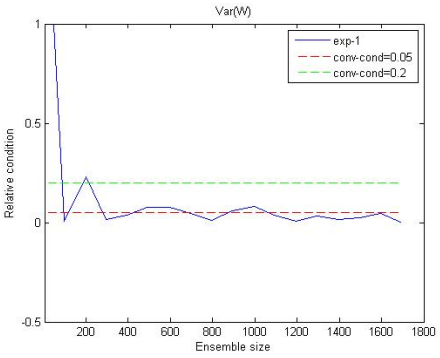
(a) Var(X)



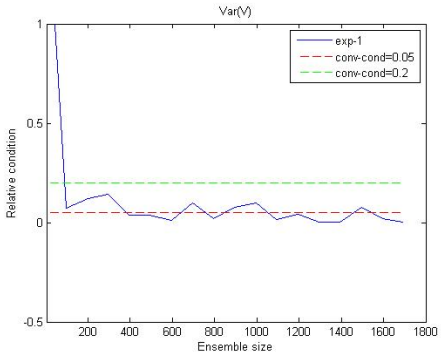
(b) Var(Y)



(c) Var(Z)

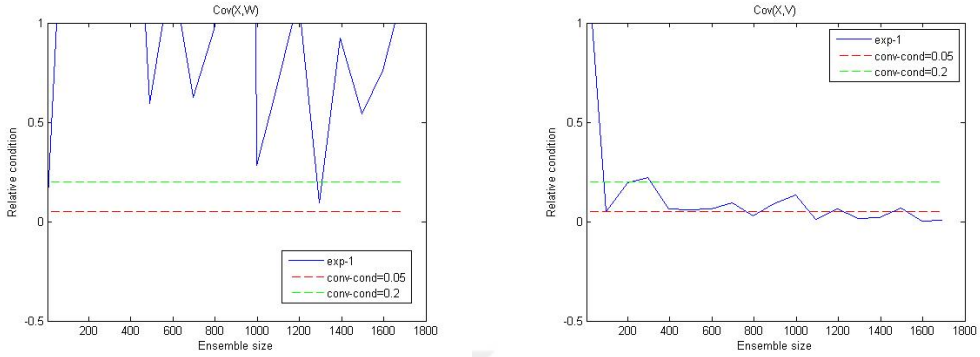


(d) Var(W)



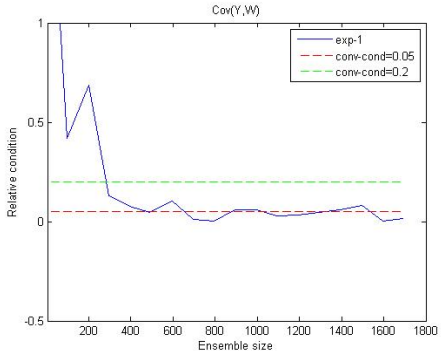
(e) Var(V)

Figure 5.2: The values of convergence condition of variances for different ensemble sizes in Experiment-1

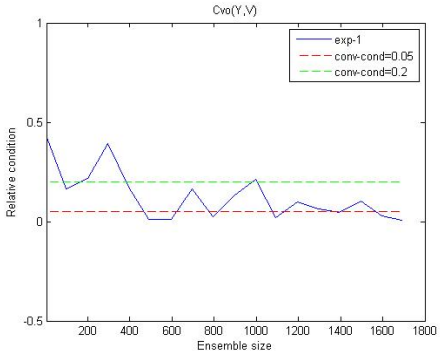


(a) Cov(X,W)

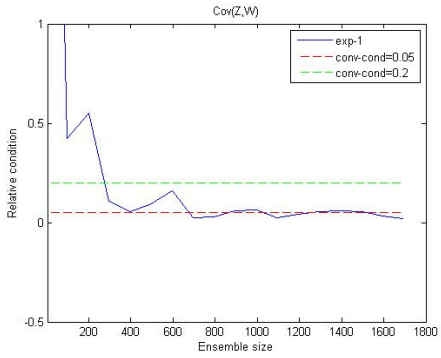
(b) Cov(X,V)



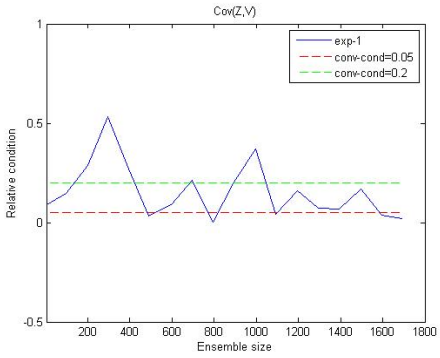
(c) Cov(Y,W)



