Techniques to account for and reduce model inadequacy in ensemble-based filters

by

Vikram Khade

Submitted to the Department of Earth, Atmospheric and Planetary sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2008

Author .....
Department of Earth, Atmospheric and Planetary sciences
April 14, 2008

Certified by .....
Kerry Emanuel
Breene M. Kerr Professor of Meteorology
Thesis Supervisor

Accepted by .....
Maria Zuber
E.A. Griswold Professor of Geophysics
Head, Department of Earth, Atmospheric and Planetary sciences

Techniques to account for and reduce model inadequacy in ensemble-based filters

by

Vikram Khade

Submitted to the Department of Earth, Atmospheric and Planetary sciences
on April 14, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

A technique for the accounting for parametric model error in the Ensemble Kalman Filter (EnKF) is investigated within the framework of Additive Error Approximation (AEA). The AEA needs an estimate of the model error covariance structure. The state-dependent model error structure is the sensitivity of the local attractor to the parameter. The Multimodel Method (MMM) and Parametric Vector Method (PVM) to estimate this state-dependent sensitivity are introduced and investigated in the low-dimensional Ikeda and L63 systems. The MMM involves assimilating data independently into multiple models. PVM aims at obtaining the estimate given by MMM using a single model. At the heart of the PVM is the concept of adjoint sensitivity which is obtained using parametric singular vectors. It is found that PVM is able to estimate the correct state-dependent model error structure if the parametric vectors are constructed over an optimization time (τ_{op}) which is equal to the state-dependent optimal time (τ_{om}). The optimal time is the time taken by a state to go from an off-attractor location to an on-attractor location. If $\tau_{op} < \tau_{om}$ then the parametric vector gives the transient sensitivity which is the incorrect model error structure. On the other hand, if $\tau_{op} > \tau_{om}$ the sensitivity obtained is non-local and tends to point in the direction of largest state error growth. The average (over the phase space) τ_{om} is calculated for the Ikeda and L63 systems. MMM and PVM give lower average analysis and forecast errors than state-independent estimates of model error structure.

Parameter estimation is a typical example of reduction of model error. The state-dependent parameter estimation (parameter tuning) in the Ikeda system is successful in partially compensating for structural model error thus resulting in lower analysis and forecast errors. However, parameter tuning is not able to completely eliminate structural model error. Nonetheless, parameter tuning can be used to identify processes in the model that have large model error. The parameters in the Emanuel convection scheme are tuned in the NOGAPS model. This parameter tuning is able to partially compensate for structural model error in the vertical flux parametrization.

Thesis Supervisor: Kerry Emanuel
Title: Breene M. Kerr Professor of Meteorology

"If a man will begin in certainties he shall end in doubts; but if he will be content to begin in doubts he shall end in certainties."

- Francis Bacon, Philosopher and essayist (1561-1626)

Acknowledgments

(Dedicated to the memory of my father, who did not leave anything to chance in my upbringing.)

I would like to express my deep sense of gratitude to Dr. Jim Hansen, Naval Research Laboratory (NRL), Monterey, for his help throughout my graduate studies. I would like to thank the other members of my thesis committee, namely Prof. Kerry Emanuel, Prof. Paola Malanotte-Rizzoli and Prof. Dennis McLaughlin, without whose advice and steadfast support this thesis would not have seen the light of the day. The expertise of the committee members in the fields of state estimation, nonlinear/chaotic dynamics and modeling was very beneficial to my work. Prof. Dennis McLaughlin has viewed my work from the vantage point of being both a theorist and practitioner of state estimation. His suggestions helped me to understand the fundamentals of state estimation. Prof. Paola Malanotte-Rizzoli, being one of the innovators in the field of oceanographic state estimation, offered valuable insights into the various aspects of state estimation. I thank Prof. Kerry Emanuel for sharing his profound insights into state estimation, chaotic dynamics and his own convection parametrization scheme.

Prof. Richard Lindzen was my advisor when I arrived at MIT for my graduate studies in the Fall of 2001. I thank him for encouraging me to pursue research in the field of my liking. His encouragement greatly facilitated my migration to the discipline of chaotic dynamics and state estimation. I have always been fascinated by chaotic systems. I was fortunate that Jim Hansen had newly joined the PAOC faculty as an assistant professor. He kindly agreed to be my advisor. I was most attracted to his work because of his unique ideas which combined elements of chaotic systems, linear algebra and state estimation. I liked his philosophy of learning lessons from low-dimensional systems with the ultimate aim of applying them to high-dimensional atmospheric models. He has a great appreciation for understanding the science of state estimation, but at the same time he addresses the practical questions which are of immediate concern from the point of view of atmospheric and oceanic state estimation. I believe his novel work of interpreting state estimation in the light of chaotic systems will continue to bear rich fruits in the future. After Jim Hansen left MIT to take up a new job at NRL, Prof. Kerry Emanuel graciously took up the responsibility of advising me. I salute him for his pivotal role as my advisor. Jim, though away from MIT, was always forthcoming on phone and email for conversations about my research. I am indebted to Jim and Kerry for the constructive roles they played during my Ph.D.

I thank my colleagues in Jim Hansen's research group, namely Andrew Eichmann, Dan Gombos, Andrew Lawrence, Greg Lawson, Jon Moskaitis and Tim Whitcomb. They were my friends, collaborators, consolers and supporters all rolled into one. I am grateful to Adel Ahanin for teaching me the basics of state estimation. I thank Prof. Edward Lorenz for his inspiring presence and stimulating discussions I had with him. My conversations with Dr. Ross Hoffman, though occasional, were refreshing. I thank Prof. Alan Plumb, Director of PAOC, for his continuous encouragement.

I had the pleasure of visiting NRL, Monterey for one month in January 2005. I thank the

authorities at NRL for arranging my visit. My collaboration with the scientists at NRL has contributed significantly to my thesis.

I was a Teaching Assistant for course 10.571/12.806 (Atmospheric chemistry and physics) in the Spring 2006 semester and for course 12.815 (Atmospheric radiation) in the Fall 2007 semester. I thank Prof. Ronald Prinn, Prof. Gregory McRae and Dr. Robert McClatchey for giving me these opportunities. I am grateful to Xue and Eunjee, who were the past Teaching Assistants 12.815, for their alacritous assistance. The teaching involved in these courses was an enriching learning experience. I was funded by Prof. Richard Lindzen, Prof. Jim Hansen, Prof. Dennis McLaughlin and Prof. Kerry Emanuel at various stages in my graduate work.

I thank Masahiro for his help with the understanding of general atmospheric sciences. I had the pleasure of working with other graduate students like Bill, Brian, Dave, Elke, Fabio, Fanny, Ian, Mike, Pablo, Payal, Roberto, Uday and Valerio. I thank all of them.

I thank Carol, Frances, Jackie, Jeff, Joel, Kristen, Mark, Mary, Roberta and Vicki for their moral and administrative support. Linda and Scott were always prompt in attending to my problems with computers, printers and so on.

I would like to express my gratitude for my teachers and professors in India. I would also like to thank my friends back in India. Last but not the least, I would like to thank my family for their love and support.

Contents

Chapter 1

Introduction and Overview

1.1 Introduction

The prediction of atmosphere, ocean and climate systems is one of the important disciplines in the earth sciences. The most important reason why a sizable subset of geophysical scientists have concentrated on forecasting is the practical application of this branch. Operational weather forecasting is a typical example of such a practical application. It has long been recognized that inaccuracies in weather forecasts result from uncertainties in the initial conditions, boundary conditions and model imperfections. The initial condition uncertainty has a number of sources, including the paucity of observations and noise in observations, making it impossible to accurately determine the state of the atmosphere. An accurate estimate means an estimate that is very close to the true state of the atmosphere where the "closeness" is with respect to a chosen norm. Though the meteorological community was aware of the initial condition uncertainty problem (Thompson, 1957), its gravity was fully appreciated only after the discovery of *chaos* (Lorenz, 1963).

The scientific breakthroughs (Lorenz, 1963) in the last few decades coupled with the advent of modern technology like faster computers and data storage devices has broadened the horizons of forecasting geophysical systems. Particularly, in the field of atmospheric sciences, the temporal scales of forecasting have increased from a few days for weather models to many decades for climate models. Presently, climate models are integrated forward in time for time scales of a few centuries to obtain a forecast of future climate. At the most basic level, the forecasting or prediction of a given system involves the integration of the dynamical equations of the system forward in time starting from an initial condition. This is called a deterministic prediction or forecast. Given a perfect model and correct boundary conditions the forecast is

imperfect or wrong only because the initial condition is incorrect.

The concept of prediction was taken to a new level by the advent of *optimal estimation*, also called *state estimation*. In the meteorology community, state estimation is also called data assimilation (DA). The seeds for the theory and practice of state estimation were sown as early as in 1809 by Gauss in his book *Theoria Motus* (Gauss, 1857). Given the equations of the system, also called the *model*, and the observations of the state, state estimation involves combining the model state and observations to obtain an optimal estimate of the state. Apart from the future state of the system, one may also be interested in knowing the current and the past state of the system. The estimation of the past, the current and the future state of a system is termed smoothing, filtering and prediction, respectively (Gelb, 1974). In meteorological terminology, these are termed reanalysis, analysis and forecasting, respectively. This work deals only with analysis and forecasting.

The theory of state estimation was steadily developed in the 20th century by pioneers like Fisher (Fisher, 1912), Wiener (Wiener, 1949), and Kolmogorov (Kolmogorov, 1941). The ground breaking work of Kalman (Kalman, 1960) led to an increase in the application of state estimation. Kalman developed a methodology, named the Kalman filter, which does not necessitate the storage of past observations. This is the most attractive feature of the Kalman filter and makes it ideally suited to real-time filtering and prediction problems. Since its inception in 1960, the Kalman filter and its variants have been widely used in several branches of engineering and applied sciences like mechanical engineering, electrical engineering and aeronautical engineering. Until recently (about a decade ago), the Kalman filter methodology could not be applied to atmospheric and oceanographic models. This was because of the high dimensionality and nonlinearity of these models. The introduction of the Ensemble Kalman filter (Evensen, 1994) has opened new avenues for applying the Kalman filter methodology to atmospheric and oceanographic models. A brief review of the evolution of Kalman filter to the Ensemble Kalman filter follows.

Consider a linear model given by $\mathbf{F}$. Let $\mathbf{F}$ be perfect. Let $\mathbf{x}$ denote the model state. Let n denote the dimension of the state $\mathbf{x}$, that is, the model is n-dimensional. In this discussion, the time denoted by t, is assumed to be discrete. Let the correct boundary conditions be given. Because the model is perfect, the only reason for uncertainty in the forecast is the uncertainty in the initial condition. Given the correct initial condition, that is, the correct value of $\mathbf{x}$ at time $t = 0$, $\mathbf{F}$ produces the perfect forecast at all future times. Let the true state at time t be denoted by $\mathbf{x}^t(t)$. The superscript in $\mathbf{x}^t(t)$ denotes *truth*. An instrument is observing the truth. This instrument provides observations at equally spaced time units. The time interval between two consecutive observations is defined to be 1 time unit. The observations are also called data or measurements. An observation of the true state at time t is denoted

by $\mathbf{y}^o(t)$. The observation is not perfect, that is, it has uncertainty. This uncertainty in $\mathbf{y}^o(t)$, assumed to be gaussian, is quantified by the covariance matrix $\mathbf{R}$. $\mathbf{R}$ could be function of time, but typically it is independent of time. The Kalman filter equations are (Houtekamer and Mitchell, 2005),

$$\mathbf{x}^f(t) = \mathbf{F}[\mathbf{x}^a(t-1)] \tag{1.1a}$$

$$\mathbf{P}^f(t) = \mathbf{F}\mathbf{P}^a(t-1)\mathbf{F}^T \tag{1.1b}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T[\mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t)]^{-1} \tag{1.1c}$$

$$\mathbf{x}^a(t) = \mathbf{x}^f(t) + \mathbf{K}(t)[\mathbf{y}^o(t) - \mathbf{H}(t)\mathbf{x}^f(t)] \tag{1.1d}$$

$$\mathbf{P}^a(t) = [\mathbf{I} - \mathbf{K}(t)\mathbf{H}(t)]\mathbf{P}^f(t) \tag{1.1e}$$

where $\mathbf{x}^a(t-1)$ is the best estimate of true state at time $t-1$. Equation (1.1a) propagates this best estimate using the model $\mathbf{F}$ from time $t-1$ to time t. Consequently $\mathbf{x}^f(t)$ is the forecast at time t. $\mathbf{P}^a(t-1)$ is a covariance matrix which is an estimate of uncertainty in $\mathbf{x}^a(t-1)$. This uncertainty is assumed to be gaussian. Equation (1.1b) propagates the uncertainty in $\mathbf{x}^a(t-1)$ from time $t-1$ to time t.

The matrix $\mathbf{H}$ is a linear operator that maps the state from the model space to the observation space. $\mathbf{H}$ is also called the measurement matrix. At time t one has two independent estimates of the truth $\mathbf{x}^t(t)$ namely $\mathbf{x}^f(t)$ and $\mathbf{y}^o(t)$, along with their uncertainties $\mathbf{P}^f(t)$ and $\mathbf{R}(t)$, respectively. The Kalman filter combines or blends these two estimates to arrive at an optimal estimate called analysis, denoted by $\mathbf{x}^a(t)$. $\mathbf{x}^a(t)$ is a weighted average of $\mathbf{x}^f(t)$ and $\mathbf{y}^o(t)$ where the weight is given by the Kalman gain $\mathbf{K}(t)$. The expression for $\mathbf{K}(t)$, given by equation (1.1c), is obtained by minimizing the variance in $\mathbf{x}^a(t)$, where $\mathbf{x}^a(t)$ is linear combination of $\mathbf{x}^f(t)$ and $\mathbf{y}^o(t)$ given by (1.1d). The closed form solution given by (1.1d) along with (1.1c) is possible because $\mathbf{F}$ is linear and $\mathbf{P}^a(t-1)$ is gaussian (McLaughlin, 2002). Note that $\mathbf{P}^f(t)$ resulting from the operation on the rhs of equation (1.1b) is gaussian. This is because $\mathbf{F}$ is linear. The uncertainty in $\mathbf{x}^a(t)$ is quantified by the covariance matrix $\mathbf{P}^a(t)$ given by equation (1.1e). $\mathbf{P}^f(t)$ and $\mathbf{P}^a(t)$ are called the forecast error covariance and analysis error covariance, respectively. Equations (1.1a) through (1.1e) constitute the Kalman filter, which will be referred to as the *standard Kalman filter*. The term *Kalman filter* will be used to describe the general methodology or algorithm constituted by equations (1.1a) through (1.1e). The observation at time t is used in the Kalman filter and discarded, that is, it need not be stored and therefore the Kalman filter is very attractive for real time applications. However, the Kalman filter involves the storage of covariance matrices whose size n is dictated by the dimension of the model. Also, inversion of the matrix of size n in equation (1.1c) is involved, which could be computationally expensive for high-dimensional models. The

implementation of equation (1.1b) is very computationally expensive for high-dimensional models (Ghil and Malanotte-Rizzoli, 1991). A comprehensive review of data assimilation in atmosphere and ocean is given in Ghil and Malanotte-Rizzoli (1991). Ghil and Malanotte-Rizzoli (1991) provides an extensive and detailed description of the theory and practice of data assimilation as prevalent in 1980s.

The model $\mathbf{F}$ in the above discussion is taken to be linear. The geophysical models are seldom linear. In the case of nonlinear models, the Kalman filter equations given by (1.1a) through (1.1e) are suitably adapted, giving rise to the Extended Kalman filter (EKF). Let the model $\mathbf{F}$ be nonlinear. Also let the measurement matrix $\mathbf{H}(t)$ be nonlinear. Let $\mathbf{H}(t)'$ be the linearization of $\mathbf{H}(t)$. The Extended Kalman filter (EKF) equations are given by (Ghil and Malanotte-Rizzoli, 1991; Houtekamer and Mitchell, 2005) as,

$$\mathbf{x}^f(t) = \mathbf{F}[\mathbf{x}^a(t-1)] \tag{1.2a}$$

$$\mathbf{P}^f(t) = \mathbf{F}'\mathbf{P}^a(t-1)\mathbf{F}'^T \tag{1.2b}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)'^T[\mathbf{H}(t)'\mathbf{P}^f(t)\mathbf{H}(t)'^T + \mathbf{R}(t)]^{-1} \tag{1.2c}$$

$$\mathbf{x}^a(t) = \mathbf{x}^f(t) + \mathbf{K}(t)[\mathbf{y}^o(t) - \mathbf{H}(t)\mathbf{x}^f(t)] \tag{1.2d}$$

$$\mathbf{P}^a(t) = [\mathbf{I} - \mathbf{K}(t)\mathbf{H}(t)']\mathbf{P}^f(t) \tag{1.2e}$$

The estimate of state $\mathbf{x}^a(t-1)$ is propagated to time t in equation (1.2a). $\mathbf{P}^a(t-1)$ is assumed to be gaussian. The forecast error covariance ($\mathbf{P}^f(t)$) is obtained by equation (1.2b) where $\mathbf{F}'$ is the linearization of $\mathbf{F}$. Equations (1.2c) to (1.2e) are in general similar to (1.1c) to (1.1e). The difference is that equations (1.2c) and (1.2e) use the linearization $\mathbf{H}'$ rather than $\mathbf{H}$. Unlike the standard Kalman filter, the Extended Kalman filter does not guarantee the optimality of $\mathbf{x}^a(t)$. This is because of the nonlinearity in $\mathbf{F}$. The most important difference between the standard Kalman filter and the EKF is the way in which the forecast error covariance is calculated. The propagation of uncertainty as given by equation (1.2b) is better understood in the framework of stochastic-dynamic prediction (Epstein, 1969).

Consider a dynamical model given by $\dot{\mathbf{x}} = \mathbf{G}(\mathbf{x}, t)$. $\dot{\mathbf{x}}$ is the time derivative of the n-dimensional state x. This derivative depends on $\mathbf{x}$ and time t. Let the uncertainty in $\mathbf{x}$ be quantified by the probability distribution function ϕ. Then the evolution of ϕ in time is given by the Liouville equation, (Epstein, 1969),

$$\frac{\partial \phi}{\partial t} + \boldsymbol{\nabla}\cdot(\dot{\mathbf{x}}\phi) = 0 \tag{1.3}$$

The Liouville equation represents the conservation of probability in the phase space defined

by $\mathbf{x}$. Therefore equation (1.3) is called the continuity equation for probability in analogy with the continuity equation for mass used in fluid mechanics (Hansen, 2002). In principle the uncertainty in $\mathbf{x}$ at a future time is quantified by the integration of equation (1.3) in time. But if the dimensionality n of the system is high, then the solution of (1.3) by direct numerical integration is prohibitively high. Epstein (1969) contended that the information contained in the first two moments of ϕ is enough to quantify the uncertainty in $\mathbf{x}$. Therefore, propagating the first two moments (mean and covariance) of ϕ serves as a good approximation to propagating ϕ. But even the propagation of the first two moments poses problems, because these moments depend upon higher moments. So a further approximation is introduced where a closure scheme is used in which all moments with orders higher than the covariance are neglected. This is called *approximate stochastic-dynamic prediction*. This approximation was introduced by Epstein (1969) and has been discussed in detail by Evensen (1994). The use of linearized $\mathbf{F}$ in equation (1.2b) implies that EKF calculates the forecast error covariance by using the approximate stochastic-dynamic prediction. Though equation (1.2b) provides a method to propagate the uncertainty in time and thus extends the standard Kalman filter to the case of nonlinear models, this implementation has a drawback. The closure scheme neglects higher order moments and consequently injects some imbalance in the estimation of the forecast error covariance matrix. This imbalance manifests itself as an unbounded growth in the error variance (Evensen, 1992). The Ensemble Kalman filter (EnKF) introduced by Evensen (1994) (and corrected by Burgers et al. (1998)) is primarily aimed at solving the problem of propagating the error covariance in time. Another practical problem with the EKF is that the adjoint of the model $\mathbf{F}$ is required for the implementation of equation (1.2b).

The core of the EnKF is the ensemble approach. Epstein (1969) suggested that the Liouville equation (1.3) can be solved more accurately than the approximate stochastic-dynamic prediction approach by using Monte Carlo methodology. Epstein proposed drawing Monte Carlo samples (called the *ensemble*) from ϕ and propagating each ensemble member forward in time under model dynamics. Then the probability distribution function defined by the propagated ensemble members at future time t gives the solution to the Liouville equation. The study by Leith (1974) confirmed that the ensemble approach could be used to obtain better forecasts even if the ensemble size, that is the number of Monte Carlo samples, is small. Since then, several researchers (Toth and Kalnay, 1993; Molteni et al., 1996) have employed the ensemble approach to probabilistic forecasting.

Evensen (Evensen, 1994) combined the ensemble approach to stochastic dynamic prediction with the Kalman filter methodology to arrive at a state estimation technique called the Ensemble Kalman filter (EnKF). The EnKF equations are given by (Evensen, 1994; Burgers

et al., 1998), for $i = 1, \ldots, N$,

$$\mathbf{x}_i^f(t) = \mathbf{F}[\mathbf{x}_i^a(t-1)] \tag{1.4a}$$

$$\mathbf{P}^f(t) = Cov(\mathbf{x}_i^f(t)) \tag{1.4b}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T[\mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t)]^{-1} \tag{1.4c}$$

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f(t) + \mathbf{K}(t)[\mathbf{y}_i^o(t) - \mathbf{H}(t)\mathbf{x}_i^f(t)] \tag{1.4d}$$

$$\mathbf{P}^a(t) = Cov(\mathbf{x}_i^a(t)) \tag{1.4e}$$

The model $\mathbf{F}$ in equation (1.4a) is non linear. The subscript i denotes the ith ensemble member and N is the ensemble size. At time $t-1$ the ensemble members $\mathbf{x}_i^a$ are sampled from the uncertainty given by $\mathbf{P}^a(t)$. Equation (1.4a) propagates each of the ensemble members from time t–1 to time t. The covariance of $\mathbf{x}_i^f(t)$ gives the forecast error covariance. Equation (1.4a), in propagating the ensemble members, also propagates the uncertainty. This overcomes the closure problem faced in the EKF. Since the full nonlinear model $\mathbf{F}$ is used to propagate each ensemble member, the higher order moments are not truncated in the propagation of uncertainty. This is the biggest advantage of the EnKF over the EKF. Equation (1.4c) calculates the Kalman gain factor similar to that in standard Kalman filter. The perturbed observations are denoted by $\mathbf{y}_i^o(t)$. These are obtained by adding perturbations drawn from $N(0, \sqrt{R})$ to $\mathbf{y}^o(t)$. Equation (1.4d) obtains the analysis by running the standard Kalman filter analysis procedure over each of the ensemble members. Finally equation (1.4e) gives an estimate of uncertainty in the analysis.

Apart from overcoming the closure problem in EKF, the EnKF offers other advantages. The EnKF does not need the construction of an adjoint model. It has been found that even when $N \ll n$ the EnKF provides good estimates of atmospheric state (Houtekamer et al., 2005). The calculation of $\mathbf{P}^f(t)$, which is a huge matrix, is not required. This is because only $\mathbf{P}^f(t)\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^f(t)\mathbf{H}^T$ are needed in the EnKF equations (Houtekamer and Mitchell, 2005). However, the EnKF has some drawbacks. The EnKF equations approximate the standard Kalman filter for a linear model for infinite sample size. In applications of EnKF, a small, finite sample size is used. A sampling error is introduced because of the finite sample size. The EnKF analysis assumes that the error statistics are Gaussian, which is an approximation. There has been a lot of research in the field of EnKF in the past decade. It appears that the EnKF is a promising technique to implement sequential data assimilation using the Kalman methodology in atmospheric, oceanic and other geophysical branches like hydrology.

A good review of the theory and practice of EnKF is given in Evensen (2005). A comprehensive discussion of EnKF and its applications is given in Evensen (2006). A comparison

of different data assimilation techniques including EnKF in the field of hydrology is given by McLaughlin (2002). A review of hydrologic data assimilation before the advent of EnKF is given by McLaughlin (1995). The implementation of EnKF in hydrology is described by Reichle et al. (2002). The problems in oceanographic state estimation and their solutions are described in Wunsch (1996).

The discussion of the standard Kalman filter, the EKF and EnKF was in the context of a perfect model. In reality, the models are always imperfect. If the model is imperfect (which is always the case), then the prediction of the system (the perfect model) is rendered wrong because of the imperfection in the initial condition and the imperfection in model. Throughout this work it is assumed that the boundary conditions are perfect. Weather and climate models are imperfect, that is, they do not accurately describe the system (i.e *nature*) they seek to represent. There are several sources for the imperfection in these models. Among these are wrong values of the parameters used, low spatial resolution of the model grid, parametrizations used to represent subgrid scale physical processes, approximations inherent in the numerical techniques, and structural imperfections in the model equations. The structural imperfection refers to wrong functional forms the model equations may have compared to the system equations. Broadly speaking, the model error can be divided into two categories, namely *parametric* and *structural* model errors. Model imperfections owing to wrong values of parameters are called parametric model error and all the other sources of model error (low resolution, parametrizations etc) could be put into the structural model error category. Model imperfection is also called *model inadequacy.* Typically, structural model error is far more serious than parametric model error. Of course, the importance of each type (parametric and structural) of model error in rendering the model imperfect is dependent on their respective magnitudes.

Given that the model is imperfect, we have to worry not only about uncertainty in forecast owing to initial conditions, but also that owing to the imperfection of the model. It appears that parametric model error is simpler to deal with than structural model error.

In section 1.2, I give an overview of the work done to address the issues discussed above.

1.2 Motivation and Overview

The work in this thesis is motivated by the following questions:

1. What are the implications of realistic prediction scenarios for linearized error dynamics vis-a-vis idealized prediction scenarios?

2. How can we mitigate the effect of model error on analyses and forecasts?

This work attempts to address these questions in the framework of ensemble-based filters. The predictability of a model is a fundamental question in the field of dynamical systems. The work done to answer question 1 explores various realistic conditions and their effect on the predictability. This work shows that the model error may be as important as initial condition error in rendering the forecast wrong. Given that geophysical models are imperfect the question 2 is a strong motivation for devising new techniques to address the problem of model error.

Some comments on the models used in this work are warranted. The ultimate aim of my work is to improve the performance of real weather and climate models. Even so, for part of this work, I have used low-dimensional chaotic models, namely the Ikeda system and the 1963 system of Lorenz (L63). This is because, low-dimensional models are easier to understand than high-dimensional atmospheric models and some aspects of these low-dimensional models may be analytically tractable. Hence innovative ideas on the research frontier can be tested out in these low-dimensional models. The level of initial condition error and model error can be varied at will in the low-dimensional models. We can also define the *truth* (the trajectory in state space defined by the perfect model along with the correct initial condition) in the low-dimensional model. The observations are generated "synthetically" by perturbing the truth trajectory. This guarantees that the observations are drawn from truth. In real atmospheric models a particular configuration of the model could be defined as the perfect model and a truth trajectory can be defined. But the observations recorded by real instruments may not be sampled from this truth trajectory. The characteristics of the observations (like observational uncertainty) can be easily changed for experimentation in the low-dimensional models. The performance of new approaches for improving the model prediction can be evaluated in the low-dimensional models. These low-dimensional systems need not represent any physical phenomenon related to the real atmosphere. But they provide us with an efficient test bed to evaluate new approaches. If the approaches are successful, then we can try to implement them in higher dimensional models. The results obtained in these low-dimensional models need not scale to high-dimensional models. But nonetheless, the results obtained with the low-dimensional models can be illuminating in that they tell us the "correct" answer. Whether we have enough information to obtain the "correct" answer in high-dimensional models is a separate question. Several authors have fruitfully used this approach (Hansen and Smith, 2001; Orrell et al., 2001; Hansen, 2002). At the same time, it is very important to devise techniques that can be used practically in real models. Therefore, I have used real atmospheric model (NOGAPS) for part of my work.

The issues raised by question 1 are investigated using linearized error dynamics. In general, predictability is defined as the divergence of nearby trajectories. The simplest quantification of predictability is given by the average of this divergence over the phase space. The divergence

of nearby trajectories is state-dependent and consequently predictability is state-dependent. Not surprisingly, it is found that estimates of the predictability of a system as given by linear error growth rates are vastly different under idealized conditions than under realistic conditions. The idealized conditions, like isotropic initial uncertainty and perfect model, give a misleading picture of the predictability of the underlying system. A procedure has been outlined to calculate the fraction of the initial condition error and model error in the total forecast error. In general, this fraction is state-dependent. The Ikeda system (Ikeda, 1979) is used for this work. Model error can have disastrous consequences for the prediction of a system and therefore we have to devise ways to handle the model error.

The work done in the context of model error addresses question 2. There are two different approaches or techniques to deal with model error. These are the *accounting for model error* and the *reduction of model error*. The basic difference between these two approaches is that in the reduction of model error approach, data are used to alter or correct the model while in the accounting for model error approach the model error information is used to suitably adjust the state (or to modify the uncertainty in the state). An example of accounting of model error is the use of model output statistics (Hansen and Emanuel, 2002).

The accounting technique is explored for a parametrically imperfect model. In this work, the *accounting for model error* approach involves including the model error information in the EnKF scheme so that the analysis and forecast are moved closer to the truth, consequently decreasing the analysis and forecast errors. Implicit in the accounting for model error approach is the self-imposed limitation that observations are not used to *correct* the model. We have some knowledge about the source of the model error, but we do not try to eliminate this source by altering the model. Rather we acknowledge that the model is wrong and try to account for the model error by incorporating this information in the EnKF. The approaches to accounting for model error in the EnKF are discussed. An approximation to account for model error in EnKF, namely the additive error approximation (AEA), is described. The AEA can be implemented only if we have the correct estimate of the state-dependent model error. The physical interpretation of the model error is given in terms of chaotic dynamics. The correct estimate of the model error is the sensitivity of the local attractor structure to model error, for a given state (Hansen, 2002). The *model error structure* is a multi-dimensional PDF whose dimensionality (say n) is that of the the model. If this PDF is assumed to be gaussian then it is visualized as an n-dimensional ellipsoid in phase space. Consequently, the *model error structure* is characterized by the *directions* of the n axes of the ellipsoid and their *sizes*. The multimodel method (MMM), introduced in this work gives a state-dependent estimate of model error. A physical interpretation of ensembles (in terms of chaotic attractor structure) has been pioneered by James Hansen (Hansen and Smith, 2001; Hansen, 2002). Recently, there has been a lot of interest in the weather and climate model community to use

multimodel approaches to improve forecasts (Krishnamurti et al., 2000). The models used by these researchers to construct the multimodel ensemble (MME) usually differ from each other in more than one way. The models could be different in resolution, parameter values, parametrizations, etc. It is very hard to interpret the sensitivity obtained using these MME. Therefore, I have used low-dimensional systems, namely Ikeda system and the Lorenz 1963 system (L63), to investigate the science and engineering of MME, where the models differ from each other only in the value of one parameter. A new method to estimate the state-dependent model error, the parametric vector method (PVM), is also introduced. The PVM is based on the linear sensitivity of the state to perturbations in parameters. The static method (SQM) is a method to obtain a state-independent estimate of the model error. Both the MMM and PVM involve interesting aspects of off-attractor (transient) and on-attractor dynamics. The MMM, SQM and PVM are intercompared in the framework of AEA using Ikeda and Lorenz 1963 models. The work about MMM follows immediately from the results and suggestions made in Hansen (2002).

The effect of model error on the analyses and forecasts can be mitigated by employing the other technique, namely the *reduction of model error*. The reduction of model error technique is investigated in the Ikeda model and the Navy Operational Global Atmospheric Prediction System (NOGAPS) model (Peng et al., 2004). *Parameter estimation* is a typical example of reduction of model error. Experiments with the Ikeda model illustrate the usefulness of parameter estimation to reduce both parametric and structural model error. Parameter estimation is successful in reducing the model error when the model is parametrically imperfect. Parameter estimation in the presence of structural model is termed *parameter tuning*. Parameter estimation basically involves augmenting the model state with parameters. In the Ikeda model, parameter tuning is able to partially offset structural model error. The state-independent (or global) tuning and state-dependent (or local) tuning is explored in the Ikeda model. The error introduced by parametrizations in real atmospheric models is a typical example of structural model error. The lessons learned from model error reduction in the Ikeda model are applied to the NOGAPS model. The structural model error in NOGAPS model is introduced by changing the diffusion coefficient for momentum in the vertical direction. The parameters in the convection parametrization are tuned to compensate for this structural model error. The NOGAPS model employs the Emanuel scheme (Emanuel, 1991) for convection parametrization. The state-dependent parameter tuning of two parameters in the Emanuel convection scheme is performed. This tuning results in a distribution of parameter values. This distribution changes with the level of structural model error introduced in the NOGAPS model. The change in the distribution of the parameters is consistent with the physical mechanisms of convection. This tuning is successful in partially compensating for the structural model error introduced by changing the diffusion coefficient.

The next section gives a brief outline of the chapters of the thesis.

1.3 Outline

Chapter 2 discusses linearized dynamics, chaotic systems and compares different indices of error growth. The structure of initial error and its impact on predictability is discussed. A comparison is drawn between theoretical idealizations and practical scenarios. The role of initial condition error and model error is mathematically treated.

Chapter 3 deals with the accounting for model error technique. The Additive Error Approximation (AEA) is presented. The multimodel method (MMM) and parametric singular vector method (PVM) to estimate state-dependent model error structure are introduced and explored. These are contrasted with the SQM which gives a state-independent estimate of the model error structure. The experiments carried out in the Ikeda and L63 models to intercompare the MMM, SQM and PVM are described and results are presented. The chaotic dynamics involved in MMM and PVM is delineated. The importance of state-dependent transient time in chaotic systems is explained.

Chapter 4 deals with the reduction of model error technique. It first describes and implements the parameter estimation/tuning technique in the Ikeda model. Then this technique is implemented in the NOGAPS model. This chapter presents a brief review of the NOGAPS model and the Emanuel scheme. Parameter estimation in the Ikeda model is considered for the cases when only parametric error is present and when only structural error is present. The implications of these results for the parametrization in atmospheric and oceanic prediction are emphasized.

Chapter 5 presents a summary of results and discusses further possible work.

Chapter 2

State-dependent predictability: Impact of uncertainty dynamics, uncertainty structure and model inadequacies

The predictability of a system is defined by the divergence of nearby trajectories with time. The distance (under some norm) between two trajectories at initial time is the error in the initial condition. The divergence of the trajectories is quantified by the growth of the initial condition error. In chaotic systems, an error in the initial condition grows, on average, exponentially fast. Any small but finite error in the initial condition tends to amplify rapidly, thus frustrating attempts to accurately forecast a chaotic system. Many *geophysical systems* appear to be chaotic.