(d) Cov(Y,V)

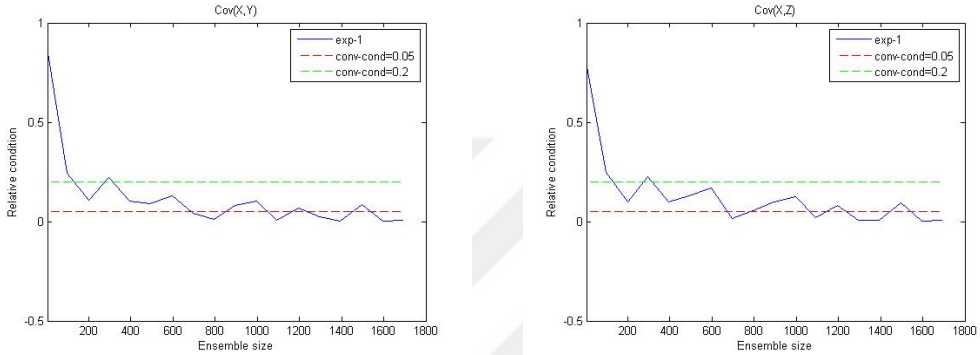


(e) Cov(Z,W)

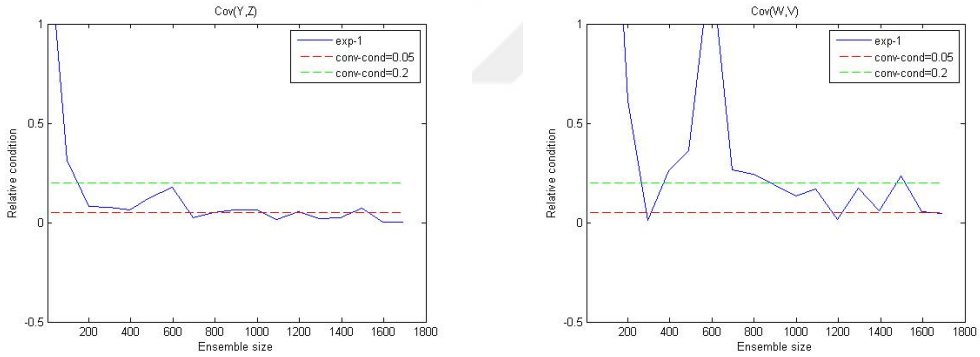


(f) Cov(Z,V)

Figure 5.3: The values of convergence condition of cross-covariances for different ensemble sizes in Experiment-1

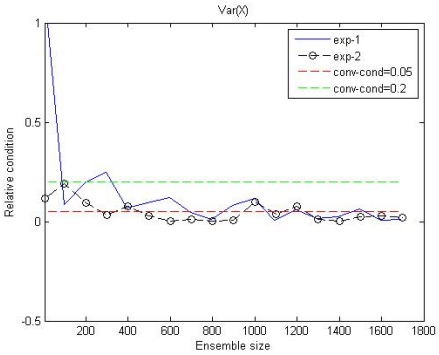


(a) Cov(X,Y) (b) Cov(X,Z)

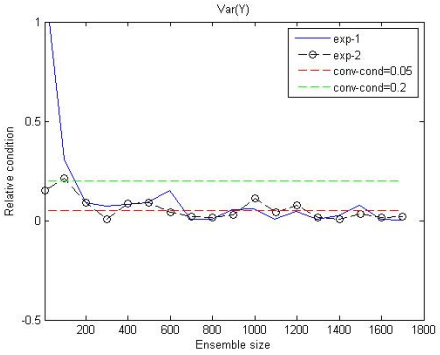


(c) Cov(Y,Z) (d) Cov(W,V)

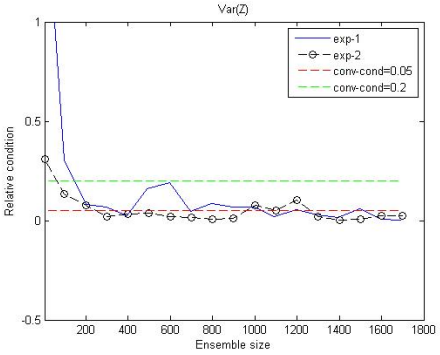
Figure 5.4: The values of convergence condition of covariances for different ensemble sizes in Experiment-1



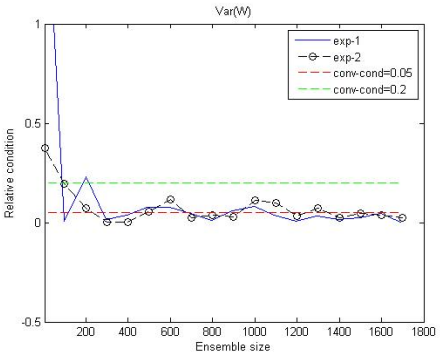
(a) Var(X)



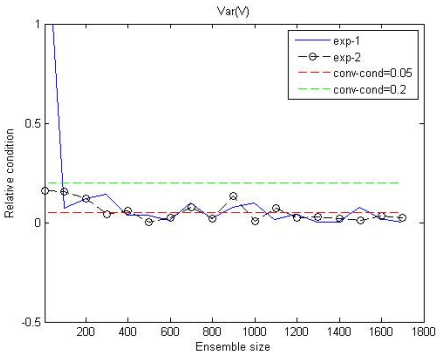
(b) Var(Y)



(c) Var(Z)

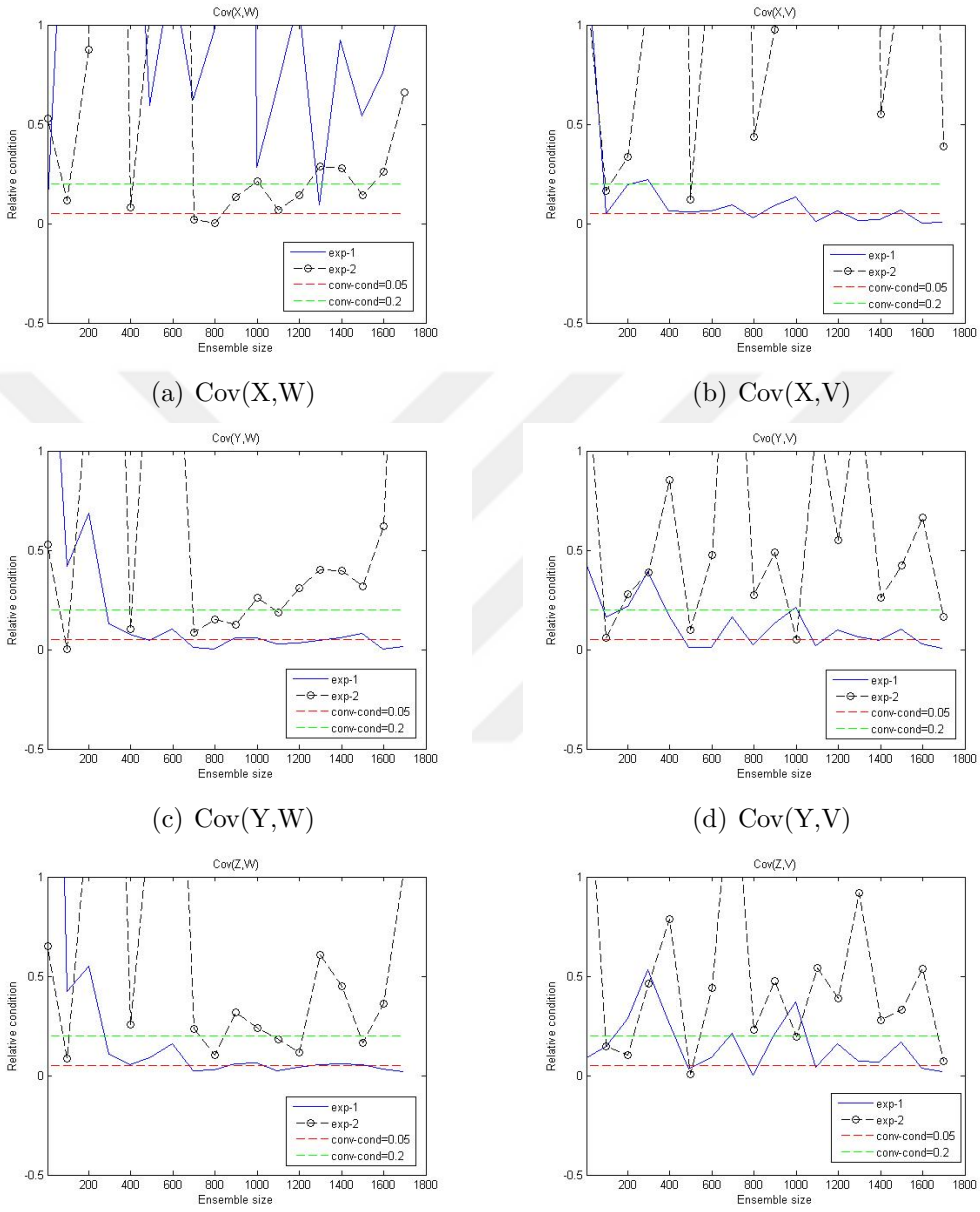


(d) Var(W)