One of the most popular diagnostics for quantifying the evolution of error in initial conditions for chaotic systems is the Lyapunov exponent (Oseledec, 1968). Several studies have used the Lyapunov exponent formulation to analyze the growth of error in dynamical systems (Smith et al., 1999). The Lyapunov exponent, as defined by Oseledec, gives the average growth rate of infinitesimal error over the whole state space, for infinite time. However, from a practical standpoint, this growth rate is not very useful, since in real life we are generally interested in short term forecasts of finite error in some localized part of state space. The concept of singular vectors and values (Strang, 1988) gives a more relevant measure for error growth than the Lyapunov exponent. E.N. Lorenz (Lorenz, 1965) pioneered the interpretation of singular vectors and values in the context of meteorology. The singular vectors are directions at a particular location in state space that experience maximum growth over a finite time interval. The singular vectors deal with localized dynamics and finite time intervals rather than global

dynamics and infinite time intervals.

2.1 Introduction

Much work on singular values and vectors in operational Numerical Weather Prediction (NWP) has been done. But the sensitivity of singular values and their distribution over state space to the shape of initial uncertainty and to model inadequacy remains by and large unexplored, though some work has been done (Barkmeijer et al., 1998). Isotropic uncertainty and perfect models are assumptions that can fail in the real world, resulting in misleading estimates of singular error growth.

In the current study, the state-dependent singular error growth over finite time has been studied in a 2-dimensional chaotic map, namely the Ikeda system (Ikeda, 1979). The Ikeda system has been chosen to allow visual demonstration of the impact of practical constraints on the distribution of the singular values. To begin with, the advantage of singular values over Lyapunov numbers is demonstrated. Next, the impact of the above mentioned simplifications (isotropic initial uncertainty and perfect model) on singular error growth is analyzed.

Section 2.2 outlines the dynamical system employed, namely the Ikeda map. It also introduces the idea of linearization and the linear propagator. Section 2.3 delineates the theoretical framework of Lyapunov and singular values. It is meant to introduce the mathematics behind these concepts. The concepts presented in section 2.3 are applied to the Ikeda system in section 2.4. Section 2.4 explains the methodology and some results of this study. Section 2.5 explores the issue of relative importance of the initial state error and the model error in determining the final state error. Finally, section 2.6 summarizes the results.

The discussion and results in sections 2.2 through 2.4 has been published in Khade and Hansen (2004).

2.2 Ikeda System and linearization

Though the exact nature of initial error growth will depend on the system under consideration, some of its generic features can be understood and possible issues elucidated by studying low-dimensional chaotic systems. Of course, the solutions obtained by using these systems need not

scale to high-dimensional and more complex models, but the generic results obtained can be illuminating. The Ikeda system has been used by several authors (Smith et al., 1999; Hansen and Smith, 2001) to elucidate predictability studies. All work in this chapter is restricted to the Ikeda system. A description of the Ikeda system and its linearization is given below.

The equations of the Ikeda system are

$$x_{s+1} = 1 + \mu(x_s \cos\theta - y_s \sin\theta) \tag{2.1a}$$

$$y_{s+1} = \mu(x_s \sin\theta + y_s \cos\theta) \tag{2.1b}$$

where

$$\theta = a - \frac{b}{(x_s^2 + y_s^2 + 1)}$$

$a = 0.4$, $b = 6$, $\mu = 0.9$ and s denotes the step number. Note that this system is a *map* as opposed to a *flow*. A map is discrete in time while a flow eg. the Lorenz 1963 system (Lorenz, 1963), is continuous.

Linearization of the Ikeda system is a crucial concept since singular and Lyapunov error growth hinge on its validity. Linearization can be explained as follows. The state space is defined by (x, y). Consider a point (x_s, y_s) in the phase space. Consider the perturbed point given by $(\tilde{x}_s, \tilde{y}_s) = (x_s + \epsilon_{x_s}, y_s + \epsilon_{y_s})$. $\epsilon_{x_s} = \tilde{x}_s - x_s$ and $\epsilon_{y_s} = \tilde{y}_s - y_s$ denote the perturbations (or the errors). The evolution of (x_s, y_s) and $(x_s + \epsilon_{x_s}, y_s + \epsilon_{y_s})$ is described by the nonlinear equations (2.1a) and (2.1b). Since x_{s+1} and y_{s+1} are functions of x_s and y_s, one can introduce the notation,

$$x_{s+1} = F_1(x_s, y_s)$$
$$y_{s+1} = F_2(x_s, y_s).$$

Using the Taylors series the evolution of linearized error is given by

$$\begin{bmatrix} \epsilon_{x_{s+1}} \\ \epsilon_{y_{s+1}} \end{bmatrix} = \begin{bmatrix} F'_{1x}(x_s, y_s) & F'_{1y}(x_s, y_s) \\ F'_{2x}(x_s, y_s) & F'_{2y}(x_s, y_s) \end{bmatrix} \begin{bmatrix} \epsilon_{x_s} \\ \epsilon_{y_s} \end{bmatrix}. \tag{2.2}$$

where

$$F'_{1x} = \frac{\partial F_1}{\partial x_s} \text{ etc.}$$

The matrix on the rhs of equation (2.2) (which is the Jacobian of equations (2.1a) and (2.1b))

is called the linear propagator,

$$\mathbf{M} = \begin{bmatrix} F'_{1x}(x_s, y_s) & F'_{1y}(x_s, y_s) \\ F'_{2x}(x_s, y_s) & F'_{2y}(x_s, y_s) \end{bmatrix}.$$

Suppose the error at the s^{th} step is

$$\boldsymbol{\epsilon}_s = \begin{bmatrix} \epsilon_{x_s} \\ \epsilon_{y_s} \end{bmatrix}.$$

Then the error at the $(s+1)^{th}$ step is approximately

$$\boldsymbol{\epsilon}_{s+1} = \mathbf{M}\boldsymbol{\epsilon}_s \tag{2.3}$$

assuming that the magnitude of $\boldsymbol{\epsilon}_s$, denoted by $\|\boldsymbol{\epsilon}_s\|$ is small enough so that the linearity assumption holds. Though this relation holds approximately for finite perturbations, it holds exactly for infinitesimal perturbations. Equation (2.3) is the linearization of the system equations about a nonlinear trajectory.

The number of steps (or time in the case when the system is a flow) over which $\mathbf{M}$ is constructed is called the optimization time τ. $\mathbf{M}$ is also called the *tangent linear model*. $\mathbf{M}$ depends on the initial conditions and the optimization time τ, that is, $\mathbf{M} = f(x_s, y_s, \tau)$. The tangent linear model is particularly useful because of the property (which is mathematically expressed by equation (2.3)) which states that the linearized error at the n^{th} step can be evaluated by knowing the initial error and the linear propagator over each of the intermediate steps. The validity of this linearization (that is, the proximity of the linear error evolution to the nonlinear error evolution) depends on the magnitude and direction of the initial error and τ. In general, for a given τ, the smaller the initial error, the better the validity. The number of steps over which linearity holds has to be ascertained before the linear propagator is actually utilized. In the current work the initial magnitude of errors has been chosen small enough so that linearity holds for the number of steps over which $\mathbf{M}$ is constructed.

2.3 Indices of error growth

A finite error in initial conditions ultimately grows nonlinearly. One needs to define *indices* of error growth to quantify this error growth. Lyapunov numbers and singular values are such indices of error growth, defined under linearized dynamics. These accurately quantify

the nonlinear growth of errors insofar as linearization holds.

2.3.1 Lyapunov numbers

Lyapunov vectors and numbers have been discussed in many papers. The following discussion has been adapted from Ziehmann et al. (1998).

Consider an m-dimensional state space. The Lyapunov exponents are defined assuming infinitesimal perturbations at time $t=0$. As discussed in section 2.2, the linear propagator $\mathbf{M}$ evolves forward the initial error (or perturbation) linearly along the nonlinear trajectory over time τ. Consider the matrix,

$$\mathbf{O}(\mathbf{x},\tau) = [\mathbf{M}(\mathbf{x},\tau)^T\mathbf{M}(\mathbf{x},\tau)]^{1/2\tau}$$

where the superscript T denotes the transpose of the matrix. Oseledec (Oseledec, 1968) showed that if the limit of $\tau \to \infty$ exists, then the m eigenvalues say p_i , $i = 1 \ldots m$ of,

$$\mathbf{O}(\mathbf{x}) = \lim_{\tau\to\infty} \mathbf{O}(\mathbf{x},\tau)$$

are independent of the state $\mathbf{x}$. Then the Lyapunov exponents are defined as

$$\Lambda_i = \frac{1}{\tau}\ln(p_i)$$

as $\tau \to \infty$ for $i = 1 \ldots m$. By convention, $\Lambda_i > \Lambda_j$ for $i < j$.

The leading Lyapunov number is defined as

$$L_1 = e^{(1/\tau)\Lambda_1} = p_1.$$

L_1 gives the average one step *factor* of growth of the fastest growing direction. i.e. if Λ_1 has been calculated for a finite time $\tau = k\Delta t$, say where k is the number of steps and Δt is the time step, then at each step the error grows by a factor of L_1, so that at the end of k steps it has grown by a factor of $L_1^{k\Delta t}$.

Figure 2-2 (page 31)(a) shows the one step growth factor as given by the Lyapunov number. The points shown are on the attractor and each one is colored by the single leading Lyapunov number. According to this picture, an error at any point in the state space would grow by a factor of 1.66 over $\tau=1$. In general it would grow by a factor of 1.66^τ over τ steps. Note that by definition the error growth given by the Lyapunov number is uniform over the whole state space.

2.3.2 Singular vectors and values

The subject of singular vectors has been treated by numerous authors in linear algebra texts and scientific literature (Strang, 1988). The singular vectors and values are defined for the linearized system of equations. Having calculated $\mathbf{M}$ over optimization time τ, the singular vectors and values are defined as follows.

$$\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{2.4}$$

where $\boldsymbol{\Sigma} = \sqrt{\boldsymbol{\Lambda}}$.

$\mathbf{U}$ and $\mathbf{V}$ are called the left (or final time) and right (or initial time) singular vectors respectively. The elements of $\boldsymbol{\Lambda}$, which is a diagonal matrix, are the square of the singular values of $\mathbf{M}$. These can be obtained by either the Singular Value Decomposition i.e. $SVD(\mathbf{M})$ or through the eigenvectors and eigenvalues of $\mathbf{M}^T\mathbf{M}$ and $\mathbf{M}\mathbf{M}^T$. Note that $\mathbf{U}$ and $\mathbf{V}$ define orthonormal bases. The Singular Value Decomposition will be referred to as SVD, for brevity.

In the context of the Ikeda system, given a linear propagator constructed over a particular τ, $\mathbf{U}$ is 2×2, $\mathbf{V}$ is 2×2 and there are 2 singular values. From equation (2.4),

$\mathbf{M}\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i$ where $i=1,2$.

The singular values can be written as,

$$\sigma_i = \frac{\|\mathbf{M}\boldsymbol{v}_i\|}{\|\boldsymbol{v}_i\|} \quad \text{where } \sigma_1 > \sigma_2.$$

The singular values σ_1 and σ_2 give the ratio of magnitudes of vectors aligned along $\boldsymbol{v}_i$ at initial time to those aligned along $\boldsymbol{u}_i$ at final time. The interpretation of these vectors and values is of immense importance from the standpoint of error growth under linearized dynamics. Given an isotropic (that is, having the same magnitude in all directions) distribution of uncertainty at initial time, the errors aligned along $\boldsymbol{v}_1$ grow largest and at optimization time align along $\boldsymbol{u}_1$ (Note that this assumes a norm, which is discussed in more detail in the section 2.3.3). The corresponding growth factor is given by σ_1. The errors aligned along $\boldsymbol{v}_2$ grow the least with the corresponding growth factor given by σ_2. All other directions grow by factors that are between σ_1 and σ_2. The leading singular value places an upper limit on the prediction of the system under linearized dynamics given an isotropic uncertainty. It provides us with an estimate of how much maximum error one can expect in the forecasts.

In the case of n dimensions, the n singular values give the n growth factors associated with the n singular directions which form an orthonormal basis. The isotropic n-dimensional sphere

evolves into an ellipsoid at final time whose n axes are given by the $\mathbf{U}$ vectors.

2.3.3 Norms

There are several ways in which *distance* can be measured between two points in a phase space. This distance is defined by a function known as the norm (Lipschutz, 1991). A given uncertainty distribution may be isotropic with respect to one norm, while being non-isotropic with respect to another. Thus it is very important to specify the norm while discussing uncertainty structure.

Let $\mathbf{X}$ be a real or complex vector space. Suppose there is a function $\|\mathbf{x}\|$ which assigns to each $\mathbf{x} \in \mathbf{X}$ a real number. This function is called a *norm* on $\mathbf{X}$ if (for $\mathbf{x}, \mathbf{y} \in \mathbf{X}$) it satisfies the following axioms:

$$\|\mathbf{x}\| \geq 0; \text{ and } \|\mathbf{x}\| = 0 \text{ iff } \mathbf{x} = 0 \tag{2.5a}$$

$$\|k\mathbf{x}\| = k\|\mathbf{x}\| \text{ where } k \text{ is a scalar} \tag{2.5b}$$

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \tag{2.5c}$$

The vector space $\mathbf{X}$ with a norm is called a *normed vector space.* There can be several functions which define different norms as long as they satisfy the above axioms. For instance, if $x_1, x_2, ..., x_n$ are components of a vector $\mathbf{x}$, then three possible norms are,

$$\|\mathbf{x}\| = max(|x_i|), i = 1 \ldots n$$

$$\|\mathbf{x}\| = |x_1| + |x_2| + ... + |x_n|$$

$$\|\mathbf{x}\| = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2}$$

These are called the L_∞, L_1 and L_2 (or Euclidean) norms respectively. The distance between two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{X}$ denoted by the function $d(\mathbf{x}, \mathbf{y})$ is dictated by the norm employed. If the L_2 norm is used then,

$$d(x, y) = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2 + ... + |x_n - y_n|^2}$$

It appears that the terms *norm* and *metric* have been used synonymously in the meteorology literature (Palmer et al., 1998). Predictability studies using singular vector analysis involves maximizing initial (i.e. analysis) perturbations (i.e. errors) over the optimization time. There are several choices available for the *physical dimension* of these perturbations, like total energy, enstrophy, streamfunction variance, etc. The state vectors can be operated on by appropriate operators to obtain the perturbations in a particular *physical dimension.* The operator

chosen defines a norm on the space (L_2 norm, energy norm, enstrophy norm, etc. (Palmer et al., 1998). In general, irrespective of the norm chosen, the analysis error structure may be non-isotropic. $SVD(\mathbf{M})$ where $\mathbf{M}$ is the linear propagator does not give the singular directions and values which are relevant to predictability, since the non-isotropic nature of the analysis error structure is not taken into consideration. This drawback can be overcome by using the covariance, say $\mathbf{P}^a$, of the analysis errors in the chosen norm's space. If $\sqrt{\mathbf{P}^{a-1}}$ is used to transform the analysis errors then the resulting analysis errors are isotropic. Thus $SVD(\mathbf{M}\sqrt{\mathbf{P}^{a-1}}^{-1})$ gives the singular directions and values relevant to predictability, because it accounts for situations in which large initial uncertainty (i.e. error) combined with small error growth factors lead to larger final time errors than small initial uncertainty combined with large error growth factors. Using $\sqrt{\mathbf{P}^{a-1}}$ as a transform is equivalent to using $\mathbf{P}^{a-1}$ as a norm. $\mathbf{P}^{a-1}$ is called the analysis error covariance norm (Barkmeijer et al., 1998). It is also called the Mahalanobis metric (Palmer et al., 1998). Usually, in realistic atmospheric models, it is numerically very expensive to calculate $\mathbf{P}^{a-1}$. Barkmeijer et al. (1998) explores some methods of obtaining the estimates of $\mathbf{P}^{a-1}$. It so happens that if perturbations are measured in energy then the analysis error distribution appears to be isotropic (Palmer, 1994), which is advantageous since then $\mathbf{P}^{a-1}$ need not be calculated. In the work that follows, the L_2 norm is used and $\sqrt{\mathbf{P}^{a-1}}$ is used to transform the non-isotropic analysis error distribution to an isotropic one.

2.4 Singular vectors and values in the Ikeda system

The concept of singular vectors and values presented in section 2.3 has been applied to the Ikeda system. Basically, the question addressed is, do the (leading) singular values change over the state space? This represents the state dependence of error growth under numerous operational constraints.

The concept of singular error growth in the Ikeda system is illustrated with the following example. Figure 2-1 (page 29)(a) illustrates singular error growth at a particular point (x_s, y_s) in the Ikeda phase space. The point (x_s, y_s) is on the system attractor, that is, it is the truth. A gaussian uncertainty distribution centered at (x_s, y_s), is generated at the initial time. The perturbations to be evolved by $\mathbf{M}$ (not shown in the figure) are sampled from this distribution. The blue circle in figure 2-1(a) shows the 2ξ bound of this gaussian distribution where ξ stands for standard deviation. (x_s, y_s) lies at the center of this blue circle. This distribution is propagated over a total optimization time of $\tau = 3$ using the tangent linear model. The propagated distribution, the singular vectors and values are saved at each time step.

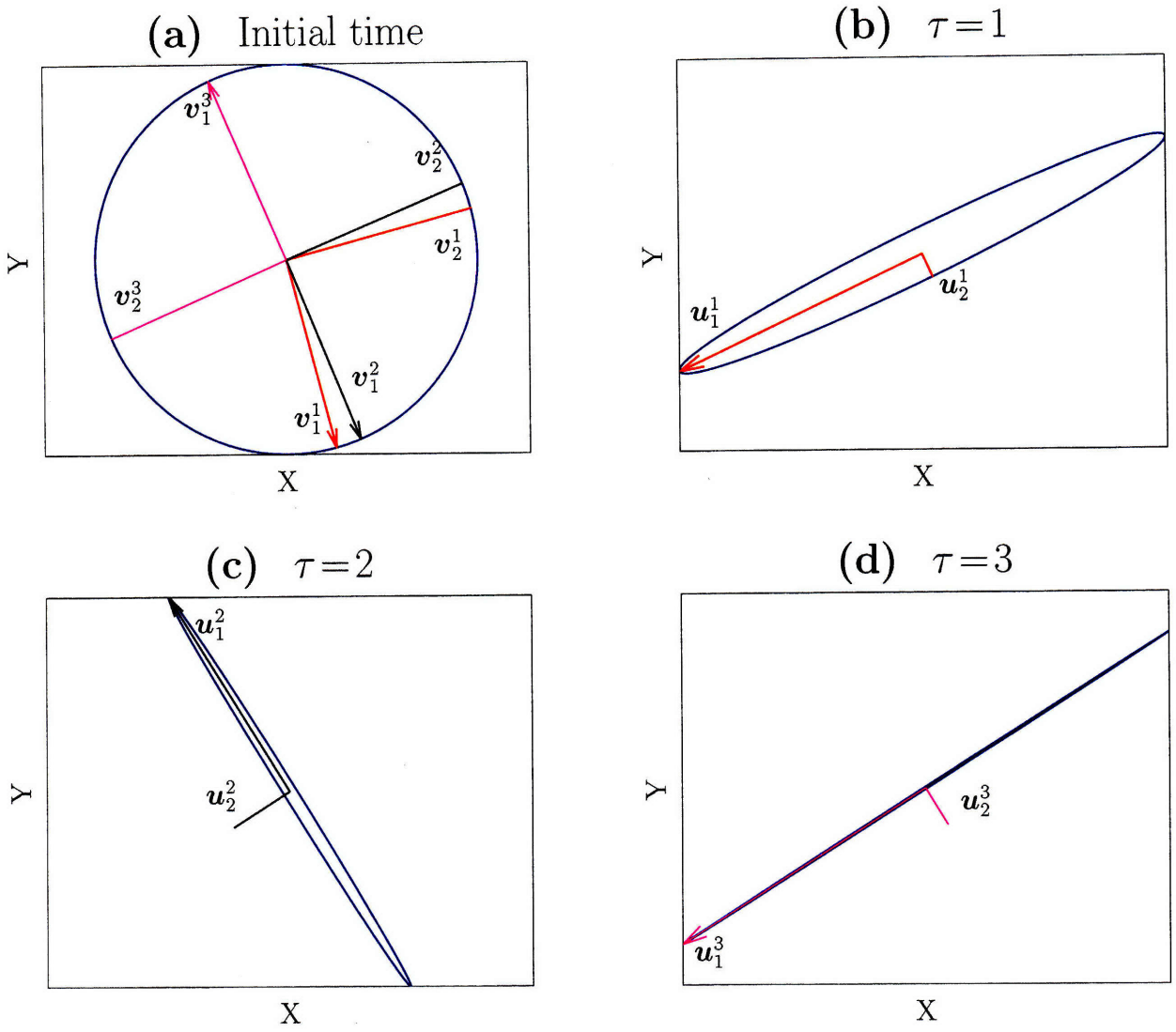


Figure 2-1: The evolution of an isotropic distribution over 3 time steps.
(a) Isotropic distribution (blue circle) at initial time and the **V** vectors.
(b) $\sigma_1=2.95, \quad \sigma_2=0.27$
(c) $\sigma_1=4.75, \quad \sigma_2=0.13$
(d) $\sigma_1=12.05, \quad \sigma_2=0.04.$

At each step the 2ξ bound of the ensemble, centered at the propagated truth, is plotted. The tangent linear model stretches (or shrinks) and rotates each of the perturbation vectors so that at each step an ellipse is generated. Figure 2-1(b), figure 2-1(c) and figure 2-1(d) show the ellipses at $\tau=1$, $\tau=2$ and $\tau=3$ respectively. First, one is interested in the direction in the future which has the maximum variance at the forecast time (or optimization time τ). The variance in this direction relative to the initial variance can be regarded as a quantitative measure of predictability. Second, one is interested in knowing the direction at initial time that grows into direction of maximum variance at τ. This knowledge can be used to help one to reduce the error in this particular initial direction so that forecast errors at final time can

be reduced.

The $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}$ are calculated by the SVD of $\mathbf{M}$ constructed over each τ, using the L_2 norm on the space and an isotropic distribution of uncertainty at initial time. The major and minor axes of the final time ellipses are given by $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, respectively. The initial directions that evolve into these directions at final time are given by $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, respectively. Note that each time step has an unique set of $\mathbf{U}$ and $\mathbf{V}$ vectors with the corresponding growth factors given by the σ's. Hence, for instance, over one step any error aligned along $\boldsymbol{v}_1^1$ direction in figure 2-1(a) (shown in red, with arrowhead) will increase in magnitude by 2.95 (which is the maximum σ over $\tau = 1$) while that aligned along $\boldsymbol{v}_2^1$ direction (shown in red, without the arrowhead) will shrink by a factor of 0.27 (which is the minimum σ over $\tau = 1$). Note that the superscripts indicate the optimization time τ while the subscripts give the index of the singular vector (subscript 1 indicates leading singular vector). At the final time, over $\tau = 1$, these directions will align themselves along $\boldsymbol{u}_1^1$ and $\boldsymbol{u}_2^1$. These final time directions for $\tau = 1$ are shown in figure 2-1(b). For $\tau = 2$ an error aligned along the direction shown in black (with arrowhead) in figure 2-1(a), will grow the most (by a factor of 4.75) and align itself along the direction $\boldsymbol{u}_1^2$, shown in figure 2-1(c). Similarly the directions shown in magenta color in figure 2-1(a), evolve to those in figure 2-1(d) over $\tau = 3$, with the maximum growth factor of 12.05 and a minimum of 0.04. These singular values, calculated by assuming an *isotropic* initial uncertainty distribution, are called the *isotropic* or the *plain* singular values. As $\tau \rightarrow \infty$ the $\mathbf{U}$ converges to the same vector. Similarly, as $\tau \rightarrow \infty$ the $\mathbf{V}$ converges to the same vector.

Thus the singular vector formulation is very remarkable in that it gives (1) the directions at initial time that would grow the most over a specified τ (2) the directions to which they would grow and (3) the factors by which they grow. The leading singular value represents the worst case scenario for forecast error under the linearization assumption.

2.4.1 Isotropic initial uncertainty

The procedure of calculating the leading singular value described in section 2.4 is applied to a number of points in the Ikeda state space so that its variation can be inspected. This motivates the superiority of singular values over the Lyapunov number in so far as finite time forecast error studies are concerned. Consider figure 2-2 (page 31)(a). The *average over space via an average over time* implicit in the definition of the Lyapunov number, gives this uniform distribution, which belies the true nature of finite time error growth, state-dependent growth as is evident in figure 2-2(b). Figure 2-2(b), shows the leading singular value for $\tau = 1$ at many points, as calculated assuming isotropic initial uncertainty. This work has been done earlier (Smith et al., 1999); but has been presented as a background to the results to follow. The L_2 norm, as defined in section 2.3.3, has been used to measure distances in the phase

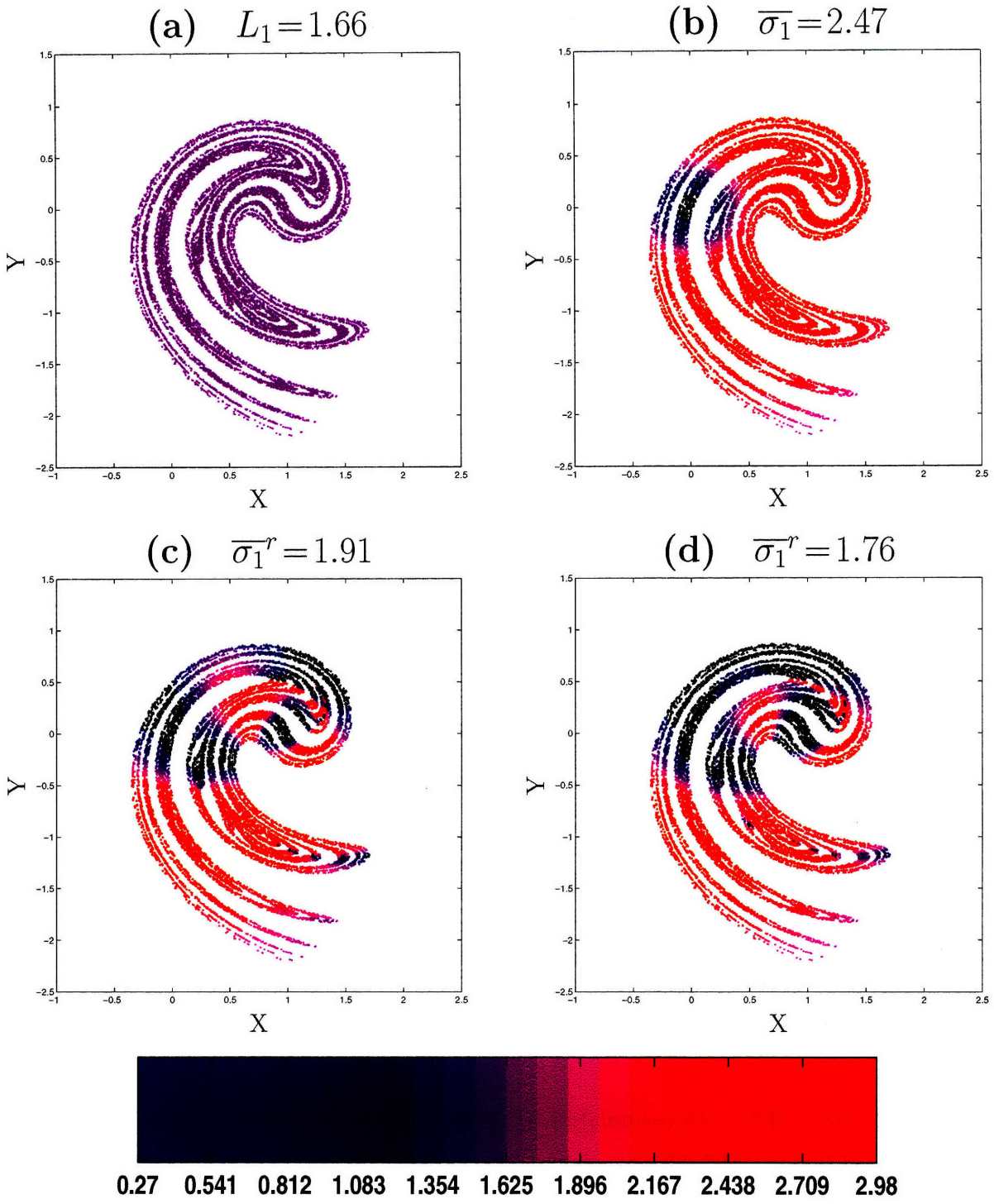


Figure 2-2: The distribution of the different indices of error growth for the perfect Ikeda model. Points in black have $\sigma_1 < 1$. **(a)** The leading Lyapunov number. The error growth is uniform over the attractor. **(b)** The leading singular value calculated using $SVD(\mathbf{M})$ i.e. by assuming that the initial uncertainty distribution is isotropic. **(c)** The relevant leading singular value calculated using $SVD(\mathbf{M}\sqrt{\mathbf{P}^{a-1}}^{-1})$. $\mathbf{P}^a$, in general is non-isotropic. **(d)** The relevant leading singular value if $\mathbf{P}^a$ is rotated by an angle θ drawn from $N(0, 20°)$.

space. The error growth factor as given by the singular value in figure 2-2(b) shows significant structure and variation. The errors in some parts of the state space could grow by as much 2.97 while in other parts they could actually shrink (these areas having $\sigma_1 < 1$ are colored in black). A comparison of figure 2-2(a) and figure 2-2(b) shows that the Lyapunov number is in fact quite a dubious measure of finite time error growth. Note that the Lyapunov number has the same value of 1.66 everywhere in the state space. In areas like those around the origin $(0, 0)$, $\sigma_1 < 1$, and consequently the forecast launched from this region for $\tau = 1$ would be very good, as is evident from figure 2-2(b). The Lyapunov number gives a false impression that the forecast will be poor in this region. On the other hand, in regions like those around $(1, 0.5)$ the Lyapunov number artificially enhances the forecast quality, while in reality the forecast will be quite poor as is evident from figure 2-2(b). The Lyapunov number is all the more misleading for longer optimization times (results not shown). For instance, over 2 steps the Lyapunov factor would be 1.66^2 while it is seen that singular value varies quite a lot over the state space. This comparison gives cogent reasons to decisively discard the Lyapunov number as a measure of finite time error growth and adopt the singular value as a far better substitute. The error growth as given by the singular value is state-dependent and forecast time-dependent. The information given by the singular values not only diagnoses the forecast quality over state space, but also could be useful for data assimilation or targeting. Thus if the initial distribution of errors is isotropic (in the L_2 norm on the space), then figure 2-2(b) indeed gives a genuine picture of forecast quality.

The discussion in this section assumed an isotropic initial uncertainty distribution. If data assimilation is used to generate the initial uncertainty distribution then the uncertainty distribution is rendered non-isotropic with respect to the L_2 norm. The next subsection (2.4.2) gives a brief overview of data assimilation using the Ensemble Kalman filter (EnKF).

2.4.2 Data assimilation

In this work, the EnKF data assimilation (Evensen, 1994; Houtekamer and Mitchell, 1998) will be used to produce the analysis. The EnKF equations can be described as follows. For $i = 1 \ldots N$,

$$\mathbf{x}_i^f(t) = \mathbf{F}[\mathbf{x}_i^a(t-1)] \tag{2.6a}$$

$$\mathbf{P}^f(t) = \frac{1}{N-1}[\mathbf{E}^f(t) - \mathbf{M}^f(t)][\mathbf{E}^f(t) - \mathbf{M}^f(t)]^T \tag{2.6b}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T[\mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t)]^{-1} \tag{2.6c}$$

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f(t) + \mathbf{K}(t)[\mathbf{y}_i^o(t) - \mathbf{H}(t)\mathbf{x}_i^f(t)] \tag{2.6d}$$

$$\mathbf{P}^a(t) = \frac{1}{N-1}[\mathbf{E}^a(t) - \mathbf{M}^a(t)][\mathbf{E}^a(t) - \mathbf{M}^a(t)]^T \tag{2.6e}$$

In these equations, N is the number of ensemble members, n is the dimension of the system and $\mathbf{x}_i^f(t)$ is the first guess generated by a short term forecast. The nonlinear model is represented by the operator $\mathbf{F}$ in equation (2.6a). $\mathbf{E}^f(t)$ and $\mathbf{E}^a(t)$ are $n \times N$ matrices which give the ensembles before and after assimilation. Thus each column of $\mathbf{E}^f(t)$ and $\mathbf{E}^a(t)$ is an ensemble member i.e. a different realization of the state and each row is a state component. $\mathbf{M}^f(t)$ and $\mathbf{M}^a(t)$ are $n \times N$ matrices having the corresponding ensemble mean of $\mathbf{E}^f(t)$ and $\mathbf{E}^a(t)$ respectively, in each of its columns. The mean is calculated by averaging over each row of $\mathbf{E}^f(t)$ and $\mathbf{E}^a(t)$. Thus a particular element in each column of $\mathbf{M}^f(t)$ (which is the same as each of other N columns) is an average over the corresponding row in $\mathbf{E}^a(t)$. $\mathbf{P}^f(t)$ (dimension $n \times n$) gives the uncertainty in the first guess in the form of its covariance matrix. $\mathbf{R}(t)$ (dimension $n \times n$, assuming all the state variables are observed) gives the uncertainty in the observations. The Kalman gain term $\mathbf{K}(t)$ (dimension $n \times n$) gives measures of confidence one should place in the first guess and the observations depending on their respective uncertainties. $\mathbf{H}(t)$ is a map from the model space to observation space. $\mathbf{y}_i^o(t)$ are the perturbed observations. $\mathbf{x}_i^a(t)$ is called analysis. $\mathbf{x}_i^a(t)$ combines the first guess and the observation, thus providing the best estimate of the initial condition. Finally $\mathbf{P}^a(t)$, is the covariance matrix that provides an estimate of uncertainty in the analysis. The equations of EnKF data assimilation methodology presented in equations (2.6a) through (2.6e) assume a perfect model, that is, the step forward equations contained in $\mathbf{F}$ are perfect.

NWP centers employ data assimilation to produce the best possible initial conditions. It is difficult to imagine that the initial condition error produced operationally by data assimilation is isotropic. This is because the first guess is produced by propagating the model state forward and the evolved uncertainty distribution will be elliptical. It would take a very special observational uncertainty distribution to render the analysis uncertainty distribution isotropic. Hence, if one wants to carry out the singular analysis in a realistic scenario, then a non-isotropic error distribution has to be assumed rather than an isotropic one (with respect to the L_2 norm in this case).

2.4.3 Impact of non-isotropic distribution

As stated above, when data are assimilated, the initial uncertainty distribution does not remain isotropic. The question then is whether the directions given by the $SVD(\mathbf{M})$ still give the *relevant* singular values and initial and final time singular vectors. The methodology to find these *relevant* values is discussed below and illustrated in figure 2-3 (page 35).

Figure 2-3(a) shows an isotropic (blue circle) and a non-isotropic (green ellipse) distribution of uncertainty at initial time (the circle and ellipse shown are the 2ξ bounds for the ensembles). The initial time singular vectors as given by the $SVD(\mathbf{M})$ satisfy,

$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ so that,

$\mathbf{M}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$ where $i = 1, 2$

The uncertainty distribution is propagated over one step ($\tau = 1$). The resulting distribution (2ξ bound) is shown in figure 2-3(c). The blue arrows, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ in figure 2-3(a) are the initial time singular vectors for the isotropic distribution (that is, by using $SVD(\mathbf{M})$). The solid blue line with arrow head is the leading $\boldsymbol{v}_1$ and the dashed blue line is $\boldsymbol{v}_2$. The corresponding singular values are given by $\sigma_1 = 2.59$ and $\sigma_2 = 0.18$. Any errors (whether in the isotropic or non-isotropic distribution) aligned along $\boldsymbol{v}_1$ in figure 2-3(a) will grow by a factor of 2.59 and align themselves along $\boldsymbol{u}_1$ in figure 2-3(c). Accordingly, the blue ellipse in figure 2-3(c) corresponding to the initial isotropic distribution in figure 2-3(a) has its major and minor axis along the $\mathbf{U}$ vectors. But the initial time non-isotropic distribution (green) ellipse in figure 2-3(a) *does not* have its major and minor axis along $\mathbf{U}$ vectors at final time in figure 2-3(c). In fact the major axis of the green ellipse in figure 2-3(c) is aligned along a different direction labeled $\tilde{\boldsymbol{u}}_1$.

The variance that is maximum along this direction ($\tilde{\boldsymbol{u}}_1$) characterizes the operationally relevant predictability, rather than that along $\boldsymbol{u}_1$. The next question is, what is the direction at initial time in figure 2-3(a) that evolved into the $\tilde{\boldsymbol{u}}_1$ direction at final time? The procedure to find these directions is delineated below.

First, the non-isotropic distribution (green ellipse) in figure 2-3(a) is operated on by $\sqrt{\mathbf{P}^{a-1}}$ where $\mathbf{P}^a$ is the covariance matrix of the non-isotropic distribution. This operation converts the non-isotropic distribution to an isotropic distribution. This isotropic distribution is shown in figure 2-3(b) as the red circle. The red circle is 2ξ bound.

The singular vectors and values are then given by the $SVD(\mathbf{M}\sqrt{\mathbf{P}^{a-1}}^{-1})$. Thus,

$\mathbf{M}\tilde{\boldsymbol{v}}_i = \tilde{\sigma}_i \tilde{\boldsymbol{u}}_i$ where $i = 1, 2$

Hence the initial direction (in the transformed space) that grows most over $\tau = 1$ is given by $\tilde{\boldsymbol{v}}_1$ (shown in figure 2-3(b) by the solid red arrow). But this direction is valid only in the transformed space. The corresponding direction in the original space is given by $\tilde{\tilde{\boldsymbol{v}}}_1 = (\sqrt{\mathbf{P}^{a-1}}^{-1})\tilde{\boldsymbol{v}}_1$.

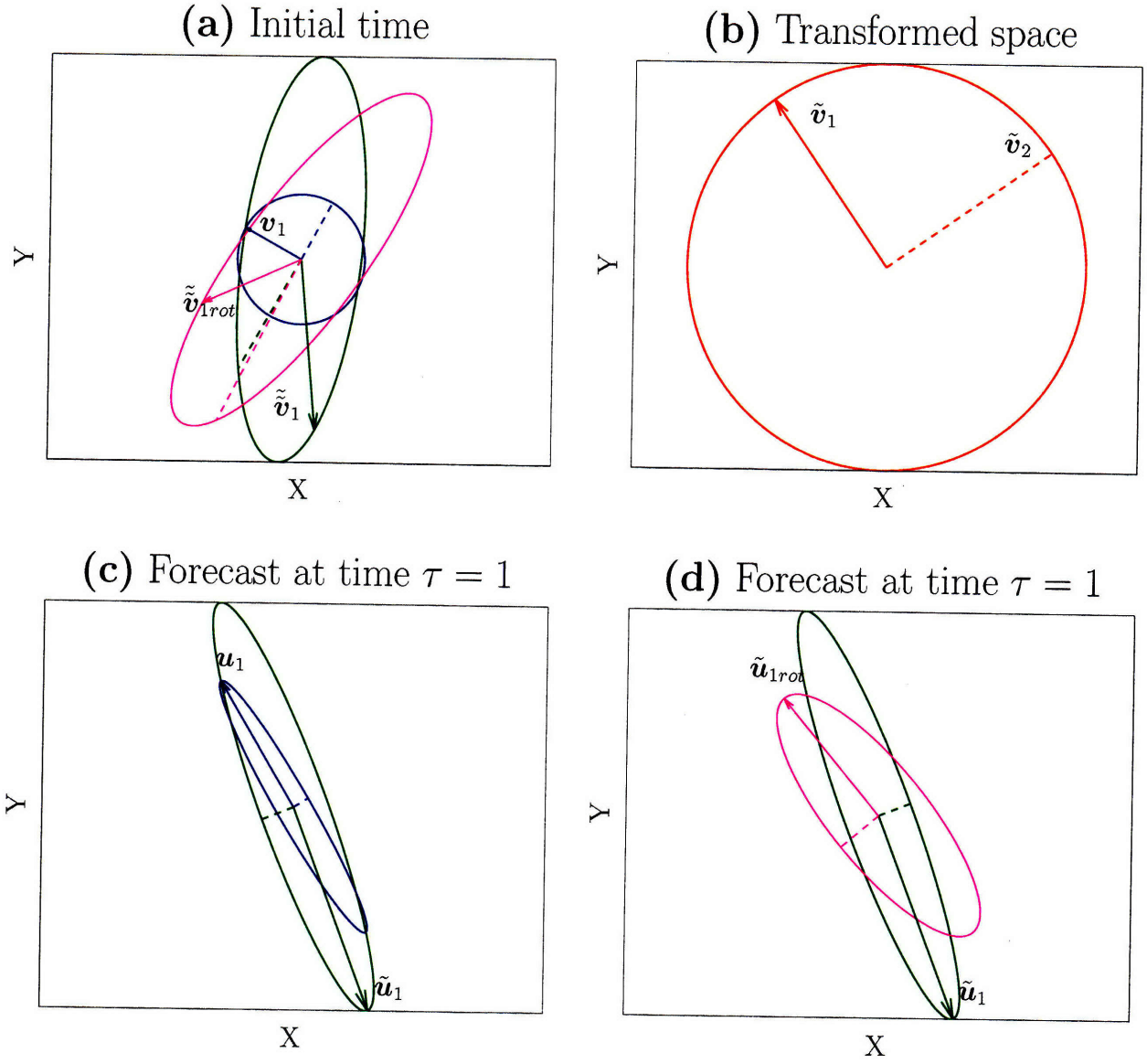


Figure 2-3: The evolution of a non-isotropic distribution. The non-isotropic distribution (green) in (a) evolves to the green ellipse in (c). The isotropic distribution (blue) in (a) evolves into the blue ellipse in (c). The magenta ellipse in (a) which is the rotated version of the green ellipse in (a), evolves to the magenta ellipse in (d). The green ellipse in (c) is reproduced in (d).

(a) $\|\boldsymbol{v}_1\| = 0.07, \quad \sigma_1 = 2.59$
$\|\tilde{\tilde{\boldsymbol{v}}}_1\| = 0.18, \quad \sigma_1^r = 1.47$
$\|\tilde{\tilde{\boldsymbol{v}}}_{1rot}\| = 0.11, \quad \sigma_{1rot}^r = 1.60$

(b) The green ellipse in (a) is the red circle in the transformed space.

(c) $\|\boldsymbol{u}_1\| = 0.07 \times 2.59 = 0.18$
$\|\tilde{\boldsymbol{u}}_1\| = 0.18 \times 1.47 = 0.26$

(d) $\tilde{\boldsymbol{u}}_{1rot} = 0.11 \times 1.60 = 0.17$

It follows that

$$\sigma_1^r = \frac{\tilde{\sigma}_1}{\|\tilde{\tilde{\boldsymbol{v}}}_1\|} \tag{2.7a}$$

$$\sigma_2^r = \frac{\tilde{\sigma}_2}{\|\tilde{\tilde{\boldsymbol{v}}}_2\|} \tag{2.7b}$$

In equations (2.7a) and (2.7b) the superscript r stands for *relevant*. The relevant singular values are also called the normed singular values.

The $\tilde{\tilde{\boldsymbol{v}}}_1$ and $\tilde{\tilde{\boldsymbol{v}}}_2$ are shown as solid green and dashed green arrows respectively in figure 2-3(a). Note that they are not orthogonal. These directions can be interpreted as follows. A perturbation along $\tilde{\tilde{\boldsymbol{v}}}_1$ grows by a factor of $\sigma_1^r = 1.47$ which is smaller than the factor by which an error aligned along $\boldsymbol{v}_1$ grows ($\sigma_1 = 2.59$). But, the error along $\tilde{\tilde{\boldsymbol{v}}}_1$ which ends up aligning itself along the $\tilde{\boldsymbol{u}}_1$ at final time happens to be larger than that along $\boldsymbol{u}_1$. This is because though $\boldsymbol{v}_1$ has a much larger growth factor than the $\tilde{\tilde{\boldsymbol{v}}}_1$ direction, the error in that direction is smaller at initial time. The size of uncertainty along a particular direction at final time is the product of the size of uncertainty along the corresponding initial time direction and the corresponding singular value. For example, the size of the solid blue arrow in figure 2-3(a), that is $\|\boldsymbol{v}_1\| = 0.07$ units. Therefore the size of solid blue arrow in figure 2-3(c) is given by $\|\boldsymbol{u}_1\| = \|\boldsymbol{v}_1\| \times \sigma_1 = 0.07 \times 2.59 = 0.18$. On the other hand, the size of the solid green arrow in figure 2-3(a), that is $\|\tilde{\boldsymbol{v}}_1\| = 0.18$ units. Therefore the size of solid green arrow in figure 2-3(c) is given by $\|\tilde{\boldsymbol{u}}_1\| = \|\tilde{\boldsymbol{v}}_1\| \times \sigma_1^r = 0.18 \times 1.47 = 0.26$.

So the direction at initial time (in figure 2-3(a)) that ends up as the direction of maximum variance at final time (in figure 2-3(c)), is dictated both by the model dynamics and the structure of the initial uncertainty. In fact, there are many directions in figure 2-3(a) which have higher growth factors than $\tilde{\tilde{\boldsymbol{v}}}_1$ and there are also directions which have larger initial uncertainty (eg. the major axis of the green ellipse in figure 2-3(a)), but $\tilde{\tilde{\boldsymbol{v}}}_1$ is the direction that has the *right* combination of these two properties and thus it aligns itself in the direction of maximum variance at the final time. Thus the relevant singular directions and values when the *initial uncertainty distribution is non-isotropic* is given by $SVD(\mathbf{M}\sqrt{\mathbf{P}^{a-1}}^{-1})$ rather than by $SVD(\mathbf{M})$. The $\tilde{\tilde{\mathbf{V}}}$, $\tilde{\mathbf{U}}$ and $\tilde{\boldsymbol{\Sigma}}^r$ take into account both the uncertainty structure and the dynamics of error growth. Note that the singular vector and values given by $SVD(\mathbf{M})$ are correct, in the sense that perturbations in the non-isotropic distribution aligned along $\boldsymbol{v}_1$ grow by a factor of σ_1; it is just that they are not relevant from the stand point of finite time forecasting for this case.

The fact that the singular vectors and values relevant for the non-isotropic case are different from those for isotropic case has important implications for forecasting. Assuming an isotropic

distribution, when it is actually not, misinforms one about the direction at initial time that will end up having maximum variance at final time and the singular values could actually lead to suboptimal estimates of forecast error. As explained in the preceding paragraph, the direction having the largest growth factor may not grow into the direction having the largest error at forecast time, because there could be a direction which has a smaller growth factor but larger error aligned along it at initial time that grows in the largest error at forecast time. $SVD(\mathbf{M})$ gives the largest growing initial time direction while $SVD(\mathbf{M}\sqrt{\mathbf{P}^{a-1}}^{-1})$ (Ehrendorfer and Tribbia, 1995) gives the initial time direction which results in the largest error at forecast time. $SVD(\mathbf{M})$ considers the dynamics but neglects the uncertainty structure i.e. assumes it to be isotropic. $SVD(\mathbf{M}\sqrt{\mathbf{P}^{a-1}}^{-1})$ gives the relevant information from the point of view of predictability — it tells what current errors are going to influence the forecast the most. $SVD(\mathbf{M}\sqrt{\mathbf{P}^{a-1}}^{-1})$ takes into account both the dynamics and non-isotropic structure of the initial uncertainty distribution.

The operational reality of non-isotropic initial uncertainty has been accounted for in figure 2-2 (page 31)(c). The EnKF procedure outlined in section 2.4.2 is implemented at each step with $N=100$. To begin with, an initial isotropic ensemble is constructed around a point (i.e. truth) with a prescribed standard deviation (1% of attractor size) in the x and y components, by drawing from a standard normal distribution. Then the step given by equation (2.6a) is implemented by propagating each of the ensemble members forward using equations (2.1a) and (2.1b) (page 23). The truth is propagated forward as well. Data are generated artificially by first displacing the truth in a random direction by 1%, and then an ensemble of observations is generated around this displaced value of truth, with a prescribed standard deviation (1% of attractor size), drawing from a standard normal distribution. The matrix $\mathbf{R}$ has the variances of x and y observational errors along its diagonal. The gain term $\mathbf{K}$ is evaluated by using $\mathbf{R}$ and the covariance of the ensemble, $\mathbf{P}^f$ ($\mathbf{H}$ is set equal to the identity matrix). Finally data are assimilated using equation (2.6d), to produce the *best* ensemble i.e. assimilated ensemble. The data are assimilated at each step ($\tau = 1$) . At each step a forecast ($\tau = 1$) is launched from the mean of the assimilated ensemble and the resulting singular vectors and values are calculated according to the method outlined above (the ensemble $\mathbf{P}^a$ is used as uncertainty norm). The ensemble mean analysis is then colored by the leading singular value, which is shown in figure 2-2 (page 31)(c).

The picture in figure 2-2(c) is profoundly different from the isotropic case in figure 2-2(b). There are many more regions where the growth factor associated with the initial direction that grows into major axis of final time ellipse, is less than 1. (This information can be used, for example, for deciding the direction in which the errors need to be reduced in order to get better forecasts). The average of the leading singular value over the attractor in figure 2-2(a) is $\overline{\sigma}_1 = 2.47$ while that in figure 2-2(b) is $\overline{\sigma}^r = 1.91$. The differences are more striking for higher

values of τ (results not shown). This is the relevant picture of the variation of 1 step forecast error over state space; not the isotropic case of figure 2-2(b).

By considering the non-isotropic uncertainty distribution, one has taken an important step toward simulating operational reality, but there are other issues that remain to be addressed. For instance it would be overly cavalier to assume that one has access to the correct covariance matrix in operational forecasting. Most probably there would be errors in the covariance matrix such that the actually accessible matrix is rotated from its original form. Of course, apart from a mere rotation of the accessible matrix, there could be other errors; in general the accessible matrix can be totally wrong. But here only one of the possible issues, that of rotation of the accessible matrix, has been considered. The direction (in magenta, with arrowhead) in figure 2-3 (page 35)(a) denoted by $\tilde{\tilde{\boldsymbol{v}}}_{1rot}$ corresponds to the leading initial time singular vector as calculated by rotating $\mathbf{P}^a$ through an angle of $\theta = 20°$ (the 2ξ bound of this rotated distribution is shown as the magenta ellipse in figure 2-3(a)). This rotated distribution is propagated forward by using the tangent linear model. The 2ξ bound of this propagated distribution is shown in figure 2-3(d) as the magenta ellipse. The corresponding final time singular vector is shown as $\tilde{\boldsymbol{u}}_{1rot}$. It is observed that though the singular directions change dramatically, the singular values remain almost the same ($\sigma_1^r \simeq \sigma_{1rot}^r$). Figure 2-2 (page 31)(d) and figure 2-4 (page 39)(a) show the distribution of σ_1^r when $\mathbf{P}^a$ is rotated through an angle drawn from $N(0, 20°)$ and $N(0, 30°)$ respectively. This means at each point from which the forecast is launched, $\mathbf{P}^a$ is rotated through a different angle chosen at random from a gaussian distribution with standard deviation 20° and 30°. Higher the angle of rotation higher is the change in the distribution of singular values.

A variant of this operational issue is when the error variances are available but the error covariances are not. In this case, the distribution of singular values will look like those in figure 2-4(b). Note that it looks intermediate to the isotropic and non-isotropic case. For this model, using variances only is better than ignoring the uncertainty structure altogether. The above discussion took into consideration some important aspects of finite time operational forecasting, which mainly concerned the initial condition uncertainty. Model inadequacy remains one of the most important stumbling blocks in the path towards accurate forecast. In fact, it is speculated that model inadequacy could be vitiating the forecast more seriously than the initial condition uncertainty (Orrell et al., 2001). The next section deals with the influence of model inadequacy on forecast errors.

2.4.4 Influence of model inadequacy

An idea about the influence of model inadequacy on predictability can be obtained by looking at its effect on the singular value distribution. The model error is broadly classified into

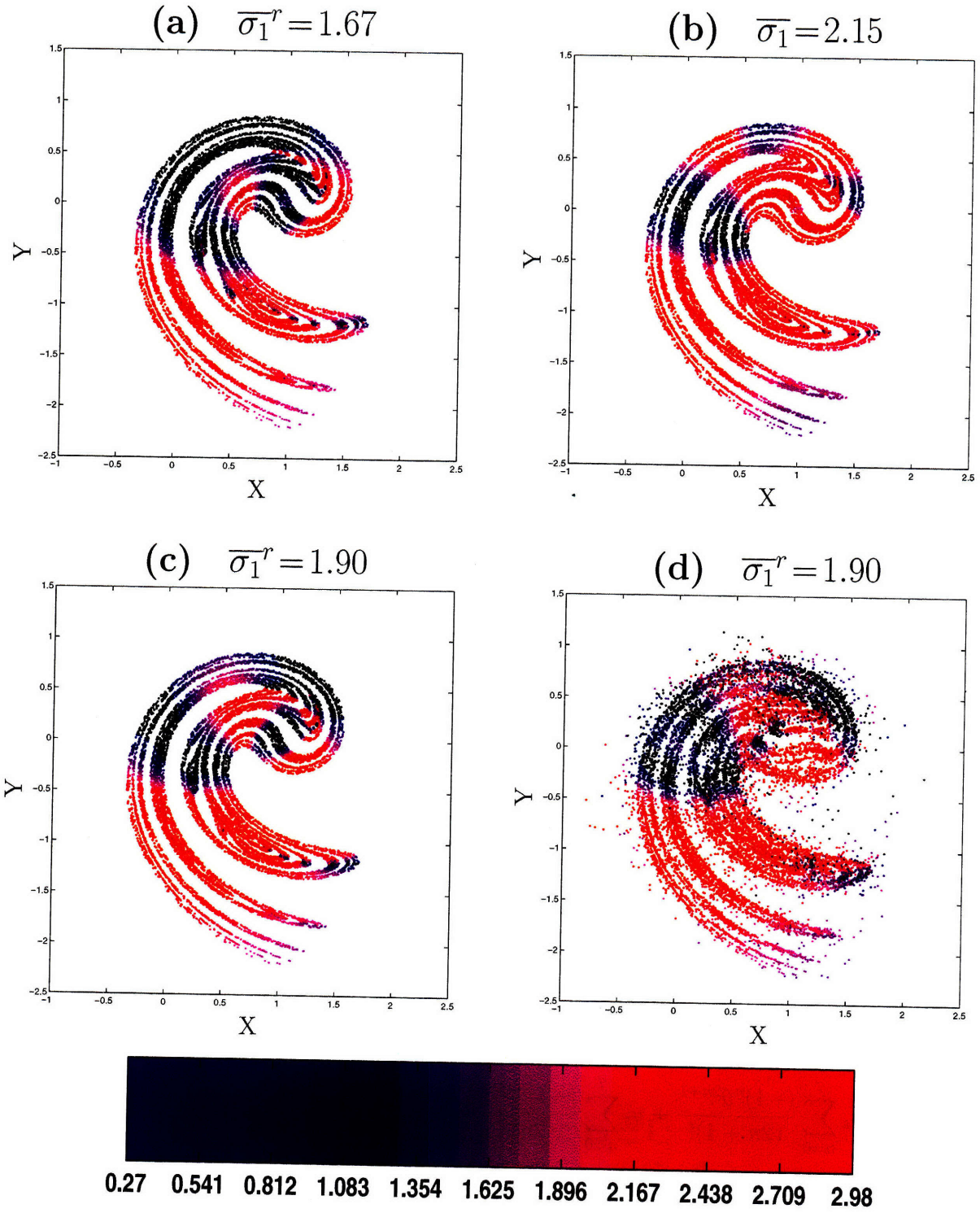


Figure 2-4: The distribution of σ_1^r under different operational scenarios. Points in black have $\sigma_1 < 1$. In **(a)** and **(b)** the model is perfect.
(a) $\mathbf{P}^a$ is rotated by an angle θ drawn from $N(0, 30°)$. **(b)** Covariances are assumed to be zero i.e. the off-diagonal entries of $\mathbf{P}^a$ are zero; $\mathbf{P}^a$ contains only the variances. **(c)** Parametrically imperfect model ($\Delta\mu = -0.05$). The initial distribution is $\mathbf{P}^a$. **(d)** Structurally imperfect model. The initial distribution is $\mathbf{P}^a$.

parametric and structural categories. *Parametric model error* pertains to having the *wrong* values of constants in a given model. In the Ikeda system, the constants are given by a, b and μ. The sensitivity of singular values to change in these parameters is studied. The Ikeda system given by equations (2.1a) and (2.1b) (page 23) is treated as a perfect model (or the system). Figure 2-4 (page 39)(c) shows the distribution of singular values calculated from the Ikeda system in which the value of parameter μ is lower by -0.05% as compared to that in the perfect model, that is, the parametric error is $\Delta\mu = -0.05$. Data assimilation is performed as in section 2.4.3, so that figure 2-4(c) shows the impact of having both an initial non-isotropic distribution and an imperfect model.

Comparing figure 2-2 (page 31)(c) with figure 2-4(c), one notes that the singular values do not change much with small changes in μ. Similarly, it is noted that for small changes in values of a and b the singular values do not change appreciably (result not shown). Hence, one can conclude that the parametric error does not affect the singular values to a large extent, for this range of parametric error for this model. The impact is of course dependent on τ. Larger impact is seen for longer τ, which means that models with parametric error are *good* models for this system for short τ. It has to be noted that, though it appears that attractor in figure 2-4(c) is same as that in figure 2-2(c), in principle the attractors are different. The system attractor is different from the model attractor.

Structural model error pertains to a situation in which the functional dependence of one or more terms in the model is different from the *correct* dependence. In the Ikeda model, such an error is introduced by replacing the cosine and sine terms in the model (equations (2.1a) and (2.1b) (page 23)) by the first 7 terms in their series expansion. Thus the model with structural error is given by

$$x_{s+1} = 1 + \mu \left(x_s \sum_{n=0}^{7} \frac{(-1)^n \theta^{2n}}{2n!} - y_s \sum_{n=0}^{7} \frac{(-1)^n \theta^{2n+1}}{(2n+1)!} \right) \tag{2.8a}$$

$$y_{s+1} = \mu \left(x_s \sum_{n=0}^{7} \frac{(-1)^n \theta^{2n+1}}{(2n+1)!} + y_s \sum_{n=0}^{7} \frac{(-1)^n \theta^{2n}}{2n!} \right) \tag{2.8b}$$

where

$$\theta = a - \frac{b}{(x_s^2 + y_s^2 + 1)^2}$$

and $a = 0.4$, $b = 6.0$, $\mu = 0.9$.

Data are assimilated as in the case of parametric error discussed is section 2.4.4. The data for assimilation are drawn from the system (that is truth) and the ensemble members are propagated using equations (2.8a) and (2.8b).

Results are shown in figure 2-4(d). Comparing figure 2-2(c) to figure 2-4(d), one sees that there is a large difference between the distributions of singular values. Particularly in the region centered around the point $(0,0)$, the regions of shrinking errors have decreased considerably. Also note that the structural error changes the attractor structure and its effect is seen clearly in the region around the point $(1,0)$. For longer optimization times, the impact of structural model error is more spectacular. Figure 2-4(d) shows the impact of having a structurally imperfect model and non-isotropic initial condition. This is the scenario which is closest to operational reality. Note the big change in the distribution of the leading singular value in going from theoretical simplification (Figure 2-2(a)) to operational reality (Figure 2-4(d)).

In section 2.4, the impact of the initial error in state ($\mathbf{x}$) was examined under various scenarios of initial error structure and model imperfections. In section 2.4.4, the *implicit* effect of model error on singular value distribution was considered. The effect of model error is implicit because the singular value only with respect to the state was considered. In the next section, we will examine the explicit influence of model error on forecast error vis-a-vis the influence of error in the initial state.

2.5 Initial condition error versus model error

In the preceding sections, the effect of the shape of initial uncertainty and model error on the predictability of the Ikeda system was explored. This exploration did not *explicitly* include the effect of model error on predictability. That is, the forecast error was solely due to initial condition error. In this section, the forecast error due to the initial condition error and the model error are considered. Particularly, the effect of initial condition error and that of model error on forecast error is compared. The linearization tools presented in sections 2.3 and 2.4 were in the context of the 2-dimensional state. In this section, these linearization tools are extended to a 3-dimensional (augmented) system so that the impact of parametric model error on final time error can be calculated under linearized dynamics.

Let us assume that the parameter μ in the Ikeda system given by equations (2.1a) and (2.1b) (page 23) is imperfect. Let the 2-dimensional Ikeda system be augmented by μ, that is, let us consider μ as a variable rather than a parameter. Consider the 3-dimensional augmented

Ikeda system,

$$x_{s+1} = F_1(x_s, y_s) = 1 + \mu(x_s \cos\theta - y_s \sin\theta) \tag{2.9a}$$

$$y_{s+1} = F_2(x_s, y_s) = \mu(x_s \sin\theta + y_s \cos\theta) \tag{2.9b}$$

$$\mu_{s+1} = F_3 = \mu_s \tag{2.9c}$$

Note that the step forward equation for μ (equation (2.9c)) holds the value of μ constant in time.

Let the 3-dimensional vector in the augmented Ikeda space be denoted by $\mathbf{w}$. The linear propagator for the augmented system given by equations (2.9a), (2.9b) and (2.9c) is

$$\mathbf{M^w} = \begin{bmatrix} F'_{1x} & F'_{1y} & F'_{1\mu} \\ F'_{2x} & F'_{2y} & F'_{2\mu} \\ F'_{3x} & F'_{3y} & F'_{3\mu} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.10}$$

Let the error in $\mathbf{w}$ at time s be $\boldsymbol{\epsilon}^{\mathbf{w}}_s = [\epsilon_{x_s}\ \epsilon_{y_s}\ \epsilon_{\mu_s}]$. The linearized error in $\mathbf{w}$ at time $s+1$ is

$$\boldsymbol{\epsilon}^{\mathbf{w}}_{s+1} = \begin{bmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{x_s} \\ \epsilon_{y_s} \\ \epsilon_{\mu_s} \end{bmatrix} = \begin{bmatrix} A_1\epsilon_{x_s} + A_2\epsilon_{y_s} + A_3\epsilon_{\mu_s} \\ B_1\epsilon_{x_s} + B_2\epsilon_{y_s} + B_3\epsilon_{\mu_s} \\ \epsilon_{\mu_s} \end{bmatrix}.$$

$$\boldsymbol{\epsilon}^{\mathbf{w}}_{s+1} = \begin{bmatrix} A_1\epsilon_{x_s} + A_2\epsilon_{y_s} \\ B_1\epsilon_{x_s} + B_2\epsilon_{y_s} \\ 0 \end{bmatrix} + \begin{bmatrix} A_3\epsilon_{\mu_s} \\ B_3\epsilon_{\mu_s} \\ \epsilon_{\mu_s} \end{bmatrix}. \tag{2.11}$$

$$\boldsymbol{\epsilon}^{\mathbf{w}}_{s+1} = \begin{bmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{x_s} \\ \epsilon_{y_s} \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \epsilon_{\mu_s} \end{bmatrix} \tag{2.12}$$

The above equation (2.11) can be written as

$$\boldsymbol{\epsilon}^{\mathbf{w}}_{s+1} = \mathbf{M^w}\boldsymbol{\epsilon}^{\mathbf{wx}}_s + \mathbf{M^w}\boldsymbol{\epsilon}^{\mathbf{w}\mu}_s = \boldsymbol{\epsilon}^{\mathbf{wx}}_{s+1} + \boldsymbol{\epsilon}^{\mathbf{w}\mu}_{s+1} \tag{2.13}$$

In the above equation (2.13), $\boldsymbol{\epsilon}^{\mathbf{wx}}_s$ should be read as the error in $\mathbf{w}$ at time s projected onto $\mathbf{x}$ at time s. Similarly $\boldsymbol{\epsilon}^{\mathbf{w}\mu}_s$ is the error in $\mathbf{w}$ at time s projected onto μ at time s.

Taking the first 2 components of $\boldsymbol{\epsilon}^{\mathbf{w}}_{s+1}$ in equation (2.13) (that is, taking the x, y components

of $\boldsymbol{\epsilon}_{s+1}^{\mathbf{w}}$ in equation (2.11)),

$$\boldsymbol{\epsilon}_{s+1}^{\mathbf{w},\mathbf{x}} = \boldsymbol{\epsilon}_{s+1}^{\mathbf{wx},\mathbf{x}} + \boldsymbol{\epsilon}_{s+1}^{\mathbf{w}\mu,\mathbf{x}} \tag{2.14}$$

For brevity, equation (2.14) is written as

$$\boldsymbol{\epsilon}_{s+1}^{\mathbf{wx}} = \boldsymbol{\epsilon}_{s+1}^{\mathbf{xx}} + \boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}} \tag{2.15}$$

In words, equation (2.15) says that the total error in the state at final time $s+1$ due to the error in $\mathbf{w}$ at time s, is the sum of the state error at final time, due to error in $\mathbf{x}$ at time s and due to error in μ at time s.

The term $\boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}}$ needs further explanation. $\boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}}$ gives the error in the final state due to the error in μ at time s namely ϵ_{μ_s}. That is, the state $\mathbf{x}$ at time s is perfect. However, as soon as the state is propagated forward with the error in μ, ϵ_{μ_s} induces an error in state. This induced error evolves with time. Therefore $\boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}}$ contains the effect of ϵ_{μ_s} and *subsequent perturbations it induces into the state.*

Each vector in equation (2.15) is 2-dimensional while each vector in equation (2.13) is 3-dimensional. The vectors in equation (2.15) are given by

$$\boldsymbol{\epsilon}_{s+1}^{\mathbf{wx}} = \begin{bmatrix} A_1\epsilon_{x_s} + A_2\epsilon_{y_s} + A_3\epsilon_{\mu_s} \\ B_1\epsilon_{x_s} + B_2\epsilon_{y_s} + B_3\epsilon_{\mu_s} \end{bmatrix} \tag{2.16}$$

$$\boldsymbol{\epsilon}_{s+1}^{\mathbf{xx}} = \begin{bmatrix} A_1\epsilon_{x_s} + A_2\epsilon_{y_s} \\ B_1\epsilon_{x_s} + B_2\epsilon_{y_s} \end{bmatrix} \tag{2.17}$$

$$\boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}} = \begin{bmatrix} A_3\epsilon_{\mu_s} \\ B_3\epsilon_{\mu_s} \end{bmatrix} \tag{2.18}$$

From equation (2.17), the square of the magnitude of $\boldsymbol{\epsilon}_{s+1}^{\mathbf{xx}}$ is

$$\|\boldsymbol{\epsilon}_{s+1}^{\mathbf{xx}}\|^2 = A_1^2\epsilon_{x_s}^2 + A_2^2\epsilon_{y_s}^2 + 2A_1A_2\epsilon_{x_s}\epsilon_{y_s} + B_1^2\epsilon_{x_s}^2 + B_2^2\epsilon_{y_s}^2 + 2B_1B_2\epsilon_{x_s}\epsilon_{y_s} \tag{2.19}$$

From equation (2.18), the square of the magnitude of $\boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}}$ is

$$\|\boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}}\|^2 = A_3^2\epsilon_{\mu_s}^2 + B_3^2\epsilon_{\mu_s}^2 \tag{2.20}$$

From equations (2.17) and (2.18) define the vector

$$\boldsymbol{\epsilon}_{s+1}^{\mathbf{xx}} - \boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}} = \begin{bmatrix} A_1\epsilon_{x_s} + A_2\epsilon_{y_s} - A_3\epsilon_{\mu_s} \\ B_1\epsilon_{x_s} + B_2\epsilon_{y_s} - B_3\epsilon_{\mu_s} \end{bmatrix} \tag{2.21}$$

Define

$$\eta^2 = \|\boldsymbol{\epsilon}^{\mathbf{xx}}_{s+1}\|^2 + \|\boldsymbol{\epsilon}^{\boldsymbol{\mu}\mathbf{x}}_{s+1}\|^2 - \|\boldsymbol{\epsilon}^{\mathbf{xx}}_{s+1} - \boldsymbol{\epsilon}^{\boldsymbol{\mu}\mathbf{x}}_{s+1}\|^2 \tag{2.22}$$

Using equations (2.19), (2.20) and (2.21) it can be shown that

$$\eta^2 \doteq 2[A_1A_3 + B_1B_3]\epsilon_{x_s}\epsilon_{\mu_s} + 2[A_2A_3 + B_2B_3]\epsilon_{y_s}\epsilon_{\mu_s} \tag{2.23}$$

From equations (2.16), (2.19), (2.20) and (2.23),

$$\|\boldsymbol{\epsilon}^{\mathbf{xx}}_{s+1}\|^2 + \|\boldsymbol{\epsilon}^{\boldsymbol{\mu}\mathbf{x}}_{s+1}\|^2 + \eta^2 = \|\boldsymbol{\epsilon}^{\mathbf{wx}}_{s+1}\|^2 \tag{2.24}$$

$\|\boldsymbol{\epsilon}^{\mathbf{wx}}_{s+1}\|$ is the magnitude of the total state error at final time. Dividing equation (2.24) by $\|\boldsymbol{\epsilon}^{\mathbf{wx}}_{s+1}\|^2$ and using

$$\zeta^2_{\mathbf{x}} = \frac{\|\boldsymbol{\epsilon}^{\mathbf{xx}}_{s+1}\|^2}{\|\boldsymbol{\epsilon}^{\mathbf{wx}}_{s+1}\|^2} \tag{2.25a}$$

$$\zeta^2_{\boldsymbol{\mu}} = \frac{\|\boldsymbol{\epsilon}^{\boldsymbol{\mu}\mathbf{x}}_{s+1}\|^2}{\|\boldsymbol{\epsilon}^{\mathbf{wx}}_{s+1}\|^2} \tag{2.25b}$$

$$\zeta^2_{\mathbf{x}\boldsymbol{\mu}} = \frac{\eta^2}{\|\boldsymbol{\epsilon}^{\mathbf{wx}}_{s+1}\|^2} \tag{2.25c}$$

one obtains

$$\zeta^2_{\mathbf{x}} + \zeta^2_{\boldsymbol{\mu}} + \zeta^2_{\mathbf{x}\boldsymbol{\mu}} = 1 \tag{2.26}$$

In equation (2.26), $\zeta^2_{\mathbf{x}}$ is the contribution to the total state error (at final time) from the initial error in $\mathbf{x}$. $\zeta^2_{\boldsymbol{\mu}}$ is the contribution from the model error in μ. From equation (2.23), it is clear that $\zeta^2_{\mathbf{x}\boldsymbol{\mu}}$ is the contribution to the model error from the interaction between the model error and initial condition error. I will refer to $\zeta^2_{\mathbf{x}}$ as the (contribution of) state error, $\zeta^2_{\boldsymbol{\mu}}$ as the (contribution of) model error and $\zeta^2_{\mathbf{x}\boldsymbol{\mu}}$ as the (contribution of) interaction error. It is important to realize that equation (2.24) (and consequently equation (2.26)) is a statement about the *square* of the magnitude of the total state error at final time. This is a basic result that has not appeared in the predictability literature.