(e) Var(V)

Figure 5.5: The values of convergence condition of variances for different ensemble sizes in Experiment-2 are compared with the results in Experiment-1

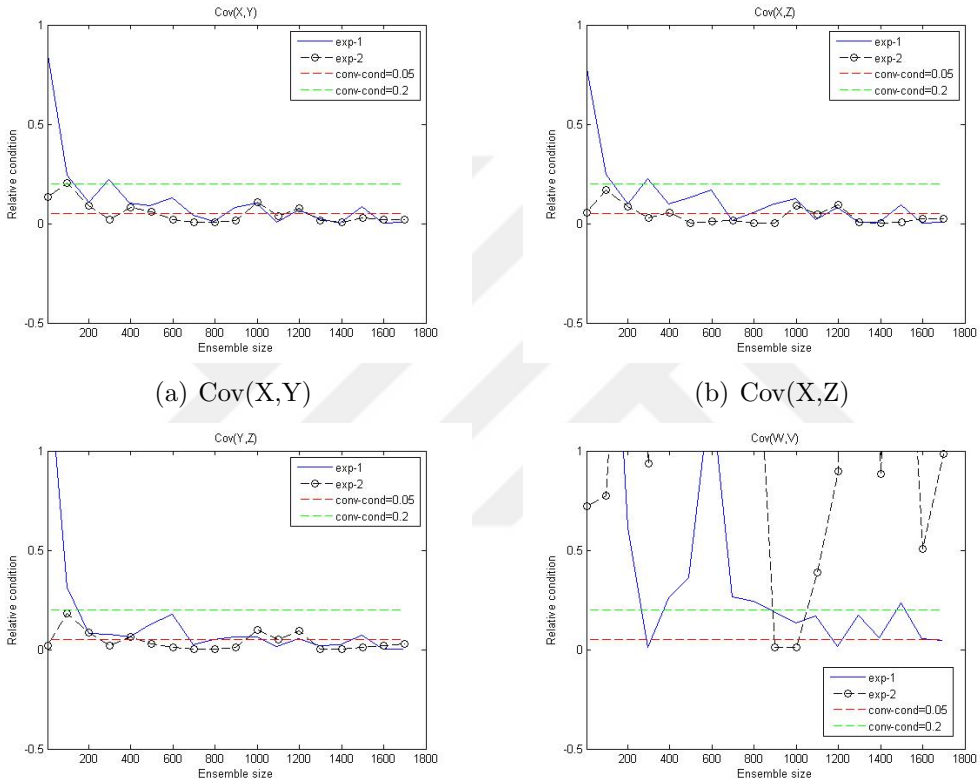


(a) Cov(X,W) (b) Cov(X,V)

(c) Cov(Y,W) (d) Cov(Y,V)

(e) Cov(Z,W) (f) Cov(Z,V)

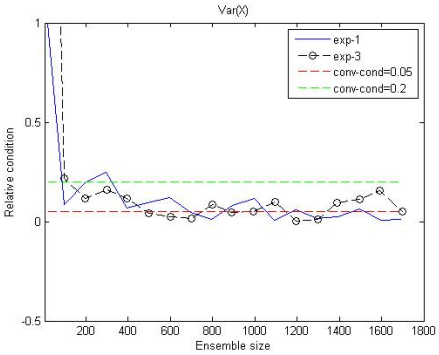
Figure 5.6: The values of convergence condition of cross-covariances for different ensemble sizes in Experiment-2 are compared with the results in Experiment-1



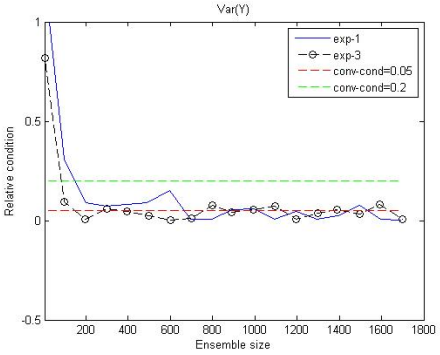
(a) Cov(X,Y) (b) Cov(X,Z)

(c) Cov(Y,Z) (d) Cov(W,V)

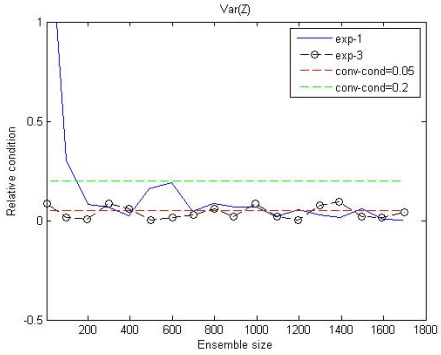
Figure 5.7: The values of convergence condition of covariances for different ensemble sizes in Experiment-2 are compared with the results in Experiment-1



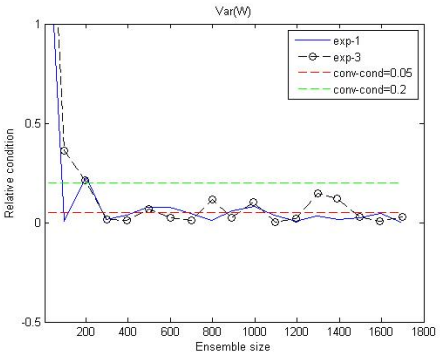
(a) Var(X)



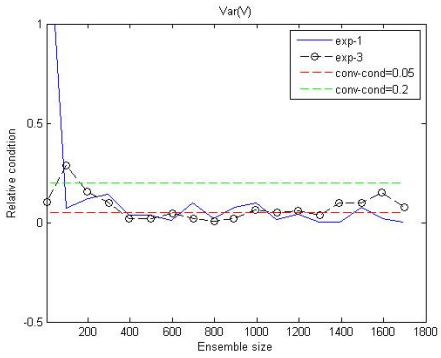
(b) Var(Y)



(c) Var(Z)