Using the axioms of norms given by equations (2.5a), (2.5b) and (2.5c) listed in section 2.3.3 (page 27), it is proved that

$$\eta^2 \geq -2\|\boldsymbol{\epsilon}^{\mathbf{xx}}_{s+1}\|\|\boldsymbol{\epsilon}^{\boldsymbol{\mu}\mathbf{x}}_{s+1}\| \tag{2.27}$$

The physical interpretation of equation (2.27) is interesting. $\eta^2 < 0$ represents the case when the error due to $\epsilon_{\mathbf{x}}$ and that due to ϵ_{μ_s} cancel out. If $\eta^2 < 0$, then from equation (2.24) the interaction error decreases the total error.

A comparison between state-dependent contributions of the state error, the model error and interaction error is presented next. The data assimilation is done as in section 2.4.4. The model is parametrically imperfect, with $\Delta\mu = -0.07$ expressed in % of $\mu = 0.9$. The analysis covariance gives the non-isotropic uncertainty structure for the initial state. This analysis covariance, $\mathbf{P}^a$ which is 2×2, is augmented with the parametric variance to obtain a 3×3 analysis covariance, namely $\mathbf{P}^a_{\mathbf{w}}$. The covariance between μ and $\mathbf{x}$ is assumed to be 0. The $SVD(\mathbf{M}^{\mathbf{w}} \sqrt{{\mathbf{P}^a_{\mathbf{w}}}^{-1}}^{-1})$, where $\mathbf{M}^{\mathbf{w}}$ is given by equation (2.10), is carried out for an optimization time of $\tau = 1$ and the relevant (leading) initial time singular vector is determined. $\zeta^2_{\mathbf{x}}$, ζ^2_{μ} and $\zeta^2_{\mathbf{x}\mu}$ corresponding to this relevant singular vector are computed. The distribution of $\zeta^2_{\mathbf{x}}$, ζ^2_{μ} and $\zeta^2_{\mathbf{x}\mu}$ over the model attractor is presented in figure 2-5 (page 46). Figure 2-5(a) shows considerable variation of the contribution of the state error. It is interesting that the regions resulting in high contribution of state error (red regions) tend to cluster together. The average value of $\zeta^2_{\mathbf{x}}$ is 0.53. Figure 2-5(b) shows the variation of contribution of the model error. The average value of ζ^2_{μ} is 0.35. This means that, on average, the contribution of model error in rendering the forecast wrong is 35%, which is quite high. Note that because these contributions are state-dependent, the average values are misleading. There are regions in the attractor (like near $(1.5, -1.25)$ in figure 2-5(b)) where the contribution of model error is almost 1. Therefore, in some regions of the attractor, the model error could be the dominant factor in rendering the forecast wrong. Figure 2-5(c) shows that the contribution of interaction error is relatively low. Finally, figure 2-5(d) shows the total contribution of the model error and interaction error taken together. In general, given a state, these contributions depend on the level (or magnitude) of model error and optimization time.

The variation of the average of $\zeta^2_{\mathbf{x}}$, ζ^2_{μ} and $\zeta^2_{\mathbf{x}\mu}$ over the attractor as a function of τ for two values of parametric model error is shown in figure 2-6. The dashed curves are for the lower model error of $\Delta\mu = -0.05$. The solid curves are for higher model error of $\Delta\mu = -0.07$. Consider the lower model error curves. As τ increases, $\zeta^2_{\mathbf{x}}$ decreases. $\zeta^2_{\mathbf{x}} > \zeta^2_{\mu}$ for all values of τ. Consider the higher model error curves. $\zeta^2_{\mathbf{x}} > \zeta^2_{\mu}$ for $\tau \leq 2$ and $\zeta^2_{\mathbf{x}} < \zeta^2_{\mu}$ for $\tau \geq 3$. Therefore the importance of model error vis-a-vis the state error in rendering the forecast wrong depends upon the magnitude of the model error and the optimization time. When the model error is decreased from $\Delta\mu = -0.07$ to $\Delta\mu = -0.05$, ζ^2_{μ} decreases by about 0.15. This decrease is approximately constant with increase in τ. The interaction error (shown in cyan color) does not change much with decrease in $\Delta\mu$. $\zeta^2_{\mathbf{x}\mu}$ tends to saturate as τ increases. Thus the increase in $\Delta\mu$ mostly leads to an increase in the model error and to a lesser extent an increase in the interaction error.

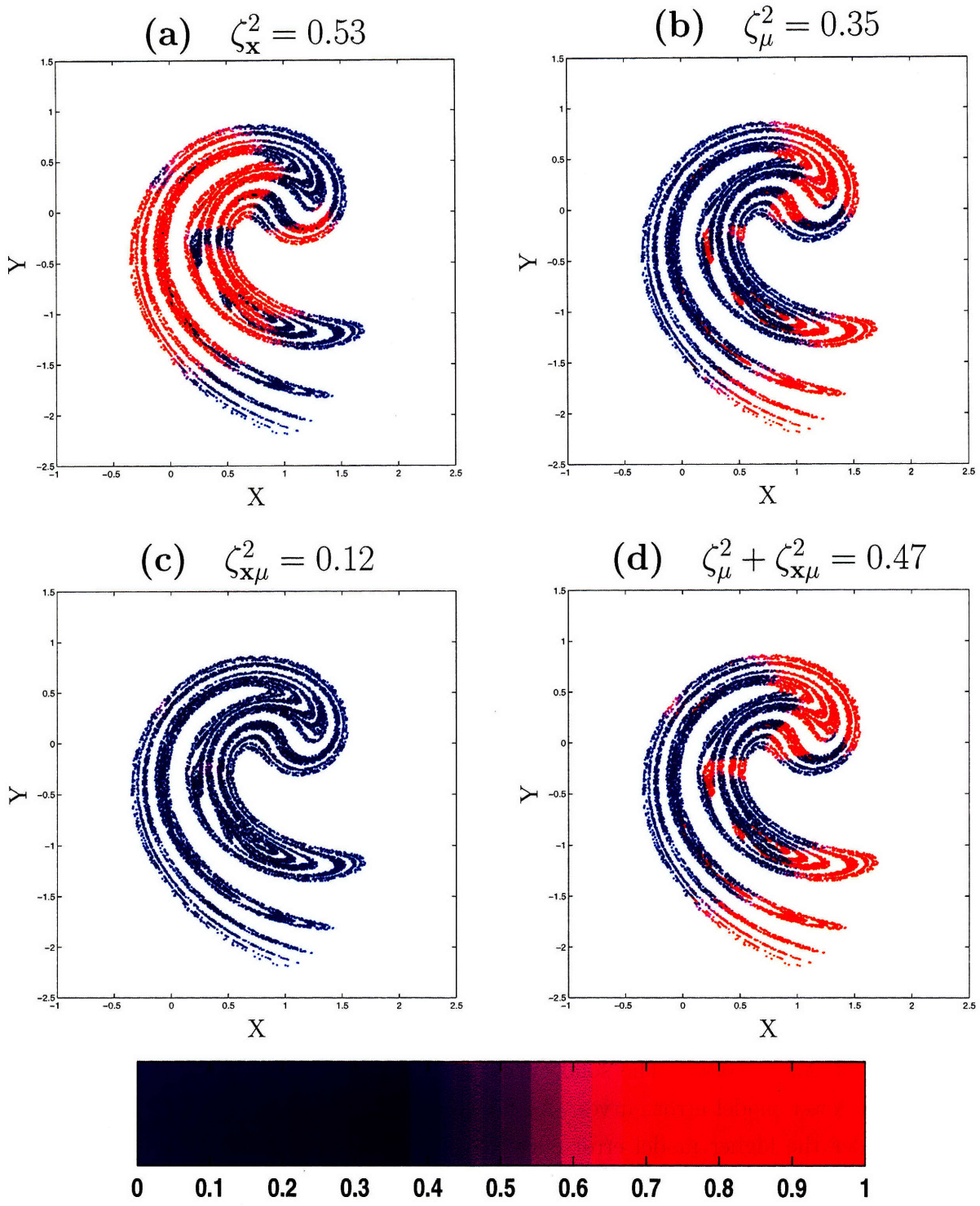


Figure 2-5: The distribution of the contribution of state error, model error and the interaction error in the total final time state error for $\Delta\mu = -0.07$ and $\tau = 1$. **(a)** The contribution of the state error. **(b)** The contribution of the model error. **(c)** The contribution of the interaction. **(d)** The sum of the contribution of the model error and interaction error.

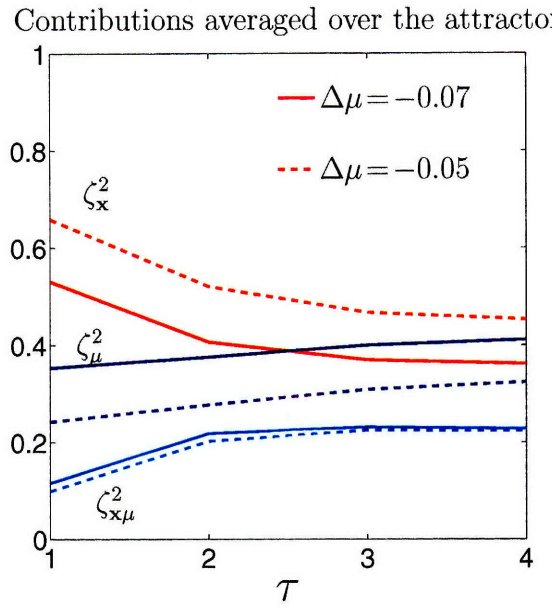


Figure 2-6: Change of average (over the attractor) of $\zeta^2_{\mathbf{x}}$ (red color), ζ^2_{μ} (blue color) and $\zeta^2_{\mathbf{x}\mu}$ (cyan color) with optimization time for two different levels of parametric model error. The solid curves are for $\Delta\mu = -0.07$ and dashed curves for $\Delta\mu = -0.05$. The contribution from the model error could be more than that from the state error depending on the level of model error and τ. The solid blue curve lies above the solid red for $\Delta\mu = -0.07$ and $\tau \geq 3$. Note that the contribution of interaction error does not change much with the model error levels compared to the change in contributions of state error and model error. The values for $\Delta\mu = -0.07$ and $\tau = 1$ correspond to figure 2-5.

Thus the model error could be as important as the state error in determining the total forecast error. In some regions of the attractor, the model error could be far more important than the state error in determining the total forecast error.

2.6 Conclusions

A cascade of predictability scenarios was presented in the context of a 2-dimensional chaotic map, ranging from theoretical idealization to something akin to operational reality, thus eliciting some of the issues impending accurate forecasts.

Singular value does a far better job in representing finite time forecast errors than does the Lyapunov number. The leading singular value calculated assuming isotropic initial uncertainty distribution changes over the phase space. Thus forecast errors under *linearized dynamics* are *state dependent*, over *finite time scales*. In some regions of the state space, all initial errors

decay, showing local enhanced forecast quality. The regions of decaying initial errors persist (though they change location) for longer optimization times. The singular error growth in most parts of the attractor is rapid, in that errors can grow almost by a factor of 3 over one step. The singular values calculated assuming an isotropic initial uncertainty are called the isotropic or plain singular values.

The initial uncertainty is always non-isotropic. The shape of this non-isotropic uncertainty is dictated by the local attractor structure. An estimate of the shape of this non-isotropic uncertainty is obtained by employing the EnKF. The leading singular direction gives the direction of maximum growth of initial errors. It is shown that these relevant directions are different from those calculated assuming an isotropic distribution. These directions depend both on the growth factor given by the local dynamics and the error magnitude given by the uncertainty structure. The method to obtain these relevant directions and growth factors is the $SVD(\mathbf{M}\sqrt{\mathbf{P}^{a-1}}^{-1})$ rather than $SVD(\mathbf{M})$. The leading singular value distribution is quite different in the non-isotropic case. Given the fact that calculation of the covariance matrix might be very expensive, using only the variances is a reasonable approximation to using the full covariance matrix. It is found that the singular values are robust to small rotations of the covariance matrix, while the singular vectors are fragile to such rotations for this system. The isotropic singular values and relevant singular values are different quantifications of predictability.

Model error changes the distribution of singular values as compared to the perfect model case. Qualitatively, it appears that for the level of parametric and structural errors specified in section 2.4.4, the structural error induces more change in the distribution of singular values as compared to the perfect model than does the parametric error for this model. It is, of course, impossible to generalize this result to other models. For longer forecast times, the perfect model assumption gives a highly misleading picture of error growth.

Though a method to find the relevant singular values and vectors for initial non-isotropic uncertainty is available, it seems difficult to get around the model inadequacy problem. When only an imperfect model is at hand, one is doomed to get the wrong singular values. The comparison of scenarios presented here clearly shows that idealizations (isotropic initial uncertainty and perfect model) can seriously hamper accurate forecasts and motivates research into ways to reduce model inadequacy.

A comparison between the contribution of parametric model error and the initial state error to the total forecast error shows that the model error could be as important as the initial state error, to say the least. For a higher level of model error, the contribution of model error to forecast error could be more than that of initial state error.

It is very important to mitigate the effect of model error. In chapter 3, the technique of accounting for model error in the EnKF is investigated. Chapter 4, explores the technique of reduction of model error in ensemble-based filters.

Chapter 3

Accounting for model error in the EnKF

The critical importance of determining the initial state accurately has prompted the application of techniques to pin down the initial state with as much certainty as possible. Prominent among these techniques are 4dVar (Le Dimet and Talagrand, 1986), the EKF (Evensen, 1992) and the EnKF (Evensen, 1994). In this chapter the modification of EnKF to account for model error is discussed. It is shown that lower analysis and forecast errors are obtained if the model uncertainty is accounted for in the EnKF. The various extant and novel methods to obtain model uncertainty are discussed and intercompared.

3.1 Introduction

Recently, there has been a lot of active research in the broad field of ensemble-based filters as tools for state estimation. Among all the different ensemble-based filters, the EnKF has occupied a central position. Typically, the EnKF has been used with atmospheric and oceanic models assuming that these models are perfect. This is not to say that the researcher believed that the models are perfect. This is rather because accounting for model error in the EnKF framework has not been as well researched as accounting for initial condition error.

As discussed in sections 2.5 (page 41) and 2.6 (page 47), model errors play an important role in rendering the model forecast wrong. This chapter addresses the question of accounting for model error in EnKF. Section 3.2 discusses the correct approach to accounting for model error in EnKF. This section also discusses a simplistic, but approximate approach called the

Additive Error Approximation (AEA). The AEA requires an estimate of the model error. Various techniques of obtaining estimates of model error are discussed in sections 3.3 (page 57), 3.4, 3.5 and 3.6. The multimodel method is introduced and discussed in section 3.5. A novel method termed the parametric vector method is introduced and discussed in section 3.6. The performance of the various methods of estimating the model error is intercompared within the AEA framework in section 3.7. This intercomparison is done in the low-dimensional Ikeda and Lorenz 1963 models. Section 3.8 (page 86) presents the conclusions of this chapter.

3.2 Approaches to accounting for model error in the EnKF

The EnKF equations were presented and discussed in section 2.4.2 (page 32) given a perfect model. In this section the issue of accounting for model error in the EnKF is discussed.

Accounting for model error in EnKF is easier to discuss if the Extended Kalman Filter (EKF) is understood (Hansen and Smith, 2001). The following discussion is adapted from Hansen and Smith (2001). The EKF equations are discussed below. Consider the system given by

$$\mathbf{x}^t(t) = \mathbf{G}[\mathbf{x}^t(t-1)] \tag{3.1}$$

where $\mathbf{x}^t$ is the *truth*. The rhs of equation (3.1) should be read as the (nonlinear) system operator $\mathbf{G}$ acting on the truth $\mathbf{x}^t$ at time $t-1$. Consider the nonlinear forecast model given by

$$\mathbf{x}^f(t) = \mathbf{F}[\mathbf{x}^a(t-1)] \tag{3.2}$$

where $\mathbf{x}^a(t-1)$ is an initial condition. Let the imperfect observation of the truth at time t be available. This observation is given by

$$\mathbf{y}^o(t) = \mathbf{H}\mathbf{x}^t(t) + \epsilon_o \tag{3.3}$$

where $E(\epsilon_o) = 0$ and $E(\epsilon_o \epsilon_o^T) = \mathbf{R}$, where E is the expectation operator. $\mathbf{H}$ maps the model space to observation space.

If the model is perfect, that is, $\mathbf{F} = \mathbf{G}$, then the Extended Kalman Filter (EKF) equations are

as follows.

$$\mathbf{x}^f(t) = \mathbf{F}[\mathbf{x}^a(t-1)] \tag{3.4a}$$

$$\mathbf{P}^f(t) = \mathbf{F}'\mathbf{P}^a(t-1)\mathbf{F}'^T \tag{3.4b}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T[\mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t)]^{-1} \tag{3.4c}$$

$$\mathbf{x}^a(t) = \mathbf{x}^f(t) + \mathbf{K}(t)[\mathbf{y}^o(t) - \mathbf{H}(t)\mathbf{x}^f(t)] \tag{3.4d}$$

$$\mathbf{P}^a(t) = \mathbf{P}^f(t) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}^f(t) \tag{3.4e}$$

Equation (3.4a) propagates the initial condition $\mathbf{x}^a$ from time $t-1$ to time t. $\mathbf{x}^f(t)$ is the forecast or the *first guess*. $\mathbf{P}^a(t-1)$ is a measure of uncertainty in $\mathbf{x}^a(t-1)$. Equation (3.4b) propagates the uncertainty in $\mathbf{x}^a$ from time $t-1$ to time t. $\mathbf{F}'$ is a linearization of $\mathbf{F}$. $\mathbf{P}^f(t)$ is called the forecast error covariance. The gain matrix $\mathbf{K}$ is calculated in equation (3.4c), which gives the relative weight given to the forecast and the observation. Equation (3.4d) "blends" the forecast with the observation $\mathbf{y}^o(t)$ to give the best estimate of state at time t, namely the analysis or the update $\mathbf{x}^a(t)$. Equation (3.4e) gives the uncertainty in the analysis. $\mathbf{P}^a(t)$ is called the analysis error covariance. $\mathbf{x}^t$, $\mathbf{x}^f$, $\mathbf{y}^o$ and $\mathbf{x}^a$ are n-dimensional vectors, where n is the dimension of the model. $\mathbf{P}^f$and $\mathbf{P}^a$ are $n \times n$ matrices. $\mathbf{R}$ is $m \times m$ matrix where m is the number of dimensions observed ($m \leq n$). Note that as the dimensionality n of the model increases, the solution of equation (3.4b) becomes computationally expensive. The EnKF (Evensen, 1994) was formulated to avoid this expensive calculation, among the other reasons discussed in section 1.1 (page 9). The EnKF equations for the perfect model were presented and discussed in the section 2.4.2. They are reproduced below to compare and contrast with the EKF equations from (3.4a) through (3.4e). For $i=1 \ldots N$,

$$\mathbf{x}_i^f(t) = \mathbf{F}[\mathbf{x}_i^a(t-1)] \tag{3.5a}$$

$$\mathbf{P}^f(t) = \frac{1}{N-1}[\mathbf{E}^f(t) - \mathbf{M}^f(t)][\mathbf{E}^f(t) - \mathbf{M}^f(t)]^T \tag{3.5b}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T[\mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t)]^{-1} \tag{3.5c}$$

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f(t) + \mathbf{K}(t)[\mathbf{y}_i^o(t) - \mathbf{H}(t)\mathbf{x}_i^f(t)] \tag{3.5d}$$

$$\mathbf{P}^a(t) = \frac{1}{N-1}[\mathbf{E}^a(t) - \mathbf{M}^a(t)][\mathbf{E}^a(t) - \mathbf{M}^a(t)]^T \tag{3.5e}$$

The EnKF propagates N samples ($\mathbf{x}_i^a$) drawn from $\mathbf{P}^a(t-1)$ under the model dynamics given by $\mathbf{F}$. These samples are called the ensemble members. The forecast ensemble members are denoted by $\mathbf{x}_i^f$. The expected value of the forecast and its uncertainty ($\mathbf{P}^f(t)$) are given by the first two moments of $\mathbf{x}_i^f$. Therefore $\mathbf{P}^f(t)$ in the EnKF is obtained without implement-

ing equation (3.4b), thus saving computational expense. The analysis is performed on each forecast ensemble member separately using perturbed observations (equation (3.5d)).

The equations of EKF ((3.4a) through (3.4e)) assume that the model is perfect (that is, $\mathbf{F}=\mathbf{G}$). If the model is imperfect, then $\mathbf{F}\neq\mathbf{G}$. The difference between $\mathbf{F}$ and $\mathbf{G}$ is quantified as follows. Let us assume that we have access to the truth at time $t-1$. Consider $\mathbf{G}[\mathbf{x}^t(t-1)]$ and $\mathbf{F}[\mathbf{x}^t(t-1)]$. Since $\mathbf{F}$ and $\mathbf{G}$ act on the same initial condition $\mathbf{x}^t(t-1)$, the difference in the forecasts must be due to difference between $\mathbf{G}$ and $\mathbf{F}$, that is, due to the model error.

Let us assume that the difference between the system and model can be written as

$$\mathbf{G}[\mathbf{x}^t(t-1)] = \mathbf{F}[\mathbf{x}^t(t-1)] + \eta(t) \tag{3.6}$$

where η is a Gaussian noise process. Let us assume that η has been obtained with some methodology that calculates the discrepancy between the system and the model. Let us assume that the first two moments of η are given by $E(\eta)=0$ and $E(\eta\eta^T)=\mathbf{Q}$. Given that $\mathbf{F}$ is imperfect the forecast uncertainty as given by equation (3.4b), should be modified to reflect the uncertainty in the model state due to the model error. If the observation error and the model error are uncorrelated, that is, $E(\eta\epsilon_o)=0$, then it is shown that,

$$\mathbf{P}^f(t) = \mathbf{F}'\mathbf{P}^a(t-1)\mathbf{F}'^T + \mathbf{Q}(t) \tag{3.7}$$

In equation (3.7) the first term on the rhs gives uncertainty in the forecast because the initial condition is wrong. The second term ($\mathbf{Q}$(t)) gives the uncertainty in forecast because the model is imperfect. It is assumed that the total uncertainty in the forecast can be obtained by linearly combining the uncertainty due to initial condition and that due to model error. How can one modify the equations of the EnKF (equations (3.5a) to (3.5e)) given an imperfect model to account for model error? The forecast covariance obtained by such a modification should be equal to that given by equation (3.7). This is possible if one adds noise to each ensemble member given by equation (3.5a), which is consistent with the model error statistics. Thus, for $i=1,\ldots,N$,

$$\mathbf{x}_i^f(t) = \mathbf{x}_i^f(t) + \varsigma_i \tag{3.8}$$

where $E(\varsigma)=0$ and $E(\varsigma\varsigma^T)=\mathbf{Q}(t)$ (Palmer and Hagedorn, 2006).

There are, broadly speaking, two different approaches to account for model error in the EnKF. The first approach is to account for the model error so that model error noise is integrated along with the ensemble of initial conditions (Hansen and Smith, 2001). This is equivalent to modifying the model $\mathbf{F}$ so that the model error noise is integrated forward in time. The

model $\mathbf{F}$ as implemented in equation (3.5a) is deterministic. The model $\mathbf{F}$ should be modified to a stochastic dynamic form to include additional forcings due to model error. This approach is equivalent to solving the Fokker-Planck equation. The Fokker-Planck equation evolves the PDF of initial condition under the model dynamics and simultaneously modifies the evolved PDF to reflect the model error. If the model is perfect then the Fokker-Planck equation reduces to the Liouville equation (Palmer and Hagedorn, 2006). This is the correct approach to accounting for model error in the EnKF. For a given model $\mathbf{F}$, the implementation of this approach would imply recasting $\mathbf{F}$ as stochastic equations. In an operational model, the equations of the operational model would have to be recoded. There has been little work done to explore this approach in operational models. Conceptually, this approach has been explained (Penland, 2003) and its potential elucidated (Palmer, 2001). Some approximate, but efficient numerical algorithms to integrate stochastic equations have been proposed (Hansen and Penland, 2006).

The second approach is to integrate the deterministic model without the stochastic model noise and add noise to the forecast ensemble at the data assimilation time (Palmer and Hagedorn, 2006). This noise should be structured so that it compensates for missing model error that would have been present if the first approach had been employed. This approach of additive noise at the data assimilation step has been used in a real weather model by Hamill and Whitaker (2005), (hereafter HW). This approach has also been tested in another model by Houtekamer et al. (2005), (hereafter HM). In HW, one of the sources of model error is the low resolution the model used. In HW, the differences between a high resolution and a low resolution model have been used to construct the additive model error noise. In HM, the additive model error has been used to compensate for model error arising from different sources. In both HW and HM, the model error (that the additive noise seeks to compensate) is potentially multiplicative. The additive noise approach appears to give good results in the EnKF framework. This approach of adding noise to forecast ensemble members at the data assimilation step is called the *additive error approximation* (AEA). The AEA entails modifying equation (3.5a) so that,

$$\mathbf{x}_i^f(t) = \mathbf{F}[\mathbf{x}_i^a(t-1)] + \varsigma_i \tag{3.9}$$

The $\mathbf{P}^f$ obtained from equation (3.5a) is different from that obtained using equation (3.9). The $\mathbf{P}^f$ obtained from equation (3.9) is in general larger and different in direction than that obtained using equation (3.5a). The first approach to account for model error in EnKF is the correct approach, but it is difficult to implement in real models. The second approach, namely AEA, is only approximate but it is easier to implement in real models.

The equations for the EnKF which employs the AEA approach to account for model error

are,

$$\mathbf{x}_i^f(t) = \mathbf{F}[\mathbf{x}_i^a(t-1)] + \varsigma_i \tag{3.10a}$$

$$\mathbf{P}^f(t) = \frac{1}{N-1}[\mathbf{E}^f(t) - \mathbf{M}^f(t)][\mathbf{E}^f(t) - \mathbf{M}^f(t)]^T \tag{3.10b}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T[\mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t)]^{-1} \tag{3.10c}$$

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f(t) + \mathbf{K}(t)[\mathbf{y}_i^o(t) - \mathbf{H}(t)\mathbf{x}_i^f(t)] \tag{3.10d}$$

$$\mathbf{P}^a(t) = \frac{1}{N-1}[\mathbf{E}^a(t) - \mathbf{M}^a(t)][\mathbf{E}^a(t) - \mathbf{M}^a(t)]^T \tag{3.10e}$$

Here ς_i is sampled from $\mathbf{Q}(t)$. $\mathbf{Q}(t)$ is the probability distribution function that quantifies the uncertainty in the forecast due to the model error. The methods to obtain $\mathbf{Q}(t)$ are discussed in sections 3.3, 3.4, 3.5 and 3.6. The current work proposes and uses a modified form of AEA, named the modified additive error approach (MAEA). The MAEA replaces equations (3.10a) and (3.10b) by the following equations:

$$\mathbf{x}_i^f(t) = \mathbf{F}[\mathbf{x}_i^a(t-1)] \tag{3.11a}$$

$$\mathbf{P}^f(t) = \frac{1}{N-1}[\mathbf{E}^f(t) - \mathbf{M}^f(t)][\mathbf{E}^f(t) - \mathbf{M}^f(t)]^T \tag{3.11b}$$

$$\mathbf{P}^f(t) = \mathbf{P}^f(t) + \mathbf{Q}(t) \tag{3.11c}$$

The ensemble members $\mathbf{x}_i^f(t)$ are resampled from the $\mathbf{P}^f(t)$ given by equation (3.11c). Then equations (3.10c) through (3.10e) are implemented. If the number of ensemble members N is large enough (compared to the dimensionality n of the model), then the AEA and MAEA give the same result, that is, the same forecast and analysis errors. In this work MAEA, is employed for accounting for model error. Because in this work, $N >> n$, using AEA would have given identical results to those presented.

The next question is, how can one estimate $\mathbf{Q}(t)$?

There are a few methods to estimate the model error which are explored in sections 3.3 through 3.6. These are the Covariance Inflation Method (CIM), Static $\mathbf{Q}$ Method (SQM), MultiModel Method (MMM) and Parametric Vector Method (PVM). The MMM and PVM are introduced in this work. The simplest of these methods is the covariance inflation method (CIM). Another method is the static $\mathbf{Q}$ denoted by $\mathbf{Q}_{st}$ and called SQM. $\mathbf{Q}_{st}$ is a state-independent estimate of the model error. The model error is a probabilistic estimate of the sensitivity of the state to the imperfections in the model. In general model error is state-dependent. The estimate of $\mathbf{Q}$ obtained by the multimodel method (MMM) is denoted by $\mathbf{Q}_{mm}$. The MMM is inspired

by suggestions made in Hansen (2002). A new method to estimate $\mathbf{Q}$, named the parametric vector method (PVM), is introduced and investigated.

3.3 Covariance inflation method (CIM)

The covariance inflation method (Anderson and Anderson, 1999), hereafter referred to as CIM, is one of the simplest methods of estimating $\mathbf{Q}$. $\mathbf{Q}$ is a $n \times n$ covariance matrix, where n is the dimensionality of the model. $\mathbf{Q}$ can be constructed if the n eigendirections (or eigenvectors) and the corresponding eigenvalues (or sizes) are known. Two different covariance matrices are said to have the same directionality if they have the same eigendirections. They may have different eigenvalues.

In the covariance inflation method, it is assumed that $\mathbf{Q}$ has the same directionality as $\mathbf{P}^f$. $\mathbf{Q}$ is factored into $\mathbf{P}^f$ with the following equation.

$$\mathbf{P}^f = \boldsymbol{\alpha}\mathbf{P}^f \tag{3.12}$$

$\boldsymbol{\alpha}$ is a vector that modifies the size of $\mathbf{P}^f$. $\boldsymbol{\alpha}$ is called the inflation factor. The elements of $\boldsymbol{\alpha}$ are chosen to suitably modify the size of $\mathbf{P}^f$ in different directions. The size of $\mathbf{P}^f$ in a particular direction can be increased by choosing the corresponding element of $\boldsymbol{\alpha} > 1$. Similarly, the size of $\mathbf{P}^f$ in a particular direction can be decreased by choosing the corresponding element of $\boldsymbol{\alpha} < 1$. The elements of $\boldsymbol{\alpha}$ are also called the *boost factors.*

The CIM is simplistic because it does not change the direction of $\mathbf{P}^f$. It assumes that the model error lies in the same subspace as that spanned by $\mathbf{P}^f$. If all the elements of $\boldsymbol{\alpha}$ are >1, then the covariance inflation method, as the name suggests, increases $\|\mathbf{P}^f\|$. In general, the model error spans a different subspace compared to that spanned by $\mathbf{P}^f$. The boosting of $\mathbf{P}^f$ could subsume the uncertainty due to the model error, if the boost factors are large enough. But this is not guaranteed, especially in higher dimensional models. If $\mathbf{Q}$ lies in the null space of $\mathbf{P}^f$, then even for very high boost factors, the resulting $\mathbf{P}^f$ will not be able to account for the model error uncertainty. The CIM can be looked upon as a method to give more weighting to the observations compared to the first guess at the analysis step (equations (3.10d)). In the next section, the SQM is discussed.

3.4 Static Q method (SQM)

$\mathbf{Q}_{st}$ is a state-independent measure of the model error covariance. The difference in the state due to the difference between the model and the system are sampled at a number of points in the phase space of the model. A covariance is constructed using these differences. This covariance, denoted by $\mathbf{Q}_{st}$ and called the static $\mathbf{Q}$, is a state-independent estimate of the uncertainty in state due to model error.

In the Ikeda system, $\mathbf{Q}_{st}$ due to the parametric model error in μ is estimated as follows. The truth $\mathbf{x}^t$ is stepped forward using equations (2.1a) and (2.1b) (page 23) with the correct value of $\mu = 0.9$ over time τ_{ob}. This results in $\mathbf{x}^t(s + \tau_{ob})$. The same initial state is also stepped forward using the model, that is, using the imperfect value of the parameter say $\mu_{st} = \mu + \Delta\mu$, over τ_{ob}. This results in $\mathbf{x}^t_{\mu_{st}}(s + \tau_{ob})$. The difference $\mathbf{x}^t(s + \tau_{ob}) - \mathbf{x}^t_{\mu_{st}}(s + \tau_{ob})$ is stored. This is repeated for several points in the phase space. The covariance of the stored values of $\mathbf{x}^t(s + \tau_{ob}) - \mathbf{x}^t_{\mu_{st}}(s + \tau_{ob})$ gives $\mathbf{Q}_{st}$. Note that the same initial condition (truth) is stepped forward using the system (i.e. correct value of μ) and the model (with μ_{st}). The estimate of model error uncertainty given by $\mathbf{Q}_{st}$ is state-independent. For a given μ_{st} and τ_{ob} there is one value of $\mathbf{Q}_{st}$. The computational expense for the calculation of $\mathbf{Q}_{st}$ is very small compared to the multimodel method (MMM) which is presented in the next section. In operational models one does not have access to the perfect model. In that case a particular version of the model can be defined to be the perfect model. Then $\mathbf{Q}_{st}$ can be constructed by sampling the differences between this perfect model and a slightly different version of the perfect model.

$\mathbf{Q}_{st}$ is gives a state-independent estimate of model error. The next section introduces the MMM which gives a state-dependent estimate of model error.

3.5 Multimodel method (MMM)

The MMM can be understood by first understanding the interpretation of model error direction. Consider the perfect Ikeda model defined by equations (2.1a) and (2.1b) (page 23) and the correct parameter value $\mu = 0.9$. The attractor corresponding to this perfect model is shown in figure 3-1(a) in black. An imperfect model is defined by equations (2.1a) and (2.1b), but with an imperfect parameter value $\mu = 0.895$. The attractor corresponding to this imperfect model is shown in red in figure 3-1(a). Note that the perfect (black) and imperfect (red) attractors, in general, do not overlap. The error in μ displaces the perfect attractor.

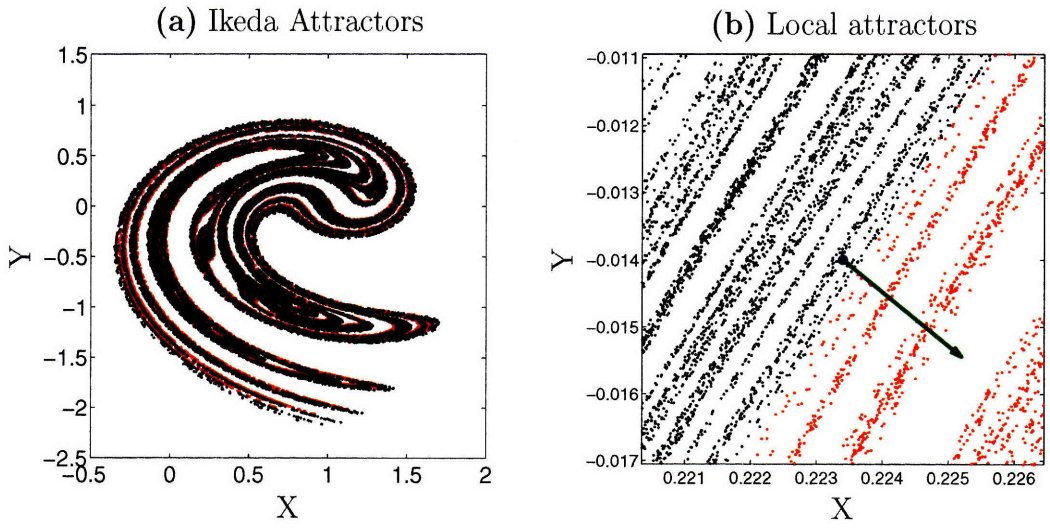


Figure 3-1: Perfect (black) and imperfect (red) Ikeda attractors. The perfect and imperfect attractors in general, do not overlap. **(a)** Black shows the perfect attractor ($\mu = 0.9$) and red shows the imperfect attractor ($\mu = 0.895$). **(a)** Local attractors in the neighborhood of the green dot in panel(a). The green arrow gives an estimate of the displacement of the local attractor due to error in μ.

This displacement is estimated by finding the dislocation of the red dots with respect to the black dots. Note that the direction of this dislocation is different at different locations of the attractor. The attractors in the neighborhood of the green dot $(0.2232, -0.014)$ are magnified and shown in figure 3-1(b). The green arrow shows an estimate of direction of the dislocation of the attractor. The green arrow is the model error direction at this location. The green arrow gives an estimate of the sensitivity of this attractor location ($\mathbf{x}$) to μ. This sensitivity ($\partial\mathbf{x}/\partial\mu$) is state-dependent. The MMM is a method to estimate the state-dependent $\partial\mathbf{x}/\partial\mu$. The calculation of $\mathbf{Q}_{mm}$ in the context of the Ikeda system is described below.

Assuming that the model imperfection is only due to a parameter, the MMM basically involves constructing an ensemble in the parameter space. At the initial time (i.e. $s = 0$), $n_{mm} = 5$ identical ensembles (say $\mathbf{E}_j$, $j = 1, \ldots, 5$) of size $N = 100$ each are constructed. An ensemble of size $n_{mm} = 5$ is constructed in the parameter (μ) space. Let the parameter ensemble members be denoted by μ'_j, $j = 1, \ldots, 5$. I have chosen μ'_j as $\mu'_j = \mu + \Delta\mu_j$ where $\Delta\mu_j = -0.01, -0.005, 0, 0.005, 0.01$. The $\Delta\mu_j$ is a % of the perfect value of $\mu = 0.9$ for $j = 1, \ldots, 5$. Therefore, $\mu'_1 = \mu + \Delta\mu_1 = 0.9 - 0.01 \times 0.9/100$ and so on. The ensemble $\mathbf{E}_j$ is stepped forward using μ'_j. The ensembles in the state space ($\mathbf{E}_j$, $j = 1, \ldots, 5$) constitute the the MultiModel Ensemble (MME). Note that $\Delta\mu_3 = 0.0$ and therefore $\mu'_3 = \mu$, that is, one of the perturbed parameter values happens to be the perfect value of the parameter.

The $j = 3$ member of the MME corresponds to the perfect model. Consequently, $\mathbf{E}_3$ will be referred to as the *perfect model ensemble.*

The EnKF as given by equations (3.5a) to (3.5e) (page 53) is implemented for each of the MME. Each $\mathbf{E}_j$ is stepped forward using equations (2.1a) and (2.1b) (page 23) using the corresponding value of μ'_j. Thus $\mathbf{E}_1$ is stepped forward using μ'_1, $\mathbf{E}_2$ is stepped forward using μ'_2 and so on. It should be remembered that each of $\mathbf{E}_j$ has $N=100$ ensemble members in the state. The stepping forward of $\mathbf{E}_1$ using μ'_1 means stepping forward each of the $N=100$ members of $\mathbf{E}_1$ using μ'_1. $\mathbf{E}_3$ is stepped forward using the correct value of $\mu=0.9$.

The Ikeda system is a discrete system. Therefore, the time is expressed in steps. Let the number of steps between consecutive observations be $\tau_{ob}=1$. Thus the data assimilation is performed at every step. The size of $\mathbf{R}$ is 1% of the attractor size. The *same observations* are assimilated into each of the $n_{mm}=5$ MMEs. The observations are drawn from truth, that is, the local attractor corresponding to $\mu=0.9$. Note that the DA cycles for each of the MMEs are completely independent of each other. In the context of operational DA and forecasting this is similar to different operational centers producing different forecast ensembles. Irrespective of whether the DA is ensemble based, these operational forecast ensembles can be considered to be a MME. Though each of the MMEs is identical at $s=0$, they diverge from each other because each is stepped forward using a different value of μ. The forecast MME at one of the locations in the Ikeda state space is shown in figure 3-2(a). The black dots show the perfect Ikeda attractor, that is, corresponding to $\mu=0.9$. The MME is shown in five different colors. The perfect model ensemble $\mathbf{E}_3$ is shown in the yellow. There are $N=100$ yellow colored dots corresponding to the $N=100$ ensemble members. The red circle (on the yellow ensemble) on the left of the red arrow is the truth $\mathbf{x}^t$. The red colored ensemble is $\mathbf{E}_1$. It shows the ensemble propagated with μ'_1. The cyan colored dots are the ensemble $\mathbf{E}_5$ propagated with μ'_5. The green ensemble ($\mathbf{E}_2$) is bounded by the red ($\mathbf{E}_1$) and yellow ($\mathbf{E}_3$) ensembles. The magenta ensemble ($\mathbf{E}_4$) is bounded by cyan ($\mathbf{E}_5$) and $\mathbf{E}_3$ ensemble. This is because $\mu'_5 < \mu'_4 < \mu'_3$. For similar reasons, the $\mathbf{E}_2$ ensemble (green dots) is bounded by $\mathbf{E}_1$ and $\mathbf{E}_3$. The mean of each of the MMEs is calculated. These means are not shown in figure 3-2(a), but are shown in figure 3-2(b) as colored stars. These forecast means are denoted by $\mathbf{m}^f_j$, $j=1,\ldots,5$. The cyan star is the mean of the cyan ensemble shown in figure 3-2(a), the magenta star is the mean of the magenta ensemble shown in figure 3-2(a) and so on. The covariance of $\mathbf{m}^f_j$, $j=1,\ldots,5$, denoted by $\mathbf{Q}_{mm}$ is computed. $\mathbf{Q}_{mm}$ is shown as the magenta ellipse in figure 3-2(b). The eigendirections of $\mathbf{Q}_{mm}$ are shown as the red and green arrows in figure 3-2(c). The leading eigendirection is shown in red and denoted by $\boldsymbol{u}_{mm}$. $\boldsymbol{u}_{mm}$ is reproduced in figure 3-2(a) as the red arrow. Note that the second leading direction (green arrow) in figure 3-2(c) is quite short compared to the leading direction. Let us try to understand the physical significance of $\boldsymbol{u}_{mm}$.

The MME shown in figure 3-2(a) is the forecast MME at time s. These forecasts were launched from the analyses MME at time $s-1$. Consider a particular member of MME, say $\mathbf{E}_3$. The analysis ensemble $\mathbf{E}_3$ at time s may not be on its attractor. The process of analysis typically

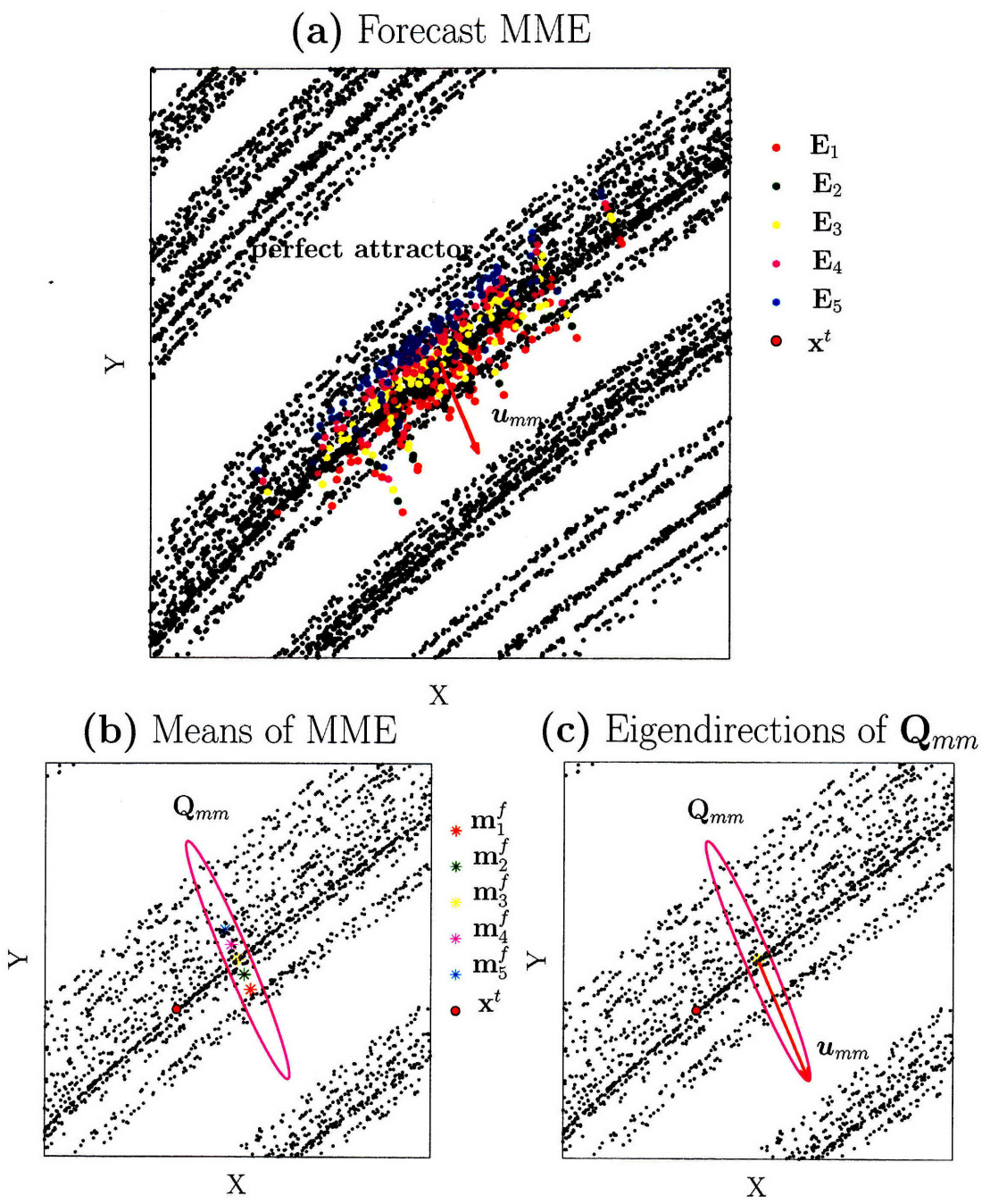


Figure 3-2: Forecast Multimodel Ensemble (MME), the means of MMEs and the multimodel direction $\boldsymbol{u}_{mm}$ (red arrow) at a particular location in the Ikeda phase space. **(a)** Different MMEs are shown in five different colors. The yellow ensemble corresponds to the perfect model. The black dots show the perfect attractor. The yellow ensemble tends to lie on the black dots. Each of the other MME tend to lie on its own attractors, which are not shown. The red arrow shows the multimodel direction $\boldsymbol{u}_{mm}$. $\boldsymbol{u}_{mm}$ emanates from the mean of the perfect ensemble that is shown as the yellow star in panel(b) and panel(c) . **(b)** The means of the forecast MMEs in panel(a) are shown as stars. The attractor region shown is the same as the region near the red arrow in panel(a), but magnified. The covariance of these means is the magenta ellipse denoted by $\mathbf{Q}_{mm}$. **(c)** The eigendirections of $\mathbf{Q}_{mm}$ are shown in red and green. $\mathbf{Q}_{mm}$ from (b) is reproduced here. The leading direction (red) $\boldsymbol{u}_{mm}$ is also shown in (a) in red. $\boldsymbol{u}_{mm}$ estimates $\partial \mathbf{x}/\partial \mu$ at this location of Ikeda attractor.

displaces the ensemble members from its attractor. When each ensemble member of $\mathbf{E}_3$ is propagated forward by equations (2.1a) and (2.1b) (page 23), it tends to fall on *its* attractor, that is, the attractor corresponding to $\mu = 0.9$. The members of the ensemble ($\mathbf{E}_3$) shown in yellow, tend to be on the perfect model attractor (shown by the black dots). Similarly, each of the other MMEs tend to lie on their *own* attractors. These attractors are not shown in figure 3-2(a). The red dots ($\mathbf{E}_1$) tend to lie on the model attractor corresponding to μ_1, the green dots tend to lie on the model attractor corresponding to μ_2 and so on. In short, the spread *across* the MMEs gives the sensitivity of the local attractor structure to the parameter μ. The sensitivity of the local attractor structure is mathematically represented by $\partial\mathbf{x}/\partial\mu$. Therefore the red arrow denoted by $\boldsymbol{u}_{mm}$ is an estimate of $\partial\mathbf{x}/\partial\mu$. $\boldsymbol{u}_{mm}$ is a linear fit to forecast means, $\mathbf{m}_j^f$. The covariance of $\mathbf{m}_j^f$ quantifies the uncertainty in the state due to imperfection in μ. This covariance, denoted by $\mathbf{Q}_{mm}$ is the multimodel $\mathbf{Q}$. $\mathbf{Q}_{mm}$ is a probabilistic estimate of $\partial\mathbf{x}/\partial\mu$. The leading eigendirection of $\mathbf{Q}_{mm}$ is $\boldsymbol{u}_{mm}$. Since the second leading direction (green arrow) in figure 3-2(c) is very short compared to the leading direction, it is assumed that $\partial\mathbf{x}/\partial\mu$ is quantified by the leading direction alone.

The interesting property of $\boldsymbol{u}_{mm}$ is that it neglects the spread within each MME. For example, the cyan ensemble has a spread on its attractor. The $\boldsymbol{u}_{mm}$ direction does not indicate this spread or for that matter the spread in the red or any other MME. In other words, $\boldsymbol{u}_{mm}$ tells how the final state would be affected if the perturbations were in μ only; it ignores the effect of initial perturbations in state on the final state.

3.6 Parametric vector method (PVM)

As seen in the last section (3.5), $\mathbf{Q}_{mm}$ gives a state-dependent estimate of $\partial\mathbf{x}/\partial\mu$. As the dispersion within the parameters used to step forward the MME decreases, $\boldsymbol{u}_{mm}$ gives the linear sensitivity of the state $\mathbf{x}$ to the parameter μ. In principle, this sensitivity should be obtainable by using the linearization tools described in section 2.3.2 (page 26). In this section, a new method (PVM) to obtain an estimate of model error based on this linear sensitivity is presented. The linearization tools presented in section 2.3.2, were used to find the sensitivity of the final time state to perturbations in the initial time state. These linearization tools are modified appropriately to find the linear sensitivity of the final state to the perturbations in parameter μ. The next section (3.6.1) introduces and describes the notion of parametric singular vectors.

3.6.1 Parametric singular vectors

The concept of SVD, singular vectors and their application to linear error dynamics in the Ikeda system was presented in section 2.3.2 (page 26). In this section, the concept of SVD and singular vectors is extended to include the parameter μ.

Consider the Ikeda system given by equations (2.1a) and (2.1b) (page 23). Consider augmenting the state with the parameter μ. The concept of augmenting the state with the parameter is briefly discussed in section 2.5 (page 41). The parameter is treated as a variable, which increases the dimensionality of the phase space by 1 (In general the dimensionality of the phase space increases by the number of parameters used to augment the state). The equations for the augmented state, denoted by $\mathbf{w}$ are given by equations (2.9a), (2.9b) and (2.9c), which are reproduced below.

$$x_{s+1} = F_1(x_s, y_s) = 1 + \mu(x_s \cos\theta - y_s \sin\theta) \tag{3.13a}$$

$$y_{s+1} = F_2(x_s, y_s) = \mu(x_s \sin\theta + y_s \cos\theta) \tag{3.13b}$$

$$\mu_{s+1} = F_3 = \mu_s \tag{3.13c}$$

Then the augmented linear propagator $\mathbf{M^w}$ is given by equation (2.10) (page 42). Consider the SVD of $\mathbf{M^w}$. Let the final time singular vector matrix, the initial time singular vector matrix and the singular value matrix resulting from the SVD($\mathbf{M^w}$) be denoted by $\mathbf{U^w}$,$\mathbf{V^w}$ and $\Sigma^{\mathbf{w}}$. Note that $\mathbf{U^w}$,$\mathbf{V^w}$ and $\Sigma^{\mathbf{w}}$ are 3×3 matrices. In general the dimensionality of $\mathbf{U^w}$,$\mathbf{V^w}$ and $\Sigma^{\mathbf{w}}$ depends upon the number of parameters used to augment the state. In the case considered here only one parameter, namely μ is used to augment the state. Then, analogous to equation (2.4) (page 26),

$$\mathbf{M^w} = \mathbf{U^w}\mathbf{\Sigma^w}\mathbf{V^{w}}^T \tag{3.14}$$

The final time leading singular vector in $\mathbf{U}$ gives the sensitivity of the augmented state to initial time perturbations in the augmented state $\mathbf{w}$. Consider the following projection operators $\mathbf{Pr}^i$ and $\mathbf{Pr}^f$.

$$\mathbf{Pr}^i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.15}$$

$$\mathbf{Pr}^f = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.16}$$

(In the case of multiple parameters $\mathbf{Pr}^i$ should contain relative weightings given to each parameter in the place of non-zero entries.) Let,

$$\mathbf{M}^{\mu\mathbf{x}} = \mathbf{Pr}^f \mathbf{M}^{\mathbf{w}} \mathbf{Pr}^i \tag{3.17}$$

i and f in $\mathbf{Pr}^i$ and $\mathbf{Pr}^f$ denote initial and final time respectively.

$$\mathbf{M}^{\mu\mathbf{x}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} F'_{1x} & F'_{1y} & F'_{1\mu} \\ F'_{2x} & F'_{2y} & F'_{2\mu} \\ F'_{3x} & F'_{3y} & F'_{3\mu} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.18}$$

Carrying out the matrix multiplication on the rhs and using equations (3.13a) and (3.13b),

$$\mathbf{M}^{\mu\mathbf{x}} = \begin{bmatrix} 0 & 0 & F'_{1\mu} \\ 0 & 0 & F'_{2\mu} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \partial x_{s+1}/\partial\mu \\ 0 & 0 & \partial y_{s+1}/\partial\mu \\ 0 & 0 & 0 \end{bmatrix} \tag{3.19}$$

$$\mathbf{M}^{\mu\mathbf{x}}\Delta\mathbf{w} = \begin{bmatrix} 0 & 0 & \partial x_{s+1}/\partial\mu \\ 0 & 0 & \partial y_{s+1}/\partial\mu \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta\mu \end{bmatrix}$$

$$\mathbf{M}^{\mu\mathbf{x}}\Delta\mathbf{w} = \begin{bmatrix} (\partial x_{s+1}/\partial\mu)\Delta x \\ (\partial y_{s+1}/\partial\mu)\Delta y \\ 0 \end{bmatrix}$$

Consider an initial time error in $\mathbf{w}$ given by $\Delta\mathbf{w} = [\Delta x\ \Delta y\ \Delta\mu]$. The application of $\mathbf{M}^{\mu\mathbf{x}}$ to $\Delta\mathbf{w}$ gives the change in $\mathbf{x}$ at final time due to the perturbation in μ, $\Delta\mu$.

Taking the first two components of the above equation and writing the vector as a row,

$$\mathbf{M}^{\mu\mathbf{x}}\Delta\mathbf{w} = \begin{bmatrix} \frac{\partial x_{s+1}}{\partial\mu}\Delta x & \frac{\partial y_{s+1}}{\partial\mu}\Delta y \end{bmatrix} \tag{3.20}$$

The rhs of equation (3.20) should be read as the vector change in $\mathbf{x}$ at final time due to the change in μ alone. $\mathbf{x}_{s+1}$ is a function of $\mathbf{x}_s$ and μ. The above equation gives the partial sensitivity of $\mathbf{x}_{s+1}$ to μ, thus ignoring the effect of perturbation in $\mathbf{x}$ at initial time. The *partial sensitivity* is also called the *adjoint sensitivity.* $\mathbf{M}^{\mu\mathbf{x}}$ could be called the adjoint operator. Henceforth $\mathbf{M}^{\mu\mathbf{x}}$ is referred to as the parametric linear propagator rather than the adjoint operator.

The vector on the rhs of the equation (3.20) can be obtained by performing $SVD(\mathbf{M}^{\mu\mathbf{x}})$. Let

the leading final time singular vector be denoted by $\boldsymbol{u}_1^{\mu\mathbf{x}}$, the leading time singular vector be denoted by $\boldsymbol{v}_1^{\mu\mathbf{x}}$ and the corresponding singular value be defined by $\sigma_1^{\mu\mathbf{x}}$. Then the following equation holds true.

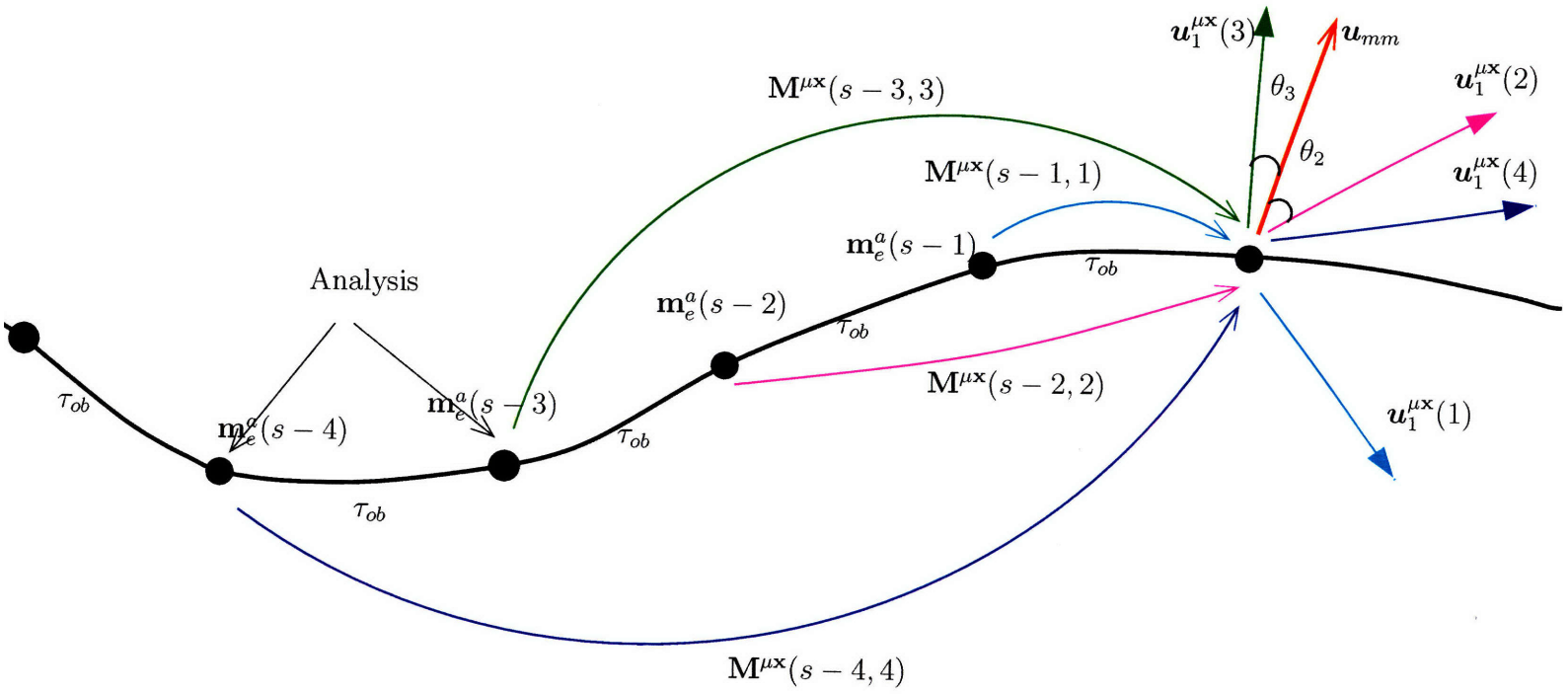
$$\mathbf{M}^{\mu\mathbf{x}}\boldsymbol{v}_1^{\mu\mathbf{x}} = \sigma_1^{\mu\mathbf{x}}\boldsymbol{u}_1^{\mu\mathbf{x}}$$

The final time leading singular vector denoted by $\boldsymbol{u}_1^{\mu\mathbf{x}}$ is called the *parametric singular vector* because the initial time perturbations are in parameter only and the final time perturbations are in state only (which is emphasized by the superscript $\mu\mathbf{x}$). If $\mathbf{M}^{\mu\mathbf{x}}$ is constructed over an optimization time of τ_{op}, then the sensitivity of the state at time $s=\tau_{op}$ to the perturbation in μ at time $t=0$ is given by $\boldsymbol{u}_1^{\mu\mathbf{x}}$. Note that the initial time singular vector is $\boldsymbol{v}_1^{\mu\mathbf{x}} = [0\ 0\ 1]$, irrespective of the τ_{op}. This is because the error at initial time is in μ only. Since the definition of $\mathbf{M}^{\mu\mathbf{x}}$ permits only one initial time singular direction $\boldsymbol{v}_1^{\mu\mathbf{x}}$, the SVD($\mathbf{M}^{\mu\mathbf{x}}$) results in only one parametric vector $\boldsymbol{u}_1^{\mu\mathbf{x}}$.

The concept of parametric vectors is further explained as follows. An ensemble in state space with $N=100$ is constructed. Let this ensemble be denoted by $\mathbf{E}_e$. The ensemble $\mathbf{E}_e$ is stepped forward with an imperfect value of μ, given by $\mu_e=\mu+\Delta\mu_e$, where $\Delta\mu_e=-0.08$. The analysis is performed after every $\tau_{ob}=1$.

The DA experiment for $\mathbf{E}_e$ is run in parallel with the multimodel DA experiment described in section 3.5 (page 58). This implies that the the initial ensemble for $\mathbf{E}_e$ is constructed at the same location in state space as that of the MME. The DA cycles for the MME are independent of the DA cycles for $\mathbf{E}_e$ except that the *same observations* are assimilated into the MME and $\mathbf{E}_e$, every $\tau_{ob}=1$. This means that the analysis MME at a given time (i.e. a given DA cycle) corresponds to the same truth as $\mathbf{E}_e$ at that time. Consequently, the forecast MME tends to lie at the same state location as the forecast $\mathbf{E}_e$, for the same forecast lead time. Let the mean of the analysis $\mathbf{E}_e$ at time s be denoted by $\mathbf{m}_e^a(s)$. As the DA cycles proceed, the $\mathbf{m}_e^a(s)$ is stored for each DA cycle. As a result at any given time, we have access to the analysis means $\mathbf{m}_e^a(s)$ in the past. To put it concretely, at time s, we have access to $\mathbf{m}_e^a(s-\tau_{ob})$, $\mathbf{m}_e^a(s-2\tau_{ob})$, $\mathbf{m}_e^a(s-3\tau_{ob})$ and so on. The τ_{ob} is chosen to be $\tau_{ob}=1$. Therefore we have access to $\mathbf{m}_e^a(s-1)$, $\mathbf{m}_e^a(s-2)$, $\mathbf{m}_e^a(s-3)$ and so on. This procedure is shown schematically in figure 3-3 (page 66). The black curve shows the trajectory of the forecasts through the Ikeda phase space. The rightmost black circle is time s. The analysis in the past are marked $\mathbf{m}_e^a(s-1)$, $\mathbf{m}_e^a(s-2)$, $\mathbf{m}_e^a(s-3)$ and $\mathbf{m}_e^a(s-4)$. The analysis at time $(s-1)$ is the latest analysis.

Consider the analysis at time $(s-1)$. The parametric linear propagator (given by equation (3.17)) linearized around $\mathbf{m}_e^a(s-1)$ is constructed over an optimization time of $\tau_{op}=1$. This linear propagator is denoted by $\mathbf{M}^{\mu\mathbf{x}}(s-1,1)$.



-3: A cartoon of the parametric vector method (PVM). The black curve shows a trajectory in the phase space. The blac
nalysis means. The parametric linear propagators are launched from analysis means in the past such that they verify a
ed arrow shows the multimodel direction. The cyan, magenta, green and blue are parametric singular vectors corresp
ent lag times.

In the notation $\mathbf{M}^{\mu\mathbf{x}}(s-1,1)$, $s-1$ denotes the time from which $\mathbf{M}^{\mu\mathbf{x}}$ is launched and 1 denotes the optimization time. $\mathbf{M}^{\mu\mathbf{x}}(s-1,1)$ verifies at time s. The parametric vector obtained from $\mathbf{M}^{\mu\mathbf{x}}(s-1,1)$ is valid at time s. This parametric vector is shown as cyan arrow, named $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ in figure 3-3. Consider the analysis at time $s-2$. $\mathbf{M}^{\mu\mathbf{x}}(s-2,2)$ is constructed linearized around $\mathbf{m}_e^a(s-2)$ over an optimization time of $\tau_{op}=2$. This linear propagator launched at time $s-2$ verifies at time s. The parametric vector corresponding to $\mathbf{M}^{\mu\mathbf{x}}(s-2,2)$ is shown in magenta color and is named $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$.

Similarly, the linear propagators $\mathbf{M}^{\mu\mathbf{x}}(s-3,3)$ and $\mathbf{M}^{\mu\mathbf{x}}(s-4,4)$ are constructed linearized around $\mathbf{m}_e^a(s-3)$ and $\mathbf{m}_e^a(s-4)$ respectively. The parametric vectors $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ and $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ are shown as green and blue arrows respectively.

In short, the parametric linear propagators, $\mathbf{M}^{\mu\mathbf{x}}(s-\tau_{op},\tau_{op})$ are constructed linearized around $\mathbf{m}_e^a(s-\tau_{op})$ for $\tau_{op}=1,\ldots,4$. The parametric vectors $\boldsymbol{u}_1^{\mu\mathbf{x}}(\tau_{op})$, for $\tau_{op}=1,\ldots,4$, are valid at time s. From the point of view of the the state location corresponding to time s, the parametric vectors have arrived with increasing time *lags* from the past. For example, the parametric vector $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ has arrived with a time lag of $\tau_{op}=1$ and $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ has arrived with a time lag of $\tau_{op}=2$.

$\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ gives the sensitivity of the state at s to perturbation in μ at time $s-1$. The $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ gives the sensitivity of the state at s to perturbation in μ at time $s-2$ and so on.

Because the MME DA cycle is run on the side, we have access to the multimodel direction $\boldsymbol{u}_{mm}$ at time s. The $\boldsymbol{u}_{mm}$ is shown as the red arrow. As described in the section 3.5, $\boldsymbol{u}_{mm}$ is an estimate of $\partial\mathbf{x}/\partial\mu$. But $\boldsymbol{u}_1^{\mu\mathbf{x}}$ is also an estimate of $\partial\mathbf{x}/\partial\mu$. How does $\boldsymbol{u}_1^{\mu\mathbf{x}}$ compare with $\boldsymbol{u}_{mm}$ as an estimate of $\partial\mathbf{x}/\partial\mu$? In fact, at a given time s we have four different $\boldsymbol{u}_1^{\mu\mathbf{x}}$ which estimate $\partial\mathbf{x}/\partial\mu$. Which of these four, if any, is correct? These questions are very interesting and are discussed further in the next section (3.6.2).

3.6.2 Optimal model error time

Let us assume that the $\boldsymbol{u}_{mm}$ is the best available estimate of $\partial\mathbf{x}/\partial\mu$. Then the question of which of the four available $\boldsymbol{u}_1^{\mu\mathbf{x}}$ best estimates $\partial\mathbf{x}/\partial\mu$ at a given state is answered by calculating the angles made by a particular $\boldsymbol{u}_1^{\mu\mathbf{x}}$ with $\boldsymbol{u}_{mm}$. The angle θ between a given $\boldsymbol{u}_1^{\mu\mathbf{x}}$ and $\boldsymbol{u}_{mm}$ is calculated. The angle between $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ and $\boldsymbol{u}_{mm}$ is denoted by θ_1. The smaller the value of θ_1 better the estimate of $\partial\mathbf{x}/\partial\mu$ as given by $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$. Similarly, the angles θ_2 (angle between $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ and $\boldsymbol{u}_{mm}$), θ_3 (angle between $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ and $\boldsymbol{u}_{mm}$) and θ_4 (angle between $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ and $\boldsymbol{u}_{mm}$) are calculated. The idea of using θ to find the best $\boldsymbol{u}_1^{\mu\mathbf{x}}$ is shown schematically in figure 3-3 (page 66), in which θ_3 and θ_2 are shown.

Consider figure 3-4 (page 69). The panel (a) shows the idea of calculating the angle between the $\boldsymbol{u}_{mm}$ and the parametric vectors at a particular DA cycle in the Ikeda model. This DA experiment was explained in the last section (3.6.1). Figure 3-4(a) shows the $\boldsymbol{u}_{mm}$ (red arrow) at a particular location in the Ikeda state space at time s. The cyan, magenta, yellow, green and red dots are the forecast MME at time s. These forecast MMEs were launched from the analysis MME at time $s-1$. The black dots show the perfect attractor. The four $\boldsymbol{u}_1^{\mu\mathbf{x}}$ are shown as the cyan, magenta, green and the blue arrow. These $\boldsymbol{u}_1^{\mu\mathbf{x}}$ have arrived from the "past" with different time lags. Consider the angle between the blue and the red arrow. This angle $\theta_4 = 6.6°$. θ_4 is the least among θ_1, θ_2, θ_3 and θ_4. For this particular state, $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ best captures the model error direction compared to $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$, $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ and $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$. This is because $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ lies closest to $\boldsymbol{u}_{mm}$ for this state. The implicit assumption is that $\boldsymbol{u}_{mm}$ is a good estimate of the model error direction. The cyan arrow $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ makes an angle of $\theta_1 = 37.6°$ with $\boldsymbol{u}_{mm}$. Among the parametric vectors, $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ is the second best estimate of the model error direction. The parametric vectors $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ and $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ do not give a good estimate of the model error direction. They make large angles with $\boldsymbol{u}_{mm}$ ($\theta_2 = 56.7°$ and $\theta_3 = 49.5°$). For this particular state $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ is the best estimate of $\partial\mathbf{x}/\partial\mu$.