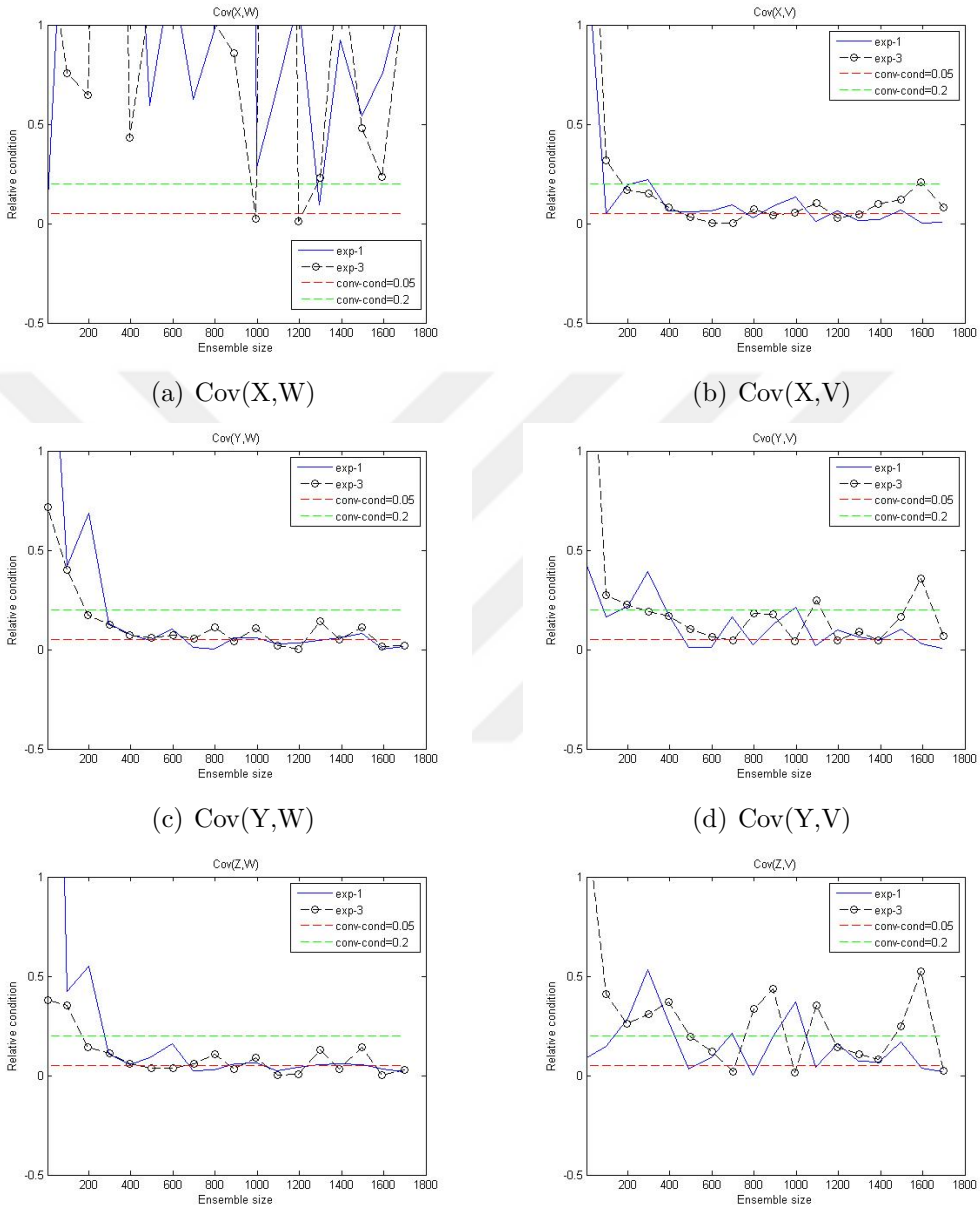


(d) Var(W)



(e) Var(V)

Figure 5.8: The values of convergence condition of variances for different ensemble sizes in Experiment-3 are compared with the results in Experiment-1

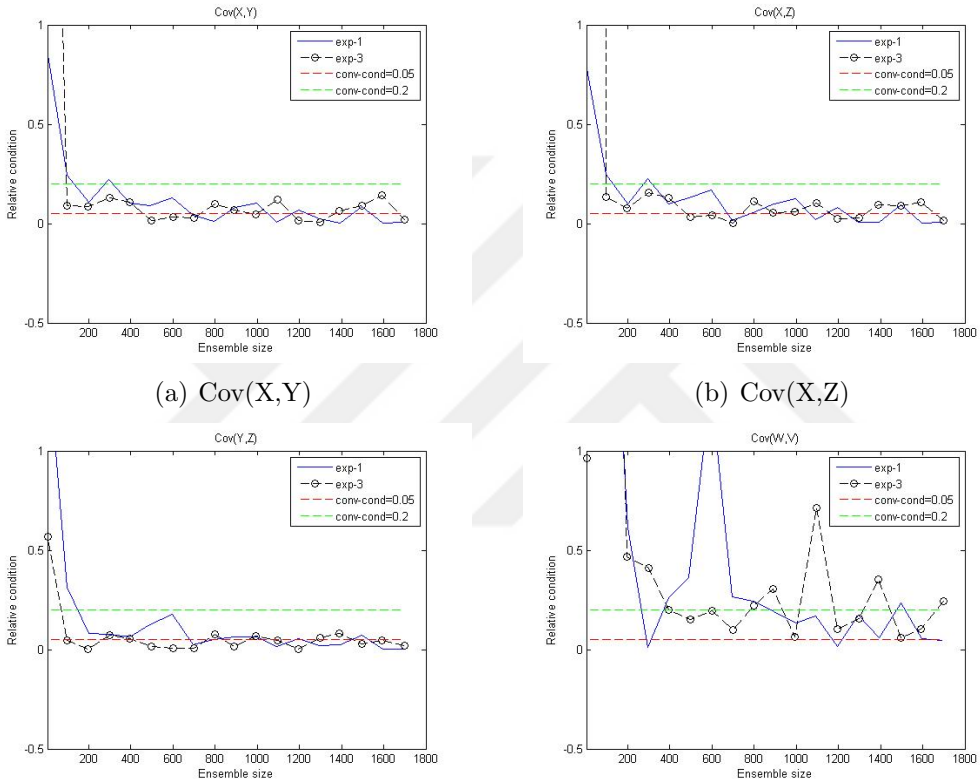


(a) Cov(X,W) (b) Cov(X,V)

(c) Cov(Y,W) (d) Cov(Y,V)

(e) Cov(Z,W) (f) Cov(Z,V)

Figure 5.9: The values of convergence condition of cross-covariances for different ensemble sizes in Experiment-3 are compared with the results in Experiment-1



(a) Cov(X,Y)

(b) Cov(X,Z)

(c) Cov(Y,Z)

(d) Cov(W,V)

Figure 5.10: The values of convergence condition of covariances for different ensemble sizes in Experiment-3 are compared with the results in Experiment-1

Chapter 6

Discussion

6.1 Summary and Conclusion

The aim of this project is to investigate following questions:

(1) How many ensemble members are needed to capture the forecast covariances correctly?