Consider another state location at a particular time s shown in the figure 3-4(b). For this particular state location $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ makes the least angle with $\boldsymbol{u}_{mm}$ ($\theta_3 = 4.5$). Therefore, $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ is the best estimate of the model error direction at this location in state space. The parametric vector $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ ($\theta_4 = 4.6$) is almost as good an estimate of the model error direction as $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$. The parametric vector $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ is the worst estimate of the model error direction ($\theta_4 = 53.4$) for this state.

Panel (c) shows a particular location in state space where $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ ($\theta_2 = 13.9$) is the best estimate of the model error direction. Panel(d) shows an example where $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ makes the smallest angle with $\boldsymbol{u}_{mm}$ ($\theta_1 = 8.6$).

The examples presented in figure 3-4 demonstrate that the *parametric vector* that best estimates the model error direction is *state-dependent*. It is possible that at some state $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ best estimates $\partial\mathbf{x}/\partial\mu$, while at some other states it is $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ and for yet some other states $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ best estimates $\partial\mathbf{x}/\partial\mu$. The optimization time required for a parametric vector to point correctly in the direction of model error is state-dependent. The optimization time required for the parametric vector to correctly capture the model error direction is called the *optimal time*. The difference between optimization time and optimal time should be noted. The optimization time is the time over which a linear propagator is constructed. The optimization time is a well established terminology in the scientific literature. The optimal time is a definition introduced in the present work. The next question is, what it the physical mechanism that decides the state-dependent optimal time? If we can understand this mechanism, then

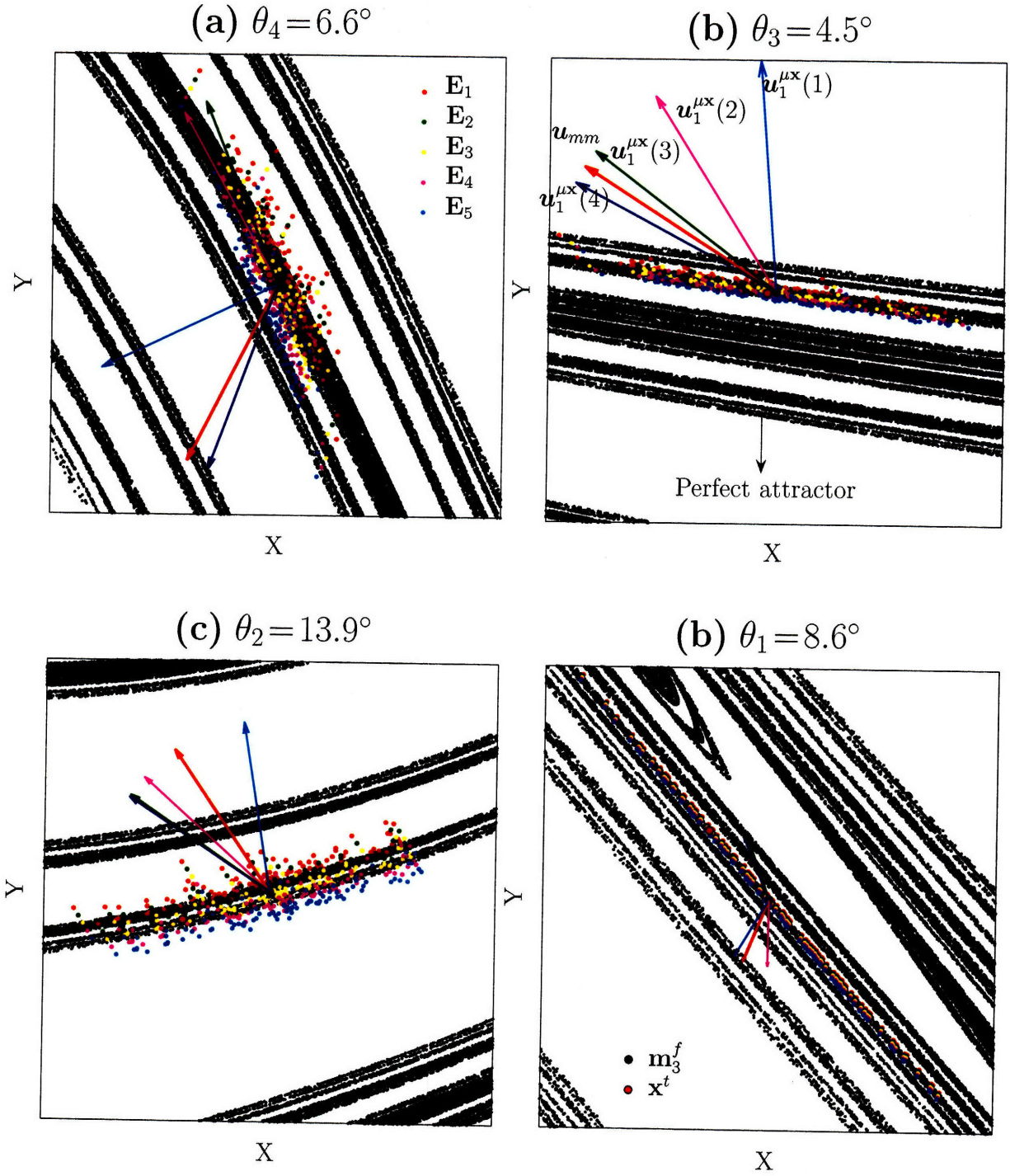


Figure 3-4: Parametric singular vectors and $\boldsymbol{u}_{mm}$ at four different locations in the Ikeda phase space. The red arrow in each panel shows the multimodel direction. Parametric vectors with different lags make the smallest angle with $\boldsymbol{u}_{mm}$ at different locations.
(a) The blue (lag 4) makes the smallest angle (6.6°). The magenta (lag 2) and green (lag 3) are post-optimal. The cyan (lag 1) is pre-optimal.**(b)** The green (lag 3) makes the smallest angle (4.5°). **(c)** The magenta (lag 2) makes the smallest angle (13.9°). **(d)** The cyan (lag 1) makes the smallest angle (8.6°).

possibly we can formulate a criterion to find the state-dependent optimal time. This physical mechanism involves interesting chaotic dynamics.

Consider a DA cycle of the EnKF for a particular model. The forecast ensemble tends to lie on the model attractor. The analysis performed on each ensemble member involves the blending of a perturbed observation and the forecast ensemble member. The analysis step typically tends to put the analysis ensemble member at an off-attractor site. This is because the analysis step moves the analysis closer to the estimate (either the observation or forecast) that has the smaller uncertainty (given by $\|\mathbf{R}\|$ and $\|\mathbf{P}^f\|$). In general, the analysis ensemble member is moved away from the forecast ensemble member. Thus analysis ensemble member is moved away from the model attractor. This is true of all analysis members and therefore the mean analysis tends to be off-attractor.

As an example, consider the analysis mean $\mathbf{m}^a(s-4)$ as shown schematically in figure 3-3 (page 66). Suppose $\mathbf{m}^a(s-4)$ is off the attractor. As it is propagated ahead in time, $\mathbf{m}^a(s-4)$ tends to fall towards the attractor. The time required for $\mathbf{m}^a(s-4)$ to get on the attractor is defined as the *transient time*. Let us suppose that the transient time for $\mathbf{m}^a(s-4)$ is 4 steps. Then $\mathbf{m}^a(s-4)$ gets on the attractor at time s, shown in figure 3-3 (page 66). Then $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ correctly estimates $\partial\mathbf{x}/\partial\mu$. We should remember that the model error direction is the displacement of the *local attractor* with respect to μ. $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ correctly estimates the model error direction because $\mathbf{m}^a(s-4)$ is *on the attractor* at time s. I have referred to the sensitivity $\partial\mathbf{x}/\partial\mu$ as the model error direction. To be more precise, $\partial\mathbf{x}/\partial\mu$, as estimated by the parametric vector, gives the model error direction at the state $\mathbf{x}$ *if* $\mathbf{x}$ *is on the attractor.* What does $\partial\mathbf{x}/\partial\mu$ give if $\mathbf{x}$ is not on the attractor?

Let us suppose that the transient time for $\mathbf{m}^a(s-4)$ is 5 steps rather than 4 steps. In this case, $\mathbf{m}^a(s-4)$ is not on the attractor at time s; it is off the attractor. Then $\partial\mathbf{x}/\partial\mu$ as estimated by $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ does not give the sensitivity of the local attractor to μ; it gives the sensitivity of the off-attractor point (wherever $\mathbf{m}^a(s-4)$ lies) to μ. The sensitivity of the off-attractor point is *not* the model error direction. The direction given by $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ is a transient sensitivity, which we are not interested in.

Let us consider figure 3-4 (page 69)(a). Among the 4 parametric vectors shown the model error direction is most correctly given by $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ (blue arrow). The optimal time for $\mathbf{m}^a(s-4)$ corresponding to this state is 4 steps. The propagated $\mathbf{m}^a(s-4)$ is on the attractor at the state location shown.

Contrast the blue arrow with the cyan arrow ($\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$), which has failed to capture the model error direction. It appears that the propagated $\mathbf{m}^a(s-1)$ is undergoing a transient; it has yet to get on the attractor. The transient time for $\mathbf{m}^a(s-1)$ is greater than 1 step. The transient

phenomenon is in evidence in each of the panels. In figure(b), $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ and $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ appear to be undergoing a transient phase while $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ and $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ give the approximately the correct model error direction because the propagated $\mathbf{m}^a(s-3)$ and $\mathbf{m}^a(s-4)$ are on the attractor.

The state-dependent optimal time is the state-dependent transient time. Let the state-dependent optimal time be denoted by τ_{om}. If the state-dependent optimal time is known then the model error direction is recovered if $\tau_{op}=\tau_{om}$ is used to construct the $\mathbf{M}^{\mu\mathbf{x}}$ linearized around the appropriate mean analysis in the past. If $\tau_{op} < \tau_{om}$, then one obtains the transient sensitivity. What happens if $\tau_{op} > \tau_{om}$?

It is observed that if $\tau_{op} > \tau_{om}$ then the parametric vector tends to align *along* the local attractor direction. Let us inspect figure 3-4 (page 69)(a), more closely. The correct model error direction is captured by $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$. The cyan arrow $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ is undergoing a transient. Look at the magenta arrow ($\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$). It is pointing in the direction of the local attractor. For $\mathbf{m}^a(s-2)$, apparently $\tau_{om} < 2$. The optimization time of $\tau_{op}=2$ is *too long* compared to τ_{om}. The direction given by $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ does not give information about the model error, rather it gives information about the direction of the local attractor. Similarly the green arrow ($\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$) tends to point in the direction of the attractor. Apparently, the transient time corresponding to $\mathbf{m}^a(s-3)$ is $\tau_{om} > 2$.

The optimization time, τ_{op} used for a particular state can be divided into 3 categories, depending on the optimal time for that particular state. If $\tau_{op} < \tau_{om}$ then the corresponding $\boldsymbol{u}_1^{\mu\mathbf{x}}$ is undergoing a transient and is termed *pre-optimal.* If $\tau_{op}=\tau_{om}$ then the corresponding $\boldsymbol{u}_1^{\mu\mathbf{x}}$ is termed *optimal*; $\boldsymbol{u}_1^{\mu\mathbf{x}}$ correctly gives the model error direction. If $\tau_{op} > \tau_{om}$ then the $\boldsymbol{u}_1^{\mu\mathbf{x}}$ is termed *post-optimal.* The directions given by pre-optimal and post-optimal parametric vectors are not useful to obtain the model error direction. The optimal parametric vector is the important direction.

Let us try to further understand the mechanism of optimal time. Consider an initial condition given by an analysis mean. As soon as the initial condition is stepped forward using a perturbation in μ, an error is induced in the state. This error is termed the *model-induced state error.* This model-induced state error grows with time and is part of the total model error signal given by the parametric vector. For a given τ_{op}, the *pure model error* is defined as the total model error signal *less* the model-induced state error. There is an interesting trade off between the model-induced state error and the pure model error signal that decides the optimal time.

As the state is stepped forward, the model-induced state error and the pure model error evolve in time. At a particular time, the pure model error maximizes with respect to the model-induced state error. This is the optimal time when the state arrives on the attractor.

During the pre-optimal phase, the pure model error undergoes transient behavior. At the optimal time, the transient behavior of $\boldsymbol{u}_1^{\mu\mathbf{x}}$ ends. For $\tau_{op} > \tau_{om}$, the model-induced state error overwhelms the pure signal and therefore the total model error signal is dominated by the model-induced state error. As time increases after τ_{om} the model-induced state error is huge compared to the pure signal and therefore $\boldsymbol{u}_1^{\mu\mathbf{x}}$ points in the direction of the local attractor structure. This can be mathematically understood in the light of the discussion in section 2.5. Equation 2.18 gives the parametric singular vector, $\boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}}$. $\|\boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}}\|^2$ is the magnitude of this parametric singular vector that is given by equation (2.20). The pure and model-induced state error signals are embedded in $\boldsymbol{\epsilon}_{s+1}^{\mu\mathbf{x}}$. It is not clear how these signals can be separated.

It has to be understood that neither the pre-optimal nor the post-optimal vector is "wrong". The pre-optimal vector gives the transient sensitivity, that is, the sensitivity of an off-attractor location to μ. But it is not the sensitivity one is interested in — one is interested in sensitivity of the on-attractor point which is the model error direction. The post-optimal vector gives the on-attractor sensitivity but it is not local. We are interested in the local sensitivity of the attractor to the parameter. The optimal vector gives the sensitivity that is both on-attractor and local.

In the next section, a methodology to construct $\mathbf{Q}$ using the concept of parametric vector is described.

3.6.3 Construction of parametric Q

The criterion to calculate the state-dependent optimal time is not known. Therefore the average optimal time for the Ikeda system is calculated by using the angle between $\boldsymbol{u}_{mm}$ and $\boldsymbol{u}_1^{\mu\mathbf{x}}$.

The DA experiment is run for 14000 steps in the Ikeda system. For each step θ_1, θ_2, θ_3 and θ_4 are calculated. The histogram of these angles is shown in figure 3-5 (page 73) from panel (a) to (d). The y-axis shows the fraction, say fr of total number of DA steps. Panel 3-5(a) shows that $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ makes an angle of about $50-60°$ for about 27% of the DA steps. $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$ is within $10°$ of $\boldsymbol{u}_{mm}$ for less that 5% of time. The fraction of time when θ is within $0°-10°$ is denoted by fr^{10}. Panel (b) shows that $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$ is within $10°$ of $\boldsymbol{u}_{mm}$ for about 20% of time. The panels (c) and (d) show the histograms for θ_3 and θ_4 respectively. The fr^{10} has increased going from θ_1 to θ_3 from about 3% to about 45%. fr^{10} has decreased going from θ_3 (45%) to θ_4 (40%). This trend of decrease in fr^{10} continues for θ_5, θ_6 and so on (results not shown). Therefore, on average, the "closeness" of $\boldsymbol{u}_1^{\mu\mathbf{x}}(\tau_{op})$ to $\boldsymbol{u}_{mm}$ has increased going from $\tau_{op}=1$ to $\tau_{op}=3$ and decreased for $\tau_{op} > 3$. This means that for the Ikeda system, it takes 3 steps for the transient to "die out", on average. The average optimal time is $\overline{\tau_{om}}=3$ when data are assimilated every

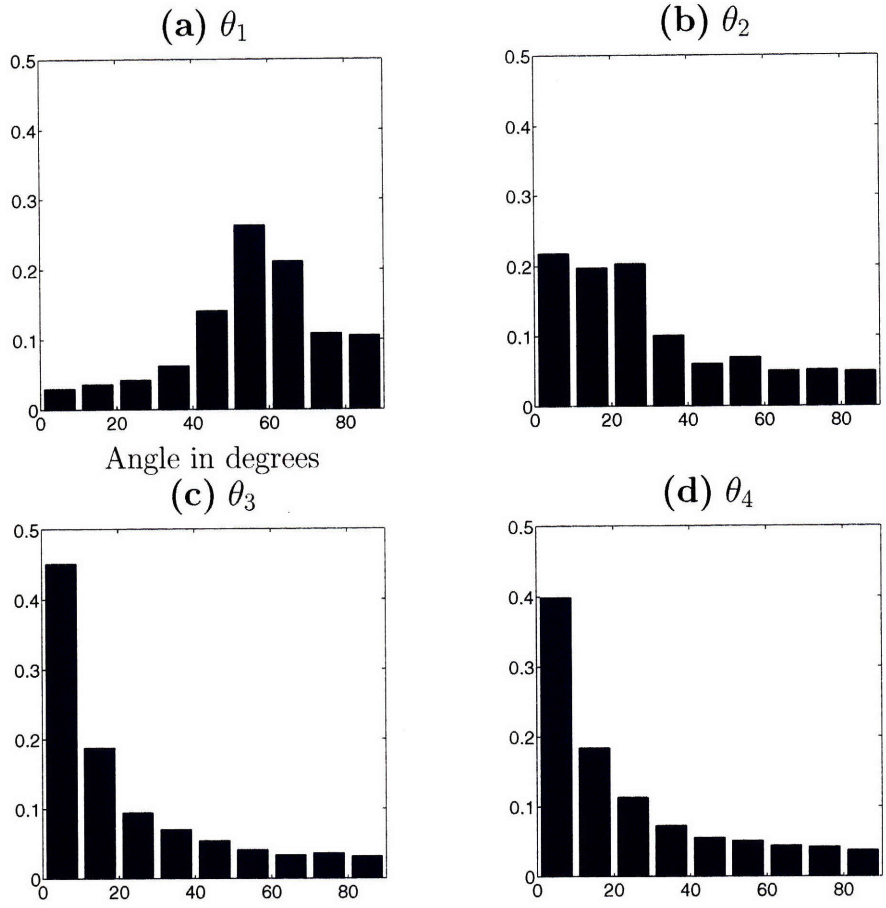


Figure 3-5: **(a)** to **(d)** show the distribution of angles made by $\boldsymbol{u}_1^{\mu\mathbf{x}}(1)$, $\boldsymbol{u}_1^{\mu\mathbf{x}}(2)$, $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ and $\boldsymbol{u}_1^{\mu\mathbf{x}}(4)$ with $\boldsymbol{u}_{mm}$ respectively, for $\tau_{ob}=1$

$\tau_{ob}=1$ steps. For a given time, on average $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ makes the least angle with $\boldsymbol{u}_{mm}$. Obviously, $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ need not always make the least angle with $\boldsymbol{u}_{mm}$. Let θ_{min} denote the least angle out of $\theta_{\tau_{om}}$, $\tau_{om}=1,\ldots,4$. Figure 3-6(a) shows the distribution of the θ_{min}. Note that the scale on the y-axis goes from 0 to 1. Comparing figures 3-5(c) and 3-6(a), $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ fails to give the correct model error direction about 25% of the time. Panel 3-6(b) shows the histogram of the angle θ_{st} made by $\boldsymbol{u}_{st}$ (the leading direction of $\mathbf{Q}_{st}$) with $\boldsymbol{u}_{mm}$. $\boldsymbol{u}_{st}$ aligns with $\boldsymbol{u}_{mm}$ about 12% of the time. Comparing figures 3-5(c) and figure 3-6(b). $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ outperforms $\boldsymbol{u}_{st}$ in giving the correct model error direction about 33% of the time.

The data assimilation experiment to determine the average transient time is performed for $\tau_{ob}=2,3,4$. The average optimal time for each of the τ_{ob} is determined using the criterion of maximum fr^{10}. The average optimal time, $\overline{\tau_{om}}$ is shown in figure 3-7(a) as a function of τ_{ob}. The red curve gives $\overline{\tau_{om}}$ for $\Delta\mu_e=-0.08$ and the dotted blue curve gives $\overline{\tau_{om}}$ for $\Delta\mu_e=-0.1$. The average optimal time is independent of the μ_e, for small μ_e. The average optimal time is either 3 or 4 steps. Panel (b) shows fr^{10} as a function of τ_{ob}. fr^{10} is about 40% to 80%. For longer τ_{ob}, one should be careful in the interpretation of fr^{10} because for longer time, the

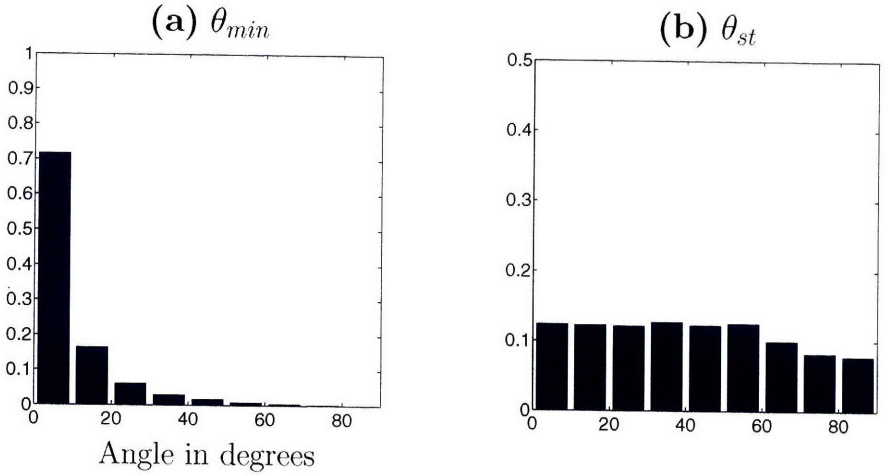


Figure 3-6: **(a)** shows the distribution of the minimum angle among θ_1, θ_2, θ_3 and θ_4, for $\tau_{ob}=1$. **(b)** shows the distribution of the angles made by $\boldsymbol{u}_{st}$ with $\boldsymbol{u}_{mm}$ for $\tau_{ob}=1$.

parametric vector is in the post-optimal phase and therefore the parametric vector tends to point in the direction of the local attractor.

The results presented in figure 3-5 suggest that if a covariance matrix is constructed using $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ as the leading eigendirection, then about 45% of the time it will perform as well as $\mathbf{Q}_{mm}$.

The next section describes the experiments performed to intercompare $\mathbf{Q}_{st}$, $\mathbf{Q}_{pv}$ and $\mathbf{Q}_{mm}$ and discusses the results.

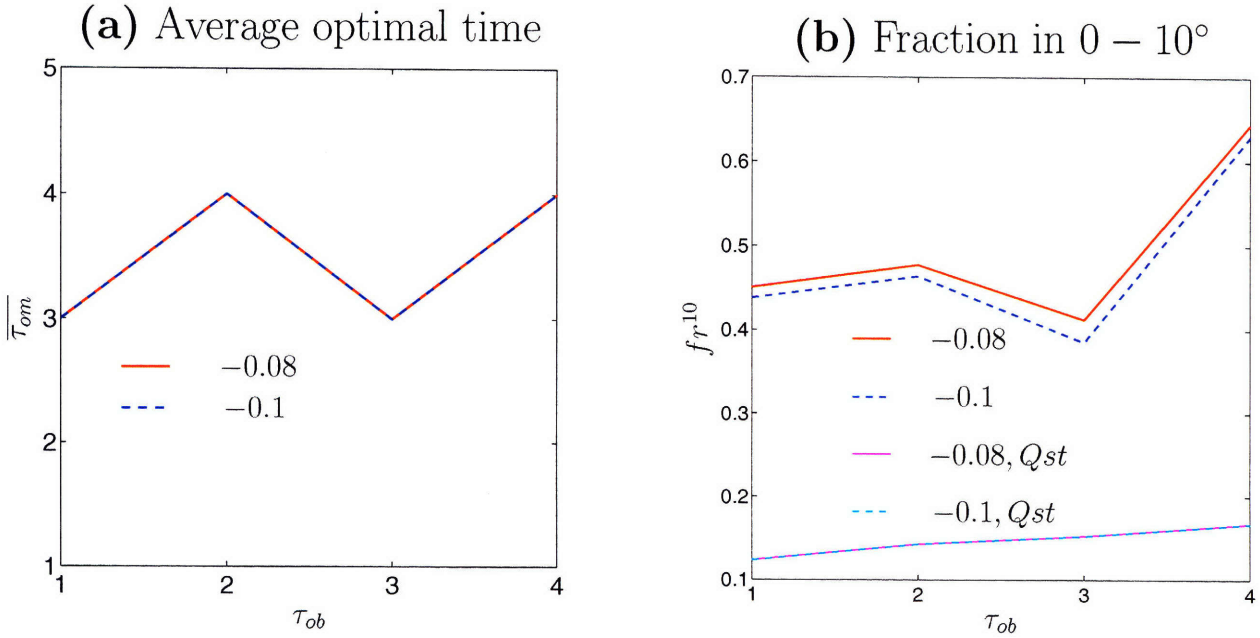


Figure 3-7: **(a)** shows the average optimal time as a function of τ_{ob} for two different error levels, $\Delta\mu_e=-0.08$ and $\Delta\mu_e=-0.1$. **(b)** shows fr^{10} corresponding to **(a)**.

3.7 Comparison and results

The performances of SQM, MMM and PVM are compared within the framework of the modified additive error approximation (MAEA). The approximation inherent in MAEA is equally applicable to the $\mathbf{Q}_{st}$, $\mathbf{Q}_{mm}$ and $\mathbf{Q}_{pv}$.

There are five EnKFs run in parallel with $\tau_{ob} = 1$. The ensemble size is $N = 100$. The first EnKF employs the perfect Ikeda model (μ=0.9). This filter is referred to as the perfect filter. The remaining four EnKFs use the Ikeda model with an imperfect μ_e where $\Delta\mu_e = -0.08$. The observations are drawn from the truth defined by the perfect model. Each of the four EnKFs using the imperfect model employs a different method to account for the model error, namely CIM (section 3.3), the SQM (section 3.4), the MMM (section 3.5) and the PVM (section 3.6). The four filters using the imperfect model use the following four different equations to replace equation (3.11c) (page 56) in the EnKF equations of MAEA.

$$\mathbf{P}^f(\alpha) = \alpha\mathbf{P}^f(\alpha) \qquad \Longrightarrow \alpha \text{ filter} \tag{3.21a}$$

$$\mathbf{P}^f(mm) = \mathbf{P}^f(mm) + \beta\mathbf{Q}_{mm} \qquad \Longrightarrow \mathbf{Q}_{mm} \text{ filter} \tag{3.21b}$$

$$\mathbf{P}^f(st) = \mathbf{P}^f(st) + \beta\mathbf{Q}_{st} \qquad \Longrightarrow \mathbf{Q}_{st} \text{ filter} \tag{3.21c}$$

$$\mathbf{P}^f(pv) = \mathbf{P}^f(pv) + \beta\mathbf{Q}_{pv} \qquad \Longrightarrow \mathbf{Q}_{pv} \text{ filter} \tag{3.21d}$$

These four filters are referred to as the imperfect filters. Among these four filters, the three filters employing $\mathbf{Q}$ are referred to as the $\mathbf{Q}$ filters. These four filters are in addition to the five MME filters run to obtain $\mathbf{Q}_{mm}$. In the current work, the boost factor used in equation (3.21a) is set equal to α=1. This means that the α filter gives the results for the imperfect model in the case when model error is not accounted for. In equations (3.21b), (3.21c) and (3.21d) β is a factor used to boost the size of $\mathbf{Q}$. β could be more or less than 1.

The MMM is used to obtain $\mathbf{Q}_{mm}$ at every DA step. An offline experiment is run to find β such that average analysis error obtained using $\mathbf{Q}_{mm}$ is minimum. The value of β is state-independent. It depends on τ_{ob} and μ_e.

At a given DA step, $\mathbf{Q}_{pv}$ and $\mathbf{Q}_{st}$ are constructed as follows. The magnitude of $\mathbf{Q}_{mm}$, that is $\|\mathbf{Q}_{mm}\|$, is calculated. The covariance matrix, $\mathbf{Q}_{pv}$, is constructed that has the leading direction given by $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ and magnitude given by $\|\mathbf{Q}_{mm}\|$. In the Ikeda system, the second eigendirection is easy to obtain because the system is 2-dimensional. Similarly $\mathbf{Q}_{st}$ is constructed so that it has a direction given by $\boldsymbol{u}_{st}$ and a magnitude given by $\|\mathbf{Q}_{mm}\|$. The boost factor β for $\mathbf{Q}$, used in equation (3.21c) and (3.21d), is the same as that used in equa-

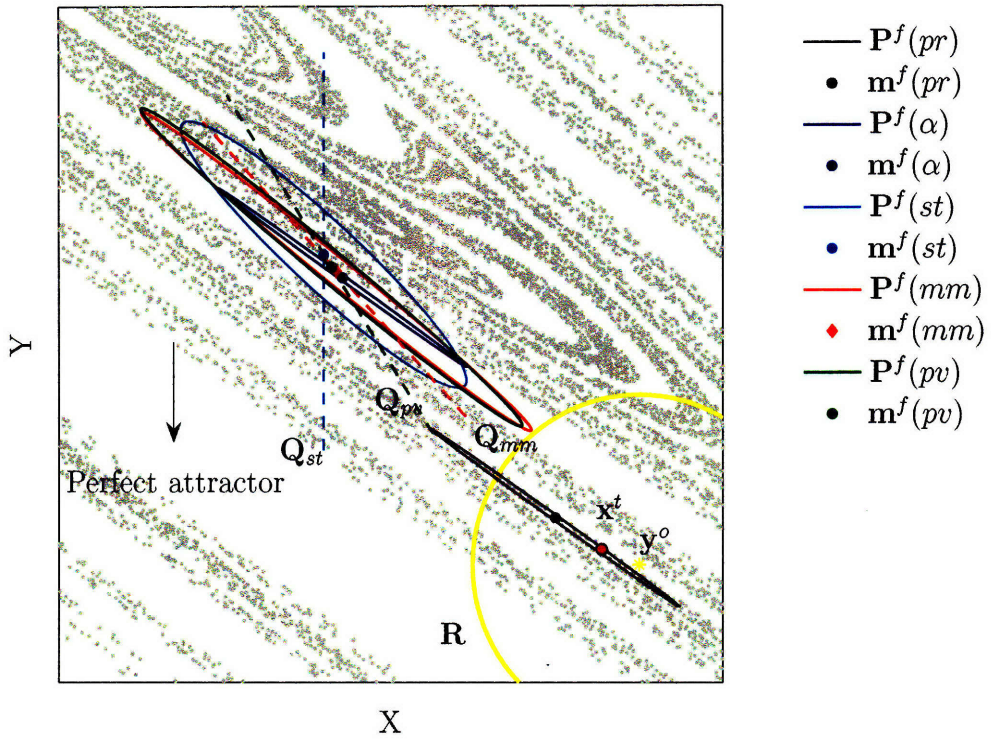
tion (3.21b). To summarize, at a given time, $\mathbf{Q}_{mm}$, $\mathbf{Q}_{pv}$ and $\mathbf{Q}_{st}$ are different in direction but they have the same *magnitude*, that is, $\|\mathbf{Q}_{pv}\| = \|\mathbf{Q}_{st}\| = \|\mathbf{Q}_{mm}\|$.

A scenario that compares the five different filters at a particular DA cycle is shown in figure 3-8. Panel (a) shows the comparison between the five filters of a forecast scenario. The gray dots show the local perfect attractor structure at the forecast location. The solid red dot in the lower right corner, marked $\mathbf{x}^t$ shows the truth. The solid yellow dot marked $\mathbf{y}^o$ shows the mean of the observations. The solid yellow dot lies at the center of the yellow circle that is the observation uncertainty $\mathbf{R}$. The black ellipse shows the $\mathbf{P}^f$ for the perfect filter. The black dot at the center of the black ellipse is the forecast mean for the perfect filter. The forecast error (fe), expressed in units of % attractor size for the perfect model, is 0.33. The blue ellipse shows the $\mathbf{P}^f(\alpha)$ and the blue dot at its center is $\mathbf{m}^f(\alpha)$. The fe for α filter is 2.23. The cyan dot gives the $\mathbf{m}^f(\text{st})$ and the cyan ellipse the corresponding $\mathbf{P}^f(\text{st})$. The green dot and green ellipse give $\mathbf{m}^f(\text{pv})$ and $\mathbf{P}^f(\text{pv})$, respectively. The red diamond between green and blue dots shows $\mathbf{m}^f(\text{mm})$. The red ellipse shows the $\mathbf{P}^f(\text{mm})$. The forecast errors for the imperfect filters are $fe(\alpha) = 2.23$, $fe(mm) = 2.28$, $fe(st) = 2.40$ and $fe(pv) = 2.32$. The forecast errors for the imperfect filters are approximately equal. The $\mathbf{P}^f$ corresponding to each of the four imperfect filters is "laterally" displaced from the perfect $\mathbf{P}^f$ shown by the black ellipse. This is because black ellipse tends to lie on the perfect attractor shown by the gray dots. The imperfect $\mathbf{P}^f$ tends to lie on the model attractor corresponding to $\mu_e = -0.08$.

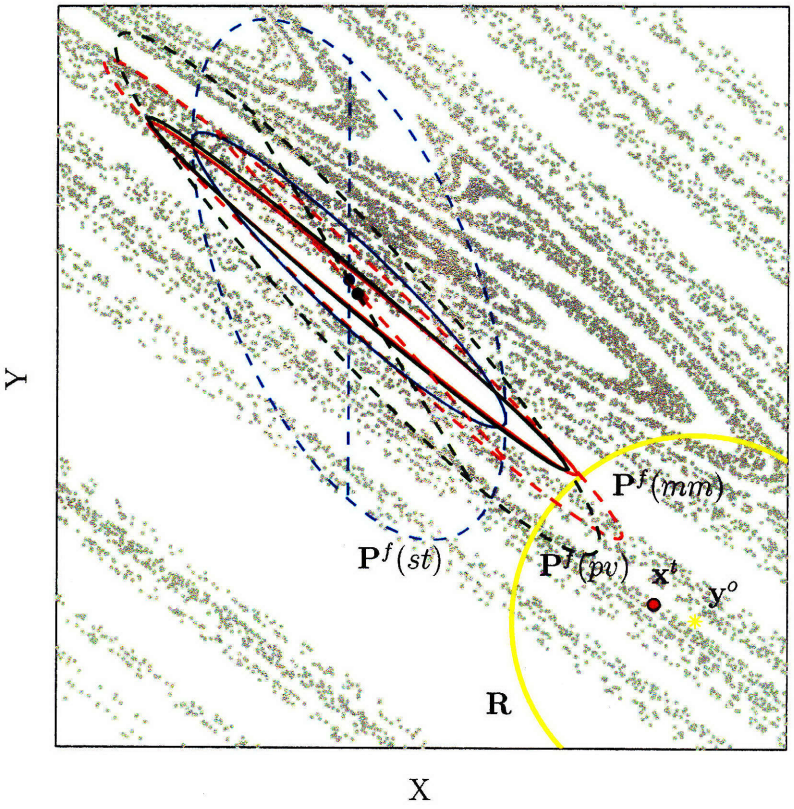
Note that the $\mathbf{P}^f(\alpha)$ (blue ellipse) is underdispersive compared to the $\mathbf{P}^f$ of the $\mathbf{Q}$ filters (red, cyan and green ellipses). Also note the minute difference between the orientation of the blue and the red ellipse in relation to the black ellipse. The blue ellipse is almost parallel to the black ellipse — the blue ellipse lies on the (imperfect) model attractor and the black ellipse on the perfect attractor. The red ellipse is not as parallel to the black ellipse as the blue ellipse is. Though the red ellipse tends to lie on the model attractor, it tends to "reach across" from the model attractor to the perfect attractor. This is because of the addition of $\mathbf{Q}_{mm}$ in the past DA cycles. Therefore, $\mathbf{P}^f(mm)$ tends to point towards truth more correctly than $\mathbf{P}^f(\alpha)$. For the same reason, $\mathbf{P}^f(\text{st})$ and $\mathbf{P}^f(\text{pv})$ tend to point towards the truth.

The dashed red, cyan and green lines show the $\mathbf{Q}_{mm}$, $\mathbf{Q}_{st}$ and $\mathbf{Q}_{pv}$ respectively (the multiplication with β is included in $\mathbf{Q}$). Though these look like lines, they are ellipses with the second eigen direction very small compared to the first eigendirection. $\mathbf{Q}_{mm}$, $\mathbf{Q}_{st}$ and $\mathbf{Q}_{pv}$ have the same magnitudes. Note that $\mathbf{Q}_{mm}$ is "correctly" pointing towards the truth (red circle). Because the leading eigendirection of $\mathbf{Q}_{pv}$ is $\boldsymbol{u}_1^{\mu\mathbf{x}}(3)$ by construction, on average $\mathbf{Q}_{pv}$ tends to align with $\mathbf{Q}_{mm}$. This is clearly seen in this case. The green dashed line tends to align with the red dashed line (the angle between them is 8.6°). Note that $\mathbf{Q}_{st}$ does not point towards the truth as compared to $\mathbf{Q}_{mm}$.

(a) Forecast, $\mathbf{P}^f$



(b) $\mathbf{P}^f + \mathbf{Q}$



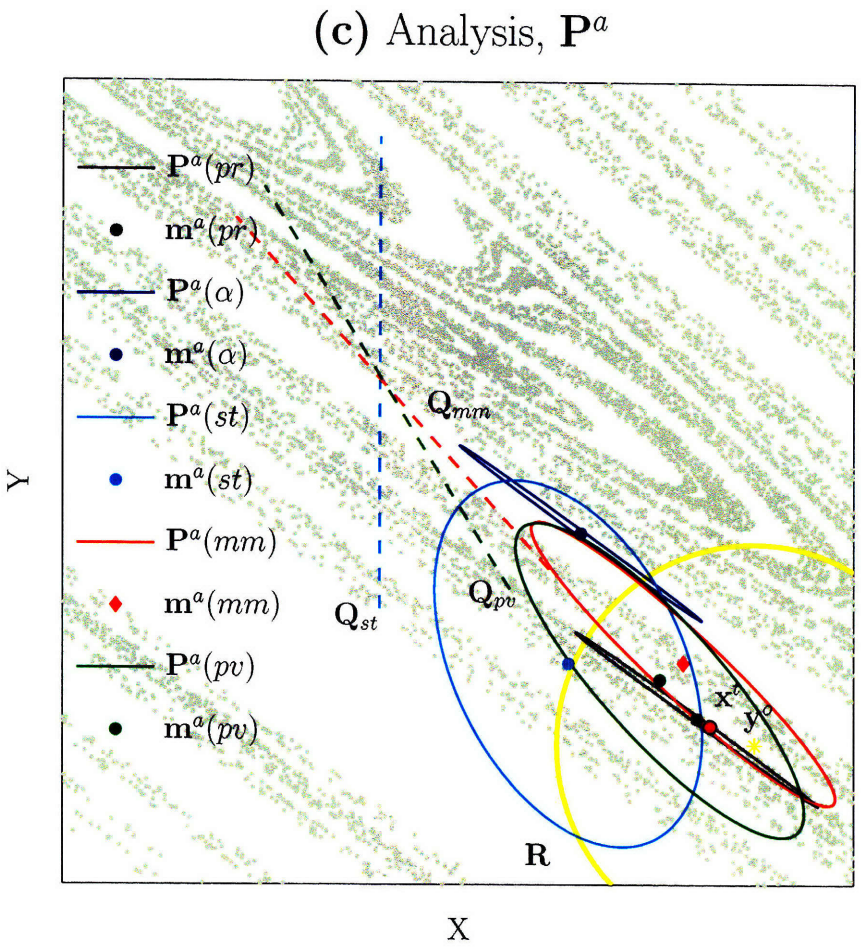


Figure 3-8: Forecast and analysis scenario of perfect and imperfect filters ($\Delta\mu_e = -0.08$) at a particular DA cycle in Ikeda model. The gray dots show the perfect attractor. The solid red circle in the lower right corner shows truth ($\mathbf{x}^t$). The yellow circle marked $\mathbf{R}$ shows the observation uncertainty. The solid yellow circle marked $\mathbf{y}^o$ is the mean observation.
(a) shows the forecast scenario. The black ellipse shows the $\mathbf{P}^f$ for perfect filter. The dotted lines show the $\mathbf{Q}_{mm}$, $\mathbf{Q}_{st}$ and $\mathbf{Q}_{pv}$.
(b) shows the forecast scenario after the addition of $\mathbf{Q}$.
(c) shows the analysis scenario. $\mathbf{Q}_{mm}$, $\mathbf{Q}_{st}$ and $\mathbf{Q}_{pv}$ are shown by the dotted lines which are the same as in (a) and (b). The observations error is 0.24% of attractor size.
The forecast error (fe) and the analysis error(ae) expressed in % attractor size for the five filters are:

	pr	α	$\mathbf{Q}_{mm}$	$\mathbf{Q}_{st}$	$\mathbf{Q}_{pv}$
fe	0.33	2.23	2.28	2.40	2.32
ae	0.07	1.16	0.34	0.78	0.34

The scenario shown in figure 3-8(a) is the forecast step for this particular DA cycle. Consider the $\mathbf{Q}_{mm}$ filter in figure 3-8(a). The red ellipse gives the sensitivity to the uncertainty in the initial condition and the red dashed ellipse gives the uncertainty due to error in μ. The next step in the DA cycle is the implementation of MAEA as discussed in section 3.2. Though $\mathbf{Q}_{mm}$ points "correctly" in the direction of truth, the important question is, what is the directionality and magnitude of $(\mathbf{P}^f + \mathbf{Q}_{mm})$?

The scenario after the implementation of MAEA is shown in figure 3-8(b). Only the $\mathbf{Q}_{mm}$, $\mathbf{Q}_{st}$ and $\mathbf{Q}_{pv}$ filters are shown. The solid ellipses and the $\mathbf{Q}$ given by the dashed lines from panel (a) are reproduced in panel (b). The dashed red ellipse, marked $\mathbf{P}^f(mm)$ shows $\mathbf{P}^f(mm) + \mathbf{Q}_{mm}$. Note that $\mathbf{P}^f(mm) + \mathbf{Q}_{mm}$ points towards truth more correctly than $\mathbf{P}^f(mm)$, that is, the addition of $\mathbf{Q}_{mm}$ has improved $\mathbf{P}^f(mm)$. Note that the green dashed ellipse, that gives $\mathbf{P}^f(pv) + \mathbf{Q}_{pv}$, is almost coincident with the red dashed ellipse. The cyan dashed ellipse gives $(\mathbf{P}^f(st) + \mathbf{Q}_{st})$.

The next step in the DA cycle is the analysis. The analysis scenario for the same DA cycle is shown in 3-8 (page 78)(c). The dashed lines which give the $\mathbf{Q}$ are reproduced from panel(b) for reference. Note that the black dot which gives the mean analysis for the perfect model lies very close to the truth. The ae for the perfect filter is 0.07. The red diamond which give $\mathbf{m}^a(mm)$ (ae=0.34) lies the closest to truth compared to the mean analysis of other imperfect filters. The $\mathbf{m}^a(pv)$ (ae=0.34) lies very close to $\mathbf{m}^a(mm)$.

The largest ae is for the α filter, the blue dot (ae=1.16). The ae for the $\mathbf{Q}_{st}$ filter is 0.78.

For this particular case, it seems that MAEA gives a good result, compared to not accounting for model error (α filter). The analysis error has improved by using either $\mathbf{Q}_{mm}$ $\mathbf{Q}_{st}$ or $\mathbf{Q}_{pv}$ method and then using the MAEA the approach.

Of course, this is a particular DA cycle. The fe and ae given correspond to this particular DA cycle. The comparative performance of these four filters can be assessed correctly by considering the average ae (and fe).

The ae is averaged for each of the filters over 14000 DA cycles for $\Delta\mu_e = -0.08$. The experiment is performed for $\tau_{ob} = 1, 2, 3, 4$. For the Ikeda system, two steps is about one error doubling time. Therefore, $\tau_{ob} = 1$ corresponds to about 50% of the error doubling time in the Ikeda system.

The average ae for the four imperfect filters and that for the perfect filter are shown in figure 3-9 (page 80)(a). The black curve, which is the lowest, gives the average ae for the perfect filter. The blue curve gives the error for the α filter. Among the filters that use $\mathbf{Q}$ to account for model error the filter using $\mathbf{Q}_{mm}$ performs the best (red curve), followed by the $\mathbf{Q}_{pv}$ filter

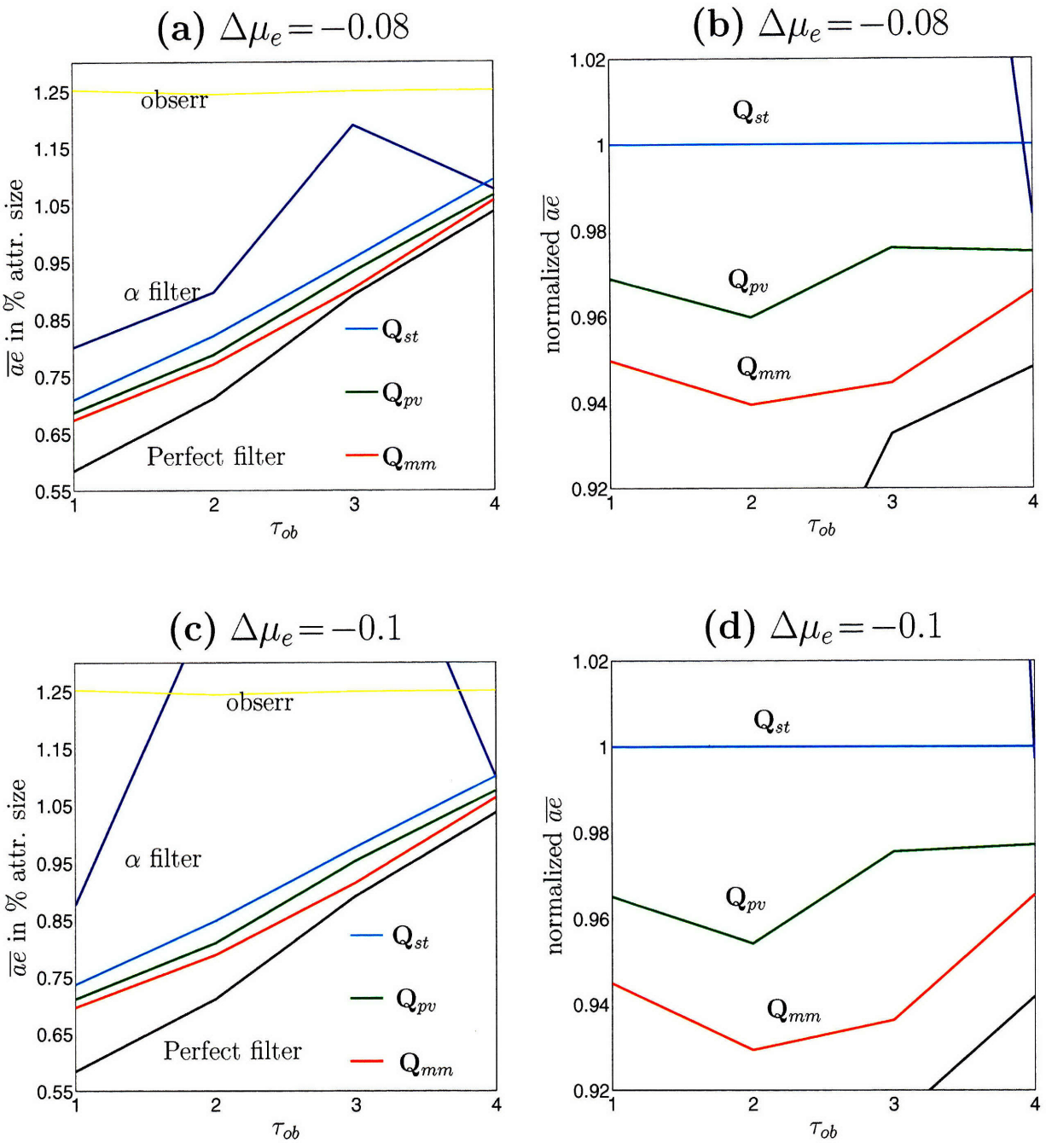


Figure 3-9: Comparison between the average analysis error of the five filters for the Ikeda system. The X-axis in all the panels is τ_{ob}. Ensemble size, $N = 100$.

(a) $\overline{ae}$ in % attractor size for $\Delta\mu_e = -0.08$.

(b) $\overline{ae}$ in (a) normalized by the $\overline{ae}(st)$ in (a). Values for α filter and perfect filter are out of scale for some τ_{ob}.

(c) $\overline{ae}$ in % attractor size for $\Delta\mu_e = -0.1$.

(d) $\overline{ae}$ in (c) normalized by the $\overline{ae}(st)$ in (c). Values for α filter and perfect filter are out of scale for some τ_{ob}.

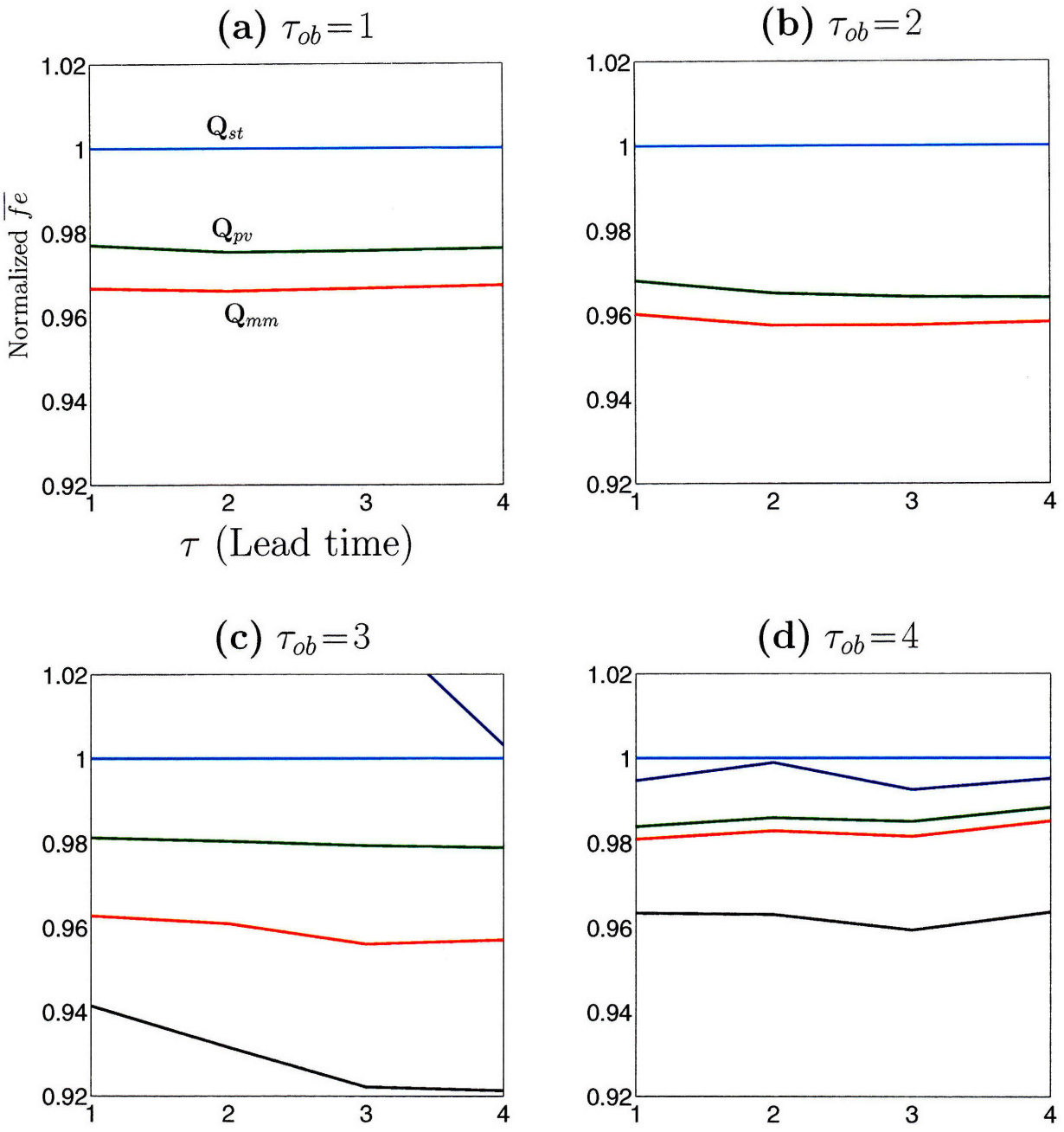


Figure 3-10: Comparison between the normalized average forecast error of the five filters for the Ikeda system for $\Delta\mu_e = -0.08$. The $\overline{fe}$ for each filter is normalized by the corresponding $\overline{fe}(st)$. The blue and black curves are for the imperfect and perfect filters, respectively. For some τ_{ob}, the blue and black curves are out of scale. Ensemble size, $N = 100$.

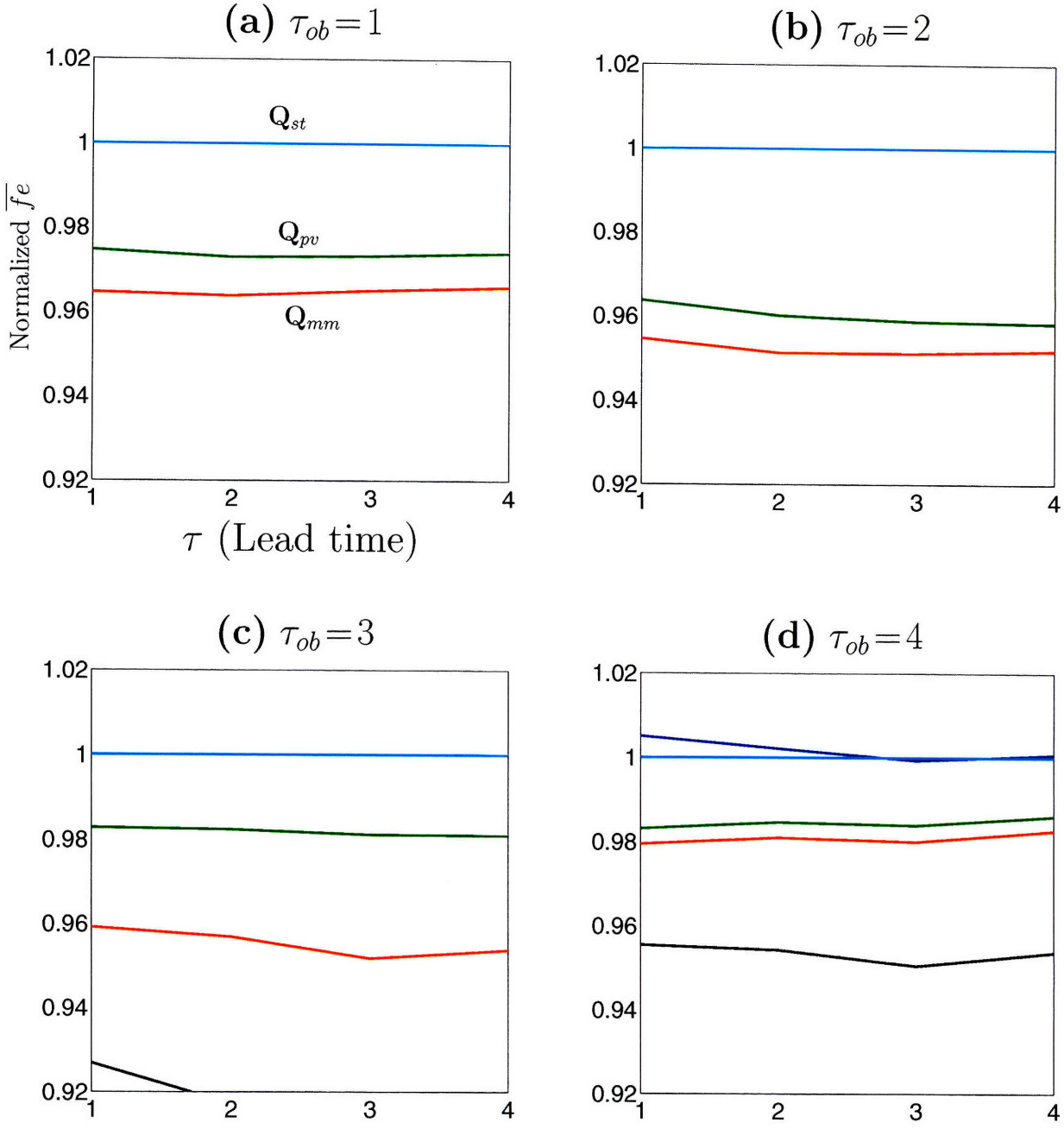


Figure 3-11: Comparison between the normalized average forecast error of the five filters for the Ikeda system for $\Delta\mu_e = -0.1$. The $\overline{fe}$ for each filter is normalized by the corresponding $\overline{fe}(st)$. The blue and black curves are for the imperfect and perfect filters, respectively. For some τ_{ob}, the blue and black curves are out of scale. Ensemble size, $N=100$.

(green curve). The $\mathbf{Q}_{st}$ filter performs the worst among the $\mathbf{Q}$ filters. In general, the $\mathbf{Q}$ filters improve the $\overline{ae}$ compared to the α filter (except for $\tau_{ob}=4$, for which $\mathbf{Q}_{st}$ gives a higher error compared to the α filter).

$\mathbf{Q}_{mm}$ performs the best because it provides state-dependent information about the model error. $\mathbf{Q}_{st}$ provides information about the model error but it is state-independent. $\mathbf{Q}_{pv}$ performs better than $\mathbf{Q}_{st}$ but worse than $\mathbf{Q}_{mm}$ because $\mathbf{Q}_{pv}$ is state-dependent but it is not correct all the time. $\mathbf{Q}_{pv}$ is correct only at those states at which the optimal time of the lagged parametric singular vectors is equal to the average optimal time for a given τ_{ob}.

Figure 3-9(b) shows the curves in figure 3-9(a) normalized by the $\mathbf{Q}_{st}$ curve in figure 3-9(a). $\mathbf{Q}_{mm}$ performs better than $\mathbf{Q}_{st}$ by about 5% while the $\mathbf{Q}_{pv}$ performs better than $\mathbf{Q}_{st}$ by about 3%. Figure 3-9(c) shows the $\overline{ae}$ for $\Delta\mu_e = -0.1$. Figure 3-9(d) shows the normalized $\overline{ae}$ corresponding to Figure 3-9(c). The results are similar to those for $\Delta\mu_e=-0.08$.

The normalized $\overline{fe}$ for $\Delta\mu_e = -0.08$ is shown in figure 3-10. The different panels are for different τ_{ob}. For a given τ_{ob}, the $\overline{fe}$ for each filter is normalized by the $\overline{fe}$ for the $\mathbf{Q}_{st}$ filter for each lead time τ. The blue and black curves give the normalized $\overline{fe}$ for imperfect and perfect filters, respectively. These curves are out of scale for some τ and τ_{ob}. Among the filters using $\mathbf{Q}$, the filter using $\mathbf{Q}_{mm}$ performs best. The $\mathbf{Q}_{pv}$ filter approximates the forecast errors of $\mathbf{Q}_{mm}$ filter well. The normalized $\overline{fe}$ for $\Delta\mu_e=-0.1$ is shown in figure 3-11. The results are similar to those for $\Delta\mu_e=-0.08$.

The average forecast and analysis errors presented for the Ikeda model suggest that the MAEA is an approach that could be used to account for model error in the EnKF. The results also support the reasoning that the correct direction of model error is the sensitivity of the local attractor to the parameter. This is the reason why the $\mathbf{Q}_{mm}$ performs the best.

We do not know whether these results will scale to higher dimensions. I have performed experiments with the Lorenz 1963 system to test MAEA and various methods to obtain $\mathbf{Q}$. In the next section 3.7.1, the results for the comparison between $\mathbf{Q}_{mm}$, $\mathbf{Q}_{st}$ and $\mathbf{Q}_{pv}$ for the Lorenz 1963 system (hereafter L63) within the framework of MAEA are presented.

3.7.1 Results for the Lorenz 1963 system

The equations of L63 are given by,

$$\frac{dx}{dt} = \sigma^{l63}(y - x) \tag{3.22a}$$

$$\frac{dy}{dt} = rx - y - xz \tag{3.22b}$$

$$\frac{dz}{dt} = xy - bz \tag{3.22c}$$

where the parameters in L63 are $\sigma^{l63} = 10$, $b = \frac{8}{3}$, $r = 28$. The parametric model error is introduced in L63 system by changing the value of σ^{l63}. The attractor of the L63 system is shown in figure 3-12.

The experiment done with L63 is similar to the experiment done with the Ikeda model explained in section 3.7. The state-dependent $\mathbf{Q}_{mm}$ is obtained by running a MME. I have chosen the $\sigma_j^{l63\prime}$ as $\sigma_j^{l63\prime} = \sigma^{l63} + \Delta\sigma_j^{l63}$ where $\Delta\sigma_j^{l63} = -0.01, -0.001, 0, 0.001, 0.01$. The results for the intercomparison between the five filters are presented in figure 3-13. The panels going from (a) to (c) show the normalized $\overline{ae}$ for increasing level of model error.

For $\Delta\sigma_e^{l63} = -0.03$ shown in panel (a), $\mathbf{Q}_{mm}$ and $\mathbf{Q}_{pv}$ clearly outperform $\mathbf{Q}_{st}$. The normalized $\overline{ae}$ for $\mathbf{Q}_{pv}$ closely shadows that from $\mathbf{Q}_{mm}$. The advantage of using state-dependent $\mathbf{Q}$ over state-independent $\mathbf{Q}$ ($\mathbf{Q}_{st}$) is more for shorter τ_{ob}. The $\mathbf{Q}_{mm}$ and $\mathbf{Q}_{pv}$ curves increases almost monotonically with τ_{ob}.

Similar trends are shown for higher levels of model error in panel (b) and (c), except that the $\mathbf{Q}_{mm}$ and $\mathbf{Q}_{pv}$ do not increase monotonically. The red and green curves decrease for $5 \geq \tau_{ob} \leq 15$ and then increase. In light of the average analysis errors obtained for the Ikeda and L63, the mechanism of MMM and PVM is considered again. The success of the MMM

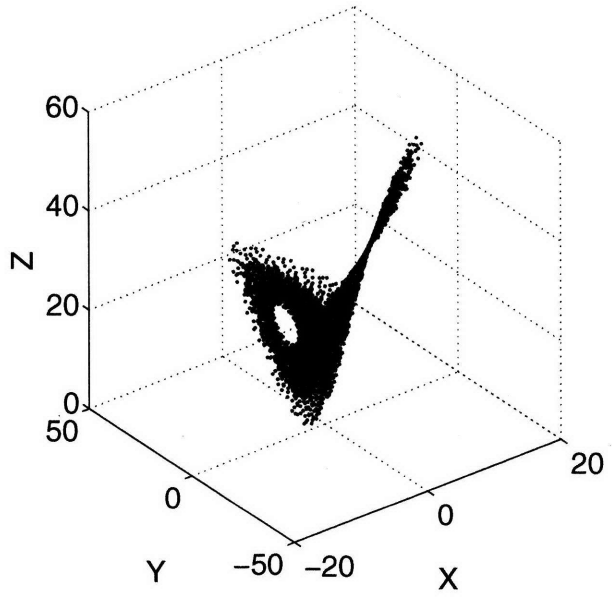


Figure 3-12: The celebrated L63 attractor.

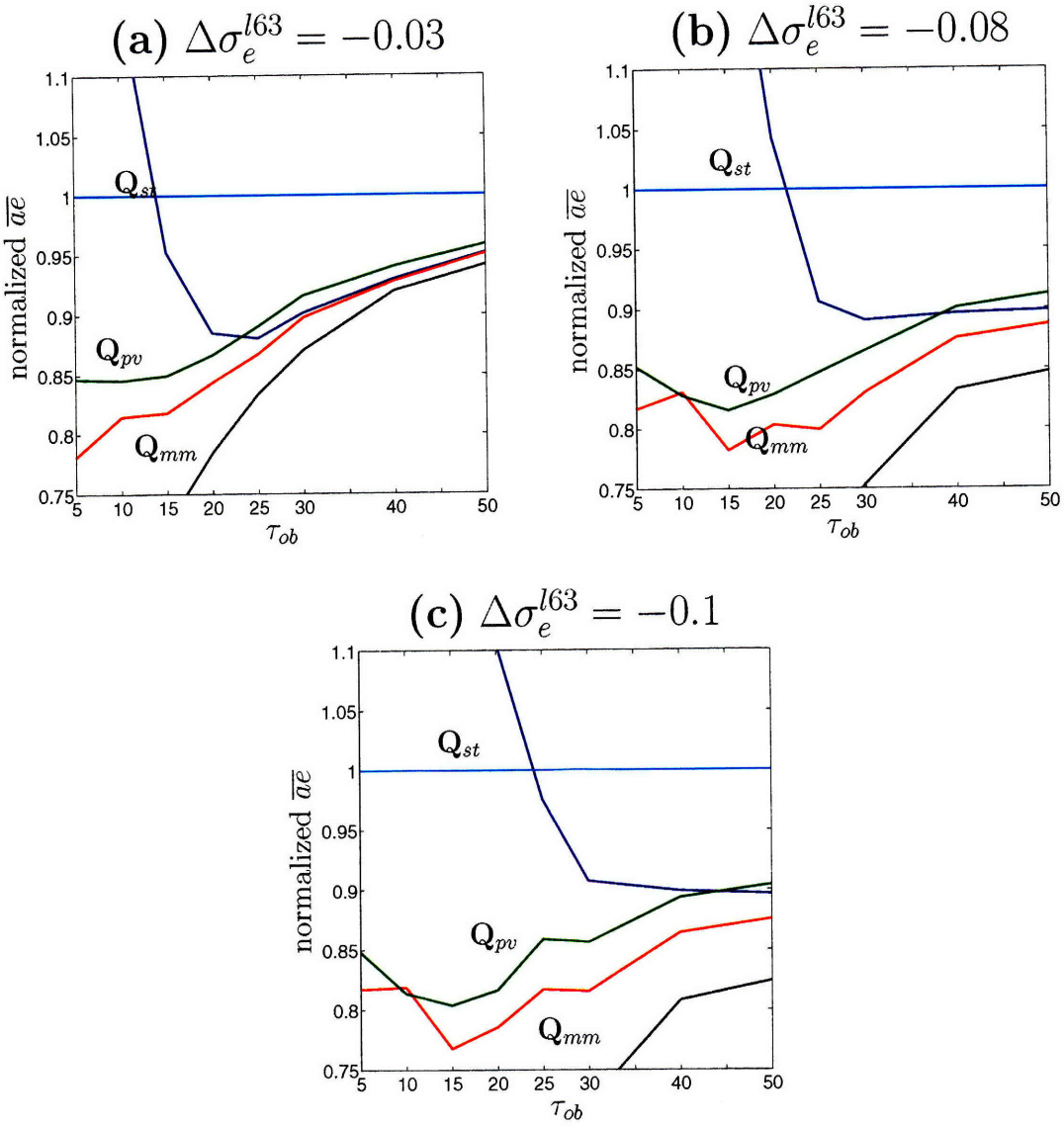


Figure 3-13: Comparison of average analysis errors between the five filters for the L63 system. The X-axis in all the panels is τ_{ob} expressed as % of the error doubling time for L63. The Y-axis gives the $\overline{ae}$ for the different filters normalized by the $\overline{ae}$ for the $\mathbf{Q}_{st}$ filter. The blue and black curves are for the imperfect and perfect filters, respectively. These curves are out of scale for some values of τ_{ob}. The ensemble size is $N = 100$.

in giving the correct model error direction depends upon each of the MME falling onto their own respective attractors. Assuming that the analysis MME are off their attractors, then the forecast MME may not fall onto their own attractors if τ_{ob} is short. The success of the MMM is at the mercy of the state-dependent transient time compared to the τ_{ob}. On the other hand, the success of the PVM depends on the optimal time, which is broadly speaking the transient time. To be more precise, the optimal time (τ_{om}) is the *parametric transient time*. If the sensitivity of the state to initial condition error is sought, then the transient time is the relevant time scale, but if the sensitivity to parametric model error is sought then the optimal time is the relevant time scale.

The next section (3.8) presents the conclusions of this chapter.

3.8 Conclusions

The EnKF is an important tool for the prediction of geophysical systems. In the past decades there has been a number of scientific papers dealing with the science and practice of EnKF. Geophysical models are imperfect. The amount of work done to deal with the model error problem is less compared to that done to deal with the initial condition error. Model error imperfection is a big impediment to accurate analysis and forecasts using the EnKF.

In this chapter, the issue of accounting for model error in the EnKF is addressed. An approach, that of additive model error (AEA), is described and explored. The AEA requires an estimate of the model error. A novel method, namely the PVM is introduced and described. The MMM and the SQM are compared with the PVM method for various parametric model errors levels in the Ikeda and the L63 system.

The correct model error direction at a particular state in the phase space of the model is the direction of displacement of the local attractor in response to model error. If the model error is due to parametric model error (PME) then the correct model error direction is the sensitivity of the local attractor structure to the parameter.