(2) What are the effects of error correlations on the estimation of forecast covariances?

(3) How these are affected by observation errors and frequencies?

To do this, first we generate a method which is an ensemble of non-incremental 4DVar methods. In this method, we generate ensemble members by perturbing background and observations with different random numbers. Then, we try to estimate the forecast error covariance matrix with different ensemble sizes. We try to look at the convergence behaviour of $\mathbf{P}_f$ compo-

nents as the ensemble size getting larger and demonstrate the effects of error correlations on the convergence of $\mathbf{P}_f$. Then, by changing the accuracies and the number of ocean observations we try to show the differences in the convergence behaviours of $\mathbf{P}_f$ components.

In our experiments, to be able to estimate $\mathbf{P}_f$, we investigate the convergence behaviour of its components as the ensemble size getting bigger. Each components of $\mathbf{P}_f$ converges to different numbers which can have different orders of magnitude. Thus, relative convergence condition is chosen to determine their convergence. Every component of $\mathbf{P}_f$ is tested separately with respect to the relative condition. We mostly focus on their convergence behaviour not the numbers they converge. We can also guess their convergence behaviour by looking at the error correlations. From our experiment results, we see that there are strong relationship between convergence behaviour of one component of $\mathbf{P}_f$ and the error correlation of its corresponding variables. The less correlation in errors the less convergence tendency in its convergence behaviour, and vice versa. It is important to know that the experimental setups used in this dissertation are not ideal ones. We use particular parameters sets. It would be useful change the parameters and do similar experiments.

In the first experiment, we try to estimate the matrix $\mathbf{P}_f$ in the system of more accurate observations with more observations in the atmosphere than in the ocean. We see that there are significant jumps on the convergence behaviour of $Cov(X, W)$ and $Cov(W, V)$. When we look at the error correlations of (X and W) and (W and V) separately, there are almost no correlation between them. The reason for the convergence behaviour of $Cov(X, W)$ can be because of the model system. There is no direct con-

nection between X and W. From this view, we expect similar convergence behaviour for $Cov(Y, V)$, $Cov(Z, W)$ and $Cov(Z, V)$, because they have also no direct connection between their corresponding variables in the model system. However, they are more or less close to the convergence condition limit even though they do not always satisfy the convergence condition from the ensemble size 6 to 1800. As we expect, they have more error correlations in their corresponding variables than $Cov(X, W)$. The differences can be from Y and Z have nonlinear formulas in the model system, while X has linear one. Nonlinear formulas can help the variables to capture the relationship between them more quickly than linear ones. This can be a reason for why $Cov(X, W)$ has higher fluctuations in its convergence behaviour than $Cov(Y, V)$, $Cov(Z, W)$ and $Cov(Z, V)$. In the case of $Cov(W, V)$, the reason for its convergence behaviour can be because of the parameters we set up. In this experiment, there are less accurate and less observations in the ocean than in the atmosphere. In the other experiments, the results are compared with the first experiment results separately.

In the second experiment, we examine the effects of ocean observation accuracies on the estimation of $\mathbf{P}_f$. The only difference in the parameters set up in the first experiment is that the ocean observations become more accurate than in the first experiment. In this experiment, most components of $\mathbf{P}_f$ show remarkable fluctuations in their convergence behaviours. It can also be checked from their error correlations. Most errors become almost uncorrelated. Interestingly, more accurate observations in ocean lead the covariances of ocean variables to show less convergence tendency with each ensemble size. As a result, estimating $\mathbf{P}_f$ becomes more difficult with more accurate ocean observations than in the case of Experiment-1.

In the third experiment, it is investigated that the effects of more ocean observations on the estimation of $\mathbf{P}_f$. We only make the number of ocean observations more than in the case of Experiment-1. Even if we make the number of ocean observations more than two times, the results are quite similar in the both experiments (Experiment-1 and 3). We can also see this result by comparing the error correlations. In both of these experiments, the error correlations of each component are almost same. As a result, estimating $\mathbf{P}_f$ are not be affected too much by the number of ocean observations.

In our experiments, we try to estimate the matrix $\mathbf{P}_f$ by investigating the convergence behaviours of its components separately as the ensemble size gets larger. Getting an exact conclusion is difficult in our case because we set up different ensemble sizes and do the experiment once for each of them. It would be useful to do same experiment more than once with the same parameters to be able to generalize the results. However, it needs more computational work in terms of time. If we do same experiments with exactly same parameters, the results can show some differences from those presented here. Random numbers can affect the results, because the ensemble members are generated by using different random numbers.

6.2 Future Work

In this dissertation, we produce the ensemble of initial analysis states by using the same background error covariance matrix $\mathbf{B}$ for one assimilation window. Then, we forecast each analysis over the assimilation window by

discretizing the model step by step. At the final time, we have an ensemble based forecast error covariance matrix $\mathbf{P}_f$ which we try to estimate in our experiments. After the estimation of the matrix, it should be used for the next assimilation window as a background error covariance matrix. Then, it would be great to generate new analysis ensemble members by using $\mathbf{P}_f$.

The matrix $\mathbf{P}_f$ includes the effects of the forecast. Investigating the convergence of $\mathbf{P}_a$, which is the analysis error covariance matrix at the initial time, would be salient to eliminate the forecast effects on this matrix. We would suggest to do same experiments for the estimation of $\mathbf{P}_a$.

We only study on the assimilation time length 1. More time length can lead the analysis trajectories spread more. Investigating the effects of the different time length on the convergence of $\mathbf{P}_f$ can be useful.

Bibliography

Anderson, J. L. and Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review, 127*(12), 2741-2758.

Bouttier, F. and Courtier, P. (2002). Data assimilation concepts and methods. *Meteorological Training Course Lecture Series, ECMWF* http://old.ecmwf.int/newsevents/training/rcourse_notes /DATA_ASSIMILATION/ASSIM_CONCEPTS/Assim_concepts2.html

Cheng, H., Jardak, M., Alexe, M., and Sandu, A. (2010). A hybrid approach to estimating error covariances in variational data assimilation. *Tellus A, 62*(3), 288-297.

Dimet, F. X. L. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A, 38*(2), 97-110.

Dubois, M. A. and Yiou, P. (1999). Testing asynchronous coupling on simple ocean-atmosphere dynamic systems. *Climate dynamics, 15*(1),

1-7.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans (19782012), 99*(C5), 10143-10162.

Fairbairn, D., Pring, S. R., Lorenc, A. C., and Roulstone, I. (2014). A comparison of 4DVar with ensemble data assimilation methods. *Quarterly Journal of the Royal Meteorological Society, 140*(678), 281-294.

Grewal, M. S., and Andrews, A. P. (2011). *Kalman filtering: theory and practice using MATLAB.* John Wiley and Sons.

Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman Filter. *Monthly Weather Review, 129*(11), 2776-2790.

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., and Raynaud, L. (2010). *Ensemble of data assimilations at ECMWF.* European Centre for Medium-Range Weather Forecasts.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory.* Academic press.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering, 82*(1), 35-45.

Kalnay, E. (2003). *Atmospheric modelling, data assimilation, and predictability.* Cambridge university press.

Lawless, A.S. (2006). Data assimilation with the Lorenz equations. http:// darc.nerc.ac.uk/, 12/08/2006, simple models

Lawless, A.S. (2013). *Variational data assimilation for very large environmental problems.* In *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences* (2013), Eds. Cullen, M.J.P., Freitag, M. A., Kindermann, S., Scheichl, R., Radon Series on Computational and Applied Mathematics 13. De Gruyter, pp. 55-90.

Liu, C., Xiao, Q., and Wang, B. (2008). An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Monthly Weather Review, 136*(9), 3363-3373.

Lorenz, E. N. (1963). Deterministic non-periodic flow. *Journal of the atmospheric sciences, 20*(2), 130-141.

Molteni, F., Ferranti, L., Palmer, T. N., and Viterbo, P. (1993). A dynamical interpretation of the global response to equatorial Pacific SST anomalies. *Journal of climate, 6*(5), 777-795.

Nichols, N. K. (2010). Mathematical concepts of data assimilation. In *Data Assimilation* (pp. 13-39). Springer Berlin Heidelberg.

Voss, R. and Sausen, R. (1996). Techniques for asynchronous and periodically synchronous coupling of atmosphere and ocean models. *Climate Dynamics, 12*(9), 605-614.