The MMM estimates the state-dependent sensitivity of the local attractor structure to the parameter. MMM outperforms SQM in giving lower average analysis and forecast errors. This is because SQM gives a state-independent estimate of model error. The PVM is based on the concept of singular vectors. The PVM estimates the linear sensitivity of the attractor to the parameter by linearizing around the analysis mean. The correctness of the sensitivity obtained using PVM depends on the parametric transient time (τ_{tp}), also called the optimal time (τ_{om}). The average τ_{om} is calculated using information from MMM. $\mathbf{Q}_{pv}$ is constructed by using the average τ_{om} as the optimization time. If the τ_{op} used to construct the parametric vector is shorter than the τ_{om} for that state, that is $\tau_{op} < \tau_{om}$, then the parametric vector gives the transient sensitivity. This sensitivity is that of an off-attractor state to the perturbation in the parameter. If $\tau_{op} < \tau_{om}$ then the parametric vector is said to be *pre-optimal*. The correct model error direction is given by the parametric vector only if $\tau_{op} = \tau_{om}$. The off-attractor point gets onto the attractor at time τ_{om} and therefore the parametric vector gives the correct model error direction. This parametric vector is said to be *optimal*. If $\tau_{op} > \tau_{om}$, then the parametric vector is said to be *post-optimal*.

To understand the post-optimal phase it is important to understand the concept of *model-induced state error*. The parametric vector gives the sensitivity of the state when it is stepped forward using two different values of the parameter. As soon as the state is stepped forward with two different values of the parameter, an error is induced in the state that evolves in time. This is called the model-induced state error. The total model error signal can be decomposed into pure model error signal and model-induced state error signal. The pure model error signal undergoes a transient during the pre-optimal phase and maximizes with respect to the induced model error at the optimal time. At the optimal time the transient behavior dies out. In the post-optimal phase the total model error signal is dominated by the model-induced state error and therefore the total model error points in the direction of the local attractor. The parametric vector in the post-optimal phase gives information about the local attractor because it is dominated by the model-induced state error rather than the pure model error. The post-optimal parametric vector is not useful in giving information about the model error.

Therefore, the important vector is the optimal parametric vector. The pre-optimal vector gives the local, off-attractor sensitivity. The post-optimal vector gives the on attractor sensitivity but it is not local. The optimal parametric vector gives the sensitivity which is both on-attractor and local.

The filter that uses $\mathbf{Q}_{pv}$ performs better than that using $\mathbf{Q}_{st}$ in that $\mathbf{Q}_{pv}$ gives lower $\overline{ae}$ and $\overline{fe}$. $\mathbf{Q}_{pv}$ closely approximates the results obtained using $\mathbf{Q}_{mm}$. In short, the PVM is a very promising method to estimate the state-dependent model error direction. It can replace the MMM method if the criterion to determine the state-dependent optimal time (τ_{om}) is identified. The optimal time is the parametric transient time. The criterion to find τ_{om} could be formulated by first formulating the criterion to find the state-dependent transient time.

The results obtained suggest that the AEA approach can be used effectively to account for model error in the EnKF provided one has the access to the correct model error structure. The AEA, though not the correct approach as discussed in section 3.2 (page 52), is simple to implement and appears to be effective in accounting for model error. The model equations used in operational weather models need not be modified/recoded to implement AEA.

Chapter 4

Reduction of model error

Chapter 3 introduced, explored and discussed a technique to account for parametric model error in the EnKF. *Accounting for model error* is a technique to address the problem of model error. The other technique is the *reduction of model error*. This chapter discusses techniques to reduce model error. The reduction of model error involves using the data to alter parameters or structure of the imperfect model so that it simulates the reality in a better way. Ideally, one would like to recover the perfect model. This may not be always possible. In that case one seeks alterations in the imperfect model that would compensate for the model error.

Models usually have one or more parameters. Typically, geophysical models have multiple parameters. The values of these parameters in nature are called their *perfect* values. One usually has an estimate of the perfect value of these parameters. Given data, this estimate of the parameter values can be improved, that is, values closer to their perfect values can be retrieved. The procedure for estimating the perfect value of parameter(s) given the data is called *parameter estimation*. A general feature of all methods that perform parameter estimation is that the parameters in the models to be estimated are treated as variables. The state of the model is augmented with parameters and this state is called the augmented state (Derber, 1989; Wunsch, 1996). Typically, observations are only of state, that is, the parameters are not observed. In this work the observations are of state only.

The Kalman filter that is described by equations 1.1 (page 11) in section 1.1 performs state estimation. This Kalman filter methodology can be used to perform parameter estimation as well. Assuming that the model is linear and uncertainties are Gaussian, the Kalman filter is implemented with the augmented state. The data are assimilated into the augmented state. The covariances between state and parameter correct the parameter estimate towards its perfect value. Though the phrase parameter estimation is used, it should be noted that it is joint parameter-state estimation. The Kalman filter has not been used for parameter

estimation in atmospheric models for the same reasons that the Kalman filter has not been used for state estimation, namely, high dimensionality and nonlinearity. The advent of EnKF has not only opened the possibility of state estimation in high-dimensional models but also that of parameter estimation in these models.

EnKF has been used recently by several workers for parameter estimation for both low and high-dimensional models. The EnKF, apart from giving the estimate of the parameter, also gives the uncertainty in the estimate. Anderson (Anderson, 2001) has shown that a variant of EnKF is able to provide a satisfactory estimate of a parameter in the 40-dimensional L96 model. In Annan and Hargreaves (2004) the parameters in the highly chaotic L63 model are successfully estimated. Within the high-dimensional atmospheric models one distinguishes between weather models and climate models. EnKF has been used to estimate parameters in both these type of models. It has been suggested that imperfection in parameters is more significant than initial condition error in rendering climate forecasts wrong. In Annan et al. (2005a), the EnKF has been used for climatological estimation of 12 independent parameters in a non-chaotic but high-dimensional intermediate complexity earth system model. The estimates of these parameters and their distributions are found to be consistent with their perfect values used in the synthetic truth run. This work has been extended in Annan et al. (2005b) in which it has been shown that the EnKF can be used to successfully estimate parameters in a realistic atmospheric GCM. In Aksoy et al. (2006a,b) multiple parameters in weather models have been estimated using the EnKF.

The Kalman filter is an example of a sequential technique of data assimilation. Apart from the sequential technique of data assimilation, there exists another technique called the variational technique of data assimilation (Talagrand, 1997). The variational data assimilation is carried out with the adjoint model equations. The variational data assimilation technique can be used for parameter estimation as well. The variational technique for parameter estimation has been found to have some drawbacks compared to ensemble-based parameter estimation (Lea et al., 2000). See Lea et al. (2000) and Lea et al. (2002) for details. In these papers, Lea et al. have presented results for sensitivity analysis of model climate to parameters rather than for parameter estimation. But the core issues in sensitivity analysis with respect to parameters are the same as those in parameter estimation. It seems that the adjoint method is vulnerable to local minima in the cost function while, on the other hand, the ensemble-based method tends to smooth out the local minima (Annan and Hargreaves, 2004).

Similar to state estimation, parameter estimation is carried out in several different disciplines. The Green's functions approach has been used by Menemenlis et al. (2005) to estimate background vertical diffusivity, vertical viscosity and other parameters in an ocean GCM. Parameter estimation is also called calibration of models by some researchers.

The parameter estimation work done earlier, which has been reviewed above, in general assumes that the model is only parametrically imperfect. The parameter estimation therefore seeks to retrieve the perfect values of the imperfect parameters. The work presented in this chapter addresses the question of parameter estimation in the framework of ensemble-based methods for parametrically and structurally imperfect models. The effect and interpretation of parameter estimation in the presence of structural model error has not been considered before. In this work it is shown that parameter estimation can be successfully employed to offset structural model error.

4.1 Introduction

Model error is divided into two types, namely *parametric* model error and *structural* model error. Parametric model error (PME) means that apart from the imperfect or incorrect values of the parameters in the models, the model is perfect. Structural model error (SME) is more serious than PME. SME includes all errors in the model not due to imperfect parameters. In real atmospheric models, parametrization is a typical example of SME. The abbreviations PIM and SIM are used to denote parametrically imperfect models and structurally imperfect models, respectively.

In this chapter, a typical technique of model error reduction — namely parameter estimation or parameter tuning is applied to the Ikeda model and a real atmospheric model (NOGAPS). The motivation is to address the problem of SME in real atmospheric models. The technique of model error reduction is first applied to the Ikeda systems and an attempt is made to identify generic methodology that might also work in real atmospheric models. In short, lessons are learned from implementing model error reduction in the Ikeda system and are then used in the NOGAPS model.

Sections 4.2 and 4.3 describe and present the results for parameter estimation/tuning in the Ikeda system for PIM and SIM, respectively. Section 4.4 summarizes the results obtained from Ikeda model and identifies some generic features of reduction of model error that are applied to a real atmospheric model, namely the NOGAPS model. Section 4.5 discusses the NOGAPS model, the Emanuel parameterization scheme and the implementation of parameter tuning in NOGAPS. Section 4.6 presents the conclusions of this chapter.

4.2 Parameter estimation in the Ikeda model

The technique of parameter estimation basically involves treating the parameter(s) as variables and using the observations to correct the parameter value(s). This section illustrates and explores parameter estimation for the Ikeda model. Section 4.2.1 explains the concept of parameter estimation and explores parameter estimation of a single parameter in the Ikeda model. Section 4.2.2 explores the simultaneous estimation of multiple parameters in the Ikeda model.

4.2.1 Estimation of single parameter

In this, section the technique of parameter estimation for PIM is laid out in detail. For simplicity it is assumed that only one parameter, is imperfect.

Consider the Ikeda system given by

$$x_{s+1} = 1 + \mu(x_s \cos\theta - y_s \sin\theta) \tag{4.1a}$$

$$y_{s+1} = \mu(x_s \sin\theta + y_s \cos\theta) \tag{4.1b}$$

where

$$\theta = a - \frac{b}{(x_s^2 + y_s^2 + 1)}$$

$a = 0.4$, $b = 6$, $\mu = 0.9$

To begin with the case when μ is imperfect is explained in detail. The parametric model error is introduced by making the parameter μ imperfect. The Ikeda is a 2-dimensional system, that is, n=2. The equations for stepping forward the state are given by equations (4.1a) and (4.1b). This 2-dimensional state is augmented by the parameter μ. The following equation is used for stepping forward the parameter μ.

$$\mu_{s+1} = \mu_s \tag{4.2}$$

Equation (4.2) is a deterministic step forward (DSF) equation because the rhs of this equation does not contain stochastic noise.

The specification of the filter used for the parameter estimation of μ, called the DSF filter, is as follows. The vector of parameters used to augment the state is denoted by $\boldsymbol{\alpha}$. The

dimension of the augmented state is denoted by n_a. In general, the state could be augmented by more than one parameters. In this particular case the state is augmented by only one parameter, namely μ. Therefore, $n_{\boldsymbol{\alpha}}=1$ and $\boldsymbol{\alpha}=\mu$. The dimension of the augmented state is $n_a=n+n_{\boldsymbol{\alpha}}=2+1=3$. Equations (4.1a) and (4.1b) along with (4.2) define the augmented Ikeda model. The EnKF is implemented using the augmented Ikeda model. The equations for the EnKF are described in chapter 2. These equations are reproduced below. For $i=1\ldots N$,

$$\mathbf{x}_i^f(t) = \mathbf{F}[\mathbf{x}_i^a(t-1)] \tag{4.3a}$$

$$\mathbf{P}^f(t) = \frac{1}{N-1}[\mathbf{E}^f(t) - \mathbf{M}^f(t)][\mathbf{E}^f(t) - \mathbf{M}^f(t)]^T \tag{4.3b}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T[\mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t)]^{-1} \tag{4.3c}$$

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f(t) + \mathbf{K}(t)[\mathbf{y}_i^o(t) - \mathbf{H}(t)\mathbf{x}_i^f(t)] \tag{4.3d}$$

$$\mathbf{P}^a(t) = \frac{1}{N-1}[\mathbf{E}^a(t) - \mathbf{M}^a(t)][\mathbf{E}^a(t) - \mathbf{M}^a(t)]^T \tag{4.3e}$$

In these equations, N is the number of ensemble members. For the work presented in this section (4.2) $N=100$. The operator $\mathbf{F}$ in equation (4.3a) contains the step forward equations for the augmented state — in this case, equations (4.1a), (4.1b) and (4.2). The ensembles $\mathbf{E}^f(t)$ and $\mathbf{E}^a(t)$ are $n_a \times N$. The first guess, $\mathbf{x}_i^f(t)$ and analysis $\mathbf{x}_i^a(t)$ are $n_a \times 1$. Though the state is augmented by the parameter μ, the observations are of state only, that is, the parameter μ is not observed. The number of dimensions that are observed is denoted by m. In this case $m=n$. Therefore each perturbed observation $\mathbf{y}_i^o(t)$ is $n \times 1$. The Kalman gain term, $\mathbf{K}(t)$ is $n \times m$. The matrices $\mathbf{P}^a(t)$ and $\mathbf{P}^f(t)$ are $n_a \times n_a$. $\mathbf{R}$ is $m \times m$. $\mathbf{H}(t)$ maps the model state to the observation space. $\mathbf{H}(t)$ is $m \times n_a$ and is given by,

$$\mathbf{H}(\mathrm{t}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{4.4}$$

In equation (4.4), the entries in the last column are 0 because the parameter μ is not observed.

At time $s=0$ an ensemble $\mathbf{E}$ of augmented states is constructed. The size of $\mathbf{E}$ is $n_a \times N$. Let $\mathbf{E}_{\boldsymbol{\alpha}}$ denote the ensemble in the parameter space. The state ensemble occupies the first two dimensions in $\mathbf{E}$ and $\mathbf{E}_{\boldsymbol{\alpha}}$ occupies the third dimension in $\mathbf{E}$. $\mathbf{E}_{\boldsymbol{\alpha}}$ is constructed as follows. The correct value of μ, that is, $\mu=0.9$ is displaced by 1% of $\mu=0.9$. This displaced value is taken as the mean of $\mathbf{E}_{\boldsymbol{\alpha}}$. Then the members of $\mathbf{E}_{\boldsymbol{\alpha}}$ are constructed by adding perturbations to the mean of $\mathbf{E}_{\boldsymbol{\alpha}}$. These perturbations are drawn from $N(0,\sigma_\mu)$. The standard deviation σ_μ is chosen to be $\sigma_\mu=1\%$ of $\mu=0.9$. (The symbol σ is used in this chapter to denote standard deviation.)

Equation (4.3a) is implemented as follows. The i^{th} state ensemble member is stepped forward with equations (4.1a) and (4.1b) using the i^{th} ensemble member of $\mathbf{E}_{\boldsymbol{\alpha}}$ as the value for μ. Then equations (4.3b) through (4.3e) are implemented, which completes one DA cycle. The data are assimilated every $s=1$ steps, that is, $\tau_{ob}=1$. This filter that uses the augmented state is called the DSF filter, because the parameter ensemble uses the deterministic step forward given by equation (4.2).

The effect of DA is to minimize the variance within $\mathbf{E}_{\boldsymbol{\alpha}}$. The mean of $\mathbf{E}_{\boldsymbol{\alpha}}$ is noted after every DA cycle. Let this analysis mean of $\mathbf{E}_{\boldsymbol{\alpha}}$ be denoted by $\hat{\mu}_d$. The subscript d in $\hat{\mu}_d$ stands for DSF. Figure 4-1(a) shows the value of $\hat{\mu}_d$ for the first 50 DA cycles in cyan. The black line shows the perfect value of $\mu = 0.9$. Note that the cyan curve converges towards the black curve. Though the parameter μ is not observed, the covariances between state and parameter act to correct the parameter. The variance within $\mathbf{E}_{\boldsymbol{\alpha}}$ decreases towards 0 as the DA cycles proceed. This variance becomes 0 at $s = 50$ and remains at 0 for $s > 50$. The value of $\hat{\mu}_d$ converges to ≈ 0.899. The value of $\hat{\mu}_d$ for all steps is shown in figure 4-1(b) in cyan color. Note that $\hat{\mu}_d$ converges towards the correct value of μ, that is $\mu=0.9$ but does not converge exactly to $\mu=0.9$. The PME is reduced but not eliminated completely. $\hat{\mu}_d$ does not converge to $\mu = 0.9$ because the variance in the parameter ensemble ($\mathbf{E}_{\boldsymbol{\alpha}}$) goes to 0 in about $s = 50$ steps and thereafter the observations are not able to correct the value of μ any further. This can be interpreted as filter divergence in the parameter space. All the ensemble members in $\mathbf{E}_{\boldsymbol{\alpha}}$ collapse to the same value, thus preventing any further correction of $\hat{\mu}_d$.

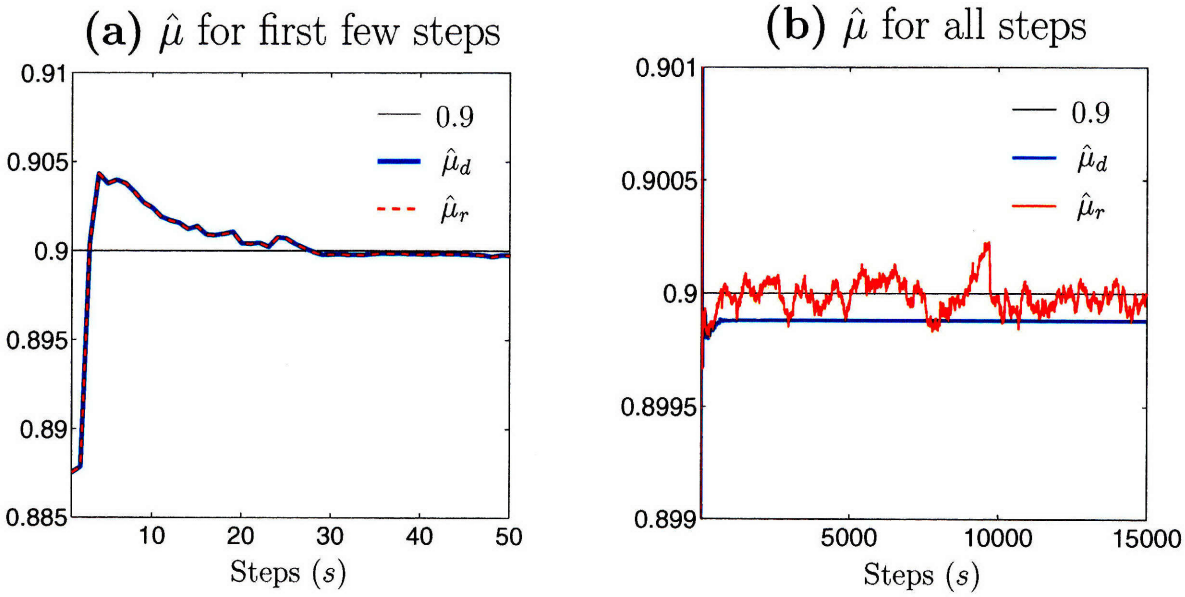


Figure 4-1: $\hat{\mu}$ is the estimate of μ given by the mean of analysis $\mathbf{E}_{\boldsymbol{\alpha}}$. The data are assimilated at every step, that is, $\tau_{ob} = 1$. Only one parmater, namely μ is imperfect. The ensemble size $N=100$.

To assess the performance of the reduction of model error, two other filters are run independently. These two filters use the perfect and imperfect models respectively. The perfect model uses $\mu=0.9$. The imperfect model uses the imperfect value of μ denoted by μ_e and is given by $\mu_e=\mu+\Delta\mu$, where $\mu=0.9$, $\Delta\mu=-0.08$. $\Delta\mu$ is in % of $\mu=0.9$. The same data are assimilated in these three filters. The perfect filter employs an ensemble that has dimension $n\times N$. For the perfect filter, the state is not augmented by the parameter μ. Similarly, for the imperfect filter, the state is not augmented by the parameter μ.

To summarize the setup of the experiment, there are three filters running in parallel, namely the DSF filter, the perfect filter and the imperfect filter. The DSF filter uses an ensemble with dimension $n_a\times N$, while the perfect and imperfect filter each use an ensemble with dimension $n\times N$. The same data are assimilated in all these three filters.

The average analysis and forecast errors denoted by $\overline{ae}$ and $\overline{fe}$ are calculated for each of these three filters. The $\overline{ae}$ and $\overline{fe}$ are expressed as a % of attractor size. The transient is removed for $s=1000$, that is, the first 1000 DA steps are not included in the calculation of the average errors. The $\overline{ae}$ for $\tau_{ob}=1$ is 0.584, 0.598 and 0.8 for the perfect, DSF and imperfect filter, respectively. The parameter estimation using DSF corrects the imperfect model and recovers approximately the perfect value of μ. This reduction of model error results in lower analysis errors for the DSF filter compared to the imperfect filter.

This experiment is carried out for different τ_{ob}, that is, the data are assimilated into the three filters at different time frequencies. The average analysis errors in % attractor size as a function of τ_{ob} is shown in figure 4-2(a). The black and the blue curves show the $\overline{ae}$ for the perfect and imperfect models, respectively. The cyan curve lies very close to the black curve and is much below the blue curve. The cyan curve approximates the black curve well for all values of τ_{ob}. The estimation of the parameter μ results in a reduction of model error, which in turn decreases the average analysis error. The advantage of the DSF over the imperfect model in reducing the model error is quantified by normalizing each of the curves by the imperfect (blue) curve. The result of this normalization is shown in figure 4-2(b). The reduction in model error decreases the $\overline{ae}$ by about 25% for $\tau_{ob}=1$. For $\tau_{ob}>1$, this decrease ranges from about 20% to about 1%. The cyan curve in panel figure 4-2(a) does not coincide exactly with the black curve. Ideally, the complete elimination of model error is desired. As explained earlier, the PME is not completely eliminated because of filter divergence in the parameter space.

The problem of filter divergence in the parameter space is addressed by employing the stochastic step forward equation for the parameter rather than using the deterministic one given by equation (4.2). The stochastic step forward (SSF) is obtained by including noise in the DSF.

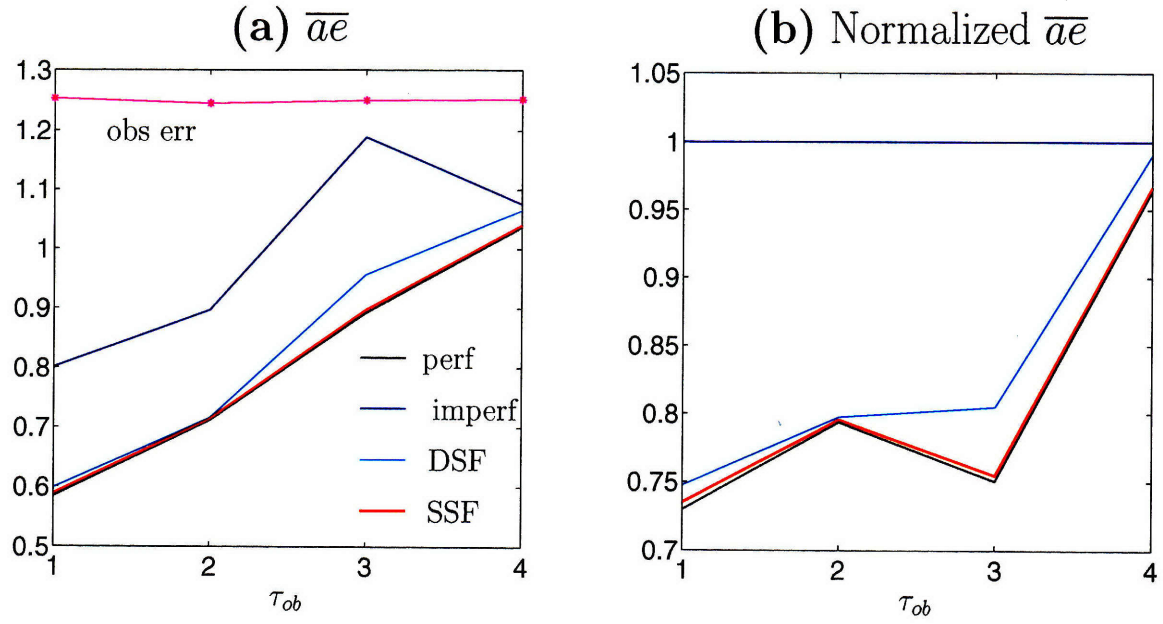


Figure 4-2: Comparison of average analysis errors for imperfect μ. **(a)** The imperfect model (blue) is defined by $\Delta\mu = -0.08$. **(b)** Curves in panel **(a)** normalized by the blue curve in panel **(a)**.

The SSF is given by

$$\mu_{s+1} = \mu_s + N(0, \sigma_r^\mu) \tag{4.5}$$

The subscript r in σ_r^μ stands for *random* or *stochastic* noise used in the SSF. The stochastic noise in the SSF is interpreted as an inflation factor in the parameter space. The magnitude of this stochastic noise is chosen to be 5×10^{-4}. The performance of SSF is assessed by running another filter which assimilates the same data as the perfect, imperfect and DSF filters. The SSF filter, similar to the DSF filter, uses the augmented state, that is, it uses an ensemble of dimension $n_a \times N$. The parameter ensemble used by SSF at time $s = 0$ is the same as that used by DSF at $s = 0$. The SSF filter uses equation (4.5) to step forward each member in the parameter ensemble. The mean of the analysis parameter ensemble, denoted by $\hat{\mu}_r$, is noted for each DA cycle. The red curve in figure 4-1(b) shows $\hat{\mu}_r$. The stochastic step forward equation prevents the parameter estimate from converging to a particular value. $\hat{\mu}_r$ tends to vary around the perfect value of μ shown by the black curve. The first few steps of this red curve are shown as the dashed red curve in figure 4-1(a).

Figure 4-3 shows $\hat{\mu}$ for DSF and SSF, for various τ_{ob}, as % deviation from the perfect value of μ=0.9. $\hat{\mu}_d$ is the value $\hat{\mu}_d$ converges to as s increases. $\overline{\hat{\mu}_r}$ is the mean of $\hat{\mu}_r$ over all steps. Note $\Delta\hat{\mu}_d$ and $\Delta\overline{\hat{\mu}_r}$ are the absolute value of the % deviations, that is, the sign of the deviations is neglected.

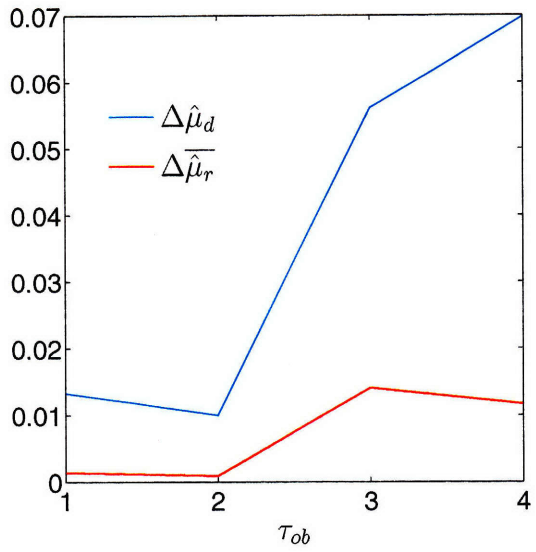


Figure 4-3: Estimates of μ as % deviations from its perfect value. Only parameter μ is imperfect.

It is seen from figure 4-3 that the SSF estimates are closer to the perfect $\mu=0.9$ than are the corresponding DSF estimates for all τ_{ob}. For example, for $\tau_{ob}=1$, $\hat{\mu}_d=0.0132$ and $\overline{\hat{\mu}_r}=0.0014$. The mean of the estimate given by SSF is closer to $\mu=0.9$ by an order of magnitude compared to the estimate given by DSF. This is seen in figure 4-1(b). The red curve bounces around the black curve while the cyan curve stabilizes to a value below the black curve.

The $\overline{ae}$ for the SSF filter is shown as the red curve in figure 4-2(a). The red curve lies very close to the black curve. Also the red curve lies below the cyan curve, implying that the SSF improves upon the performance of DSF. Figure 4-2(b) shows that the SSF has about $1-5\%$ advantage over the DSF in the average analysis errors. This is the consequence of the SSF being able to reduce the model error more (on average) than the DSF, as seen in figure 4-3. For $\tau_{ob}=1,2$, the estimate of SSF is an order of magnitude closer to $\mu=0.9$ than that of DSF.

The red curve shown in figure 4-2(a) uses $\sigma_r^\mu=5\times10^{-4}$. The value of σ_r^μ that gives the lowest analysis errors cannot be chosen a priori. Therefore, the SSF filter is run for different values of σ_r^μ. It is found that the lowest $\overline{ae}$ is obtained for $\sigma_r^\mu=5\times10^{-4}$ for all τ_{ob}. $\Delta\overline{\hat{\mu}_r}$ shown in figure 4-3 correspond to $\sigma_r^\mu=5\times10^{-4}$.

It is clear that parameter estimation reduces model error and consequently decreases the average analysis error. The SSF performs better than the DSF in decreasing the average analysis errors. The performance of the DSF and SSF in regards to the forecast errors is considered next. For a given τ_{ob}, forecasts are launched with lead time of $\tau=1,2,3,4$ for each of the four filters. These forecasts are launched from the corresponding analysis. For the

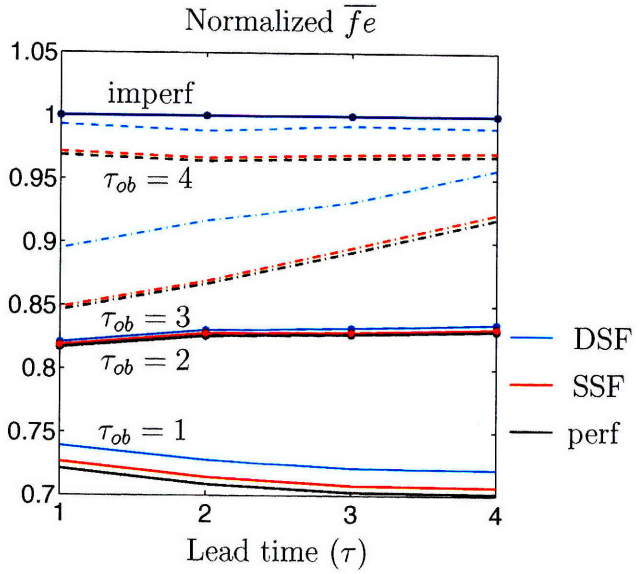


Figure 4-4: Forecast errors when μ is estimated. Each line type is for a particular τ_{ob}.

SSF filter the magnitude of stochastic noise is chosen to be $\sigma_r^{\mu}=5\times10^{-4}$, because it gives the minimum analysis error.

Figure 4-4 shows the average forecast errors ($\overline{fe}$) for each filter normalized by the $\overline{fe}$ for the imperfect filter, as function of lead time. Each line type shows the result for a different τ_{ob}. Clearly, parameter estimation decreases the average forecast errors compared to the imperfect forecast errors. The SSF parameter estimation outperforms the DSF parameter estimation for all τ_{ob}, for all lead times. The gains by employing parameter estimation are higher for the shorter τ_{ob}. For $\tau_{ob}=1$ the gain by using DSF or SSF is $\approx 27\%$ for all τ. The SSF outperforms DSF in forecast errors as it did in the analysis errors. The advantage of SSF over DSF varies between 0.5% and 5% depending on τ_{ob}. It appears that given a τ_{ob}, the advantage of SSF over DSF is approximately constant for different τ.

SSF has an advantage over DSF in both the analysis and forecast errors when μ is imperfect. μ is one of the three parameters in the Ikeda model. The other two parameters are a and b. Next, the estimation of a and b separately is considered. The setup of these experiments is similar to that described for μ except that $\boldsymbol{\alpha} = a$ and $\boldsymbol{\alpha} = b$ for estimation of a and b, respectively.

The experiment for estimation of a is as follows. The imperfect model is defined by the imperfect value of a given by $a_e = a + \Delta a$, where $a = 0.4$ and $\Delta a = 0.5$. b and μ are taken to be perfect, that is $b = 6.0$ and $\mu = 0.9$, respectively. The augmented ensembles for DSF and SSF are constructed in a manner analogous to the description given for estimation of μ, except that in this case the parameter to be estimated is a rather than μ, that is, $\boldsymbol{\alpha}=a$. The correct value of a, that is, $a=0.4$ is displaced by 1% of 0.4. This displaced value is taken as

the mean of $\mathbf{E}_{\boldsymbol{\alpha}}$. The members of $\mathbf{E}_{\boldsymbol{\alpha}}$ are constructed by adding perturbations to the mean of $\mathbf{E}_{\boldsymbol{\alpha}}$. These perturbations are drawn from $N(0, \sigma_a)$, where $\sigma_a = 1\%$ of 0.4. The DSF uses the following equation for stepping forward the ensemble members in $\mathbf{E}_{\boldsymbol{\alpha}}$.

$$a_{s+1} = a_s \tag{4.6}$$

The SSF uses the following equation for stepping forward the ensemble members in $\mathbf{E}_{\boldsymbol{\alpha}}$.

$$a_{s+1} = a_s + N(0, \sigma_r^a) \tag{4.7}$$

The magnitude of the stochastic noise (σ_r^a) in SSF that would result in lowest analysis errors is not known a priori. Therefore experiments are run with different values of σ_r^a. The values of σ_r^a that result in lowest $\overline{ae}$ are 5×10^{-4}, 5×10^{-4}, 5×10^{-2} and 5×10^{-3} for $\tau_{ob} = 1, 2, 3, 4$, respectively. The results of SSF for estimation of a presented in this section correspond to these σ_r^a. Figure 4-5(b) shows the estimates of a as given by DSF and SSF. As in the case of μ, the estimates of a given by SSF are, on average, closer to $a = 0.4$ than those given by DSF. This results in average analysis errors for SSF that are lower than the DSF. The average analysis errors are shown in Figure 4-5(a). The SSF outperforms the DSF both in average analysis and forecast errors (result not shown) when a is imperfect.

Next, the experiment for the estimation of b is described. The intercomparison experiments for imperfect b are run similarly to that for imperfect a. The imperfect model is defined by the imperfect value of b given by $b_e = b + \Delta b$, where b=6.0 and Δb=0.1. a and μ are taken to be perfect, that is a=0.4 and μ=0.9, respectively. The filter that estimates b uses $\boldsymbol{\alpha} = b$ and perfect value for a and μ. The correct value of b, that is, $b = 6.0$ is displaced by 1% of 6.0. This displaced value is taken as the mean of $\mathbf{E}_{\boldsymbol{\alpha}}$. Then the members of $\mathbf{E}_{\boldsymbol{\alpha}}$ are constructed by adding perturbations to the mean of $\mathbf{E}_{\boldsymbol{\alpha}}$. These perturbations are drawn from $N(0, \sigma_b)$, where $\sigma_b = 1\%$ of 6.0. The DSF uses the following equation for stepping forward the ensemble members in $\mathbf{E}_{\boldsymbol{\alpha}}$.

$$b_{s+1} = b_s \tag{4.8}$$

The SSF uses the following equation for stepping forward the ensemble members in $\mathbf{E}_{\boldsymbol{\alpha}}$.

$$b_{s+1} = b_s + N(0, \sigma_r^b) \tag{4.9}$$

The estimates of b as obtained by DSF and SSF are shown in figure 4-5(d). SSF performs better in reducing the model error due to b in that its estimates are closer to the perfect value of b than those given by DSF. The SSF gives better $\overline{ae}$ (shown in figure 4-5(c)).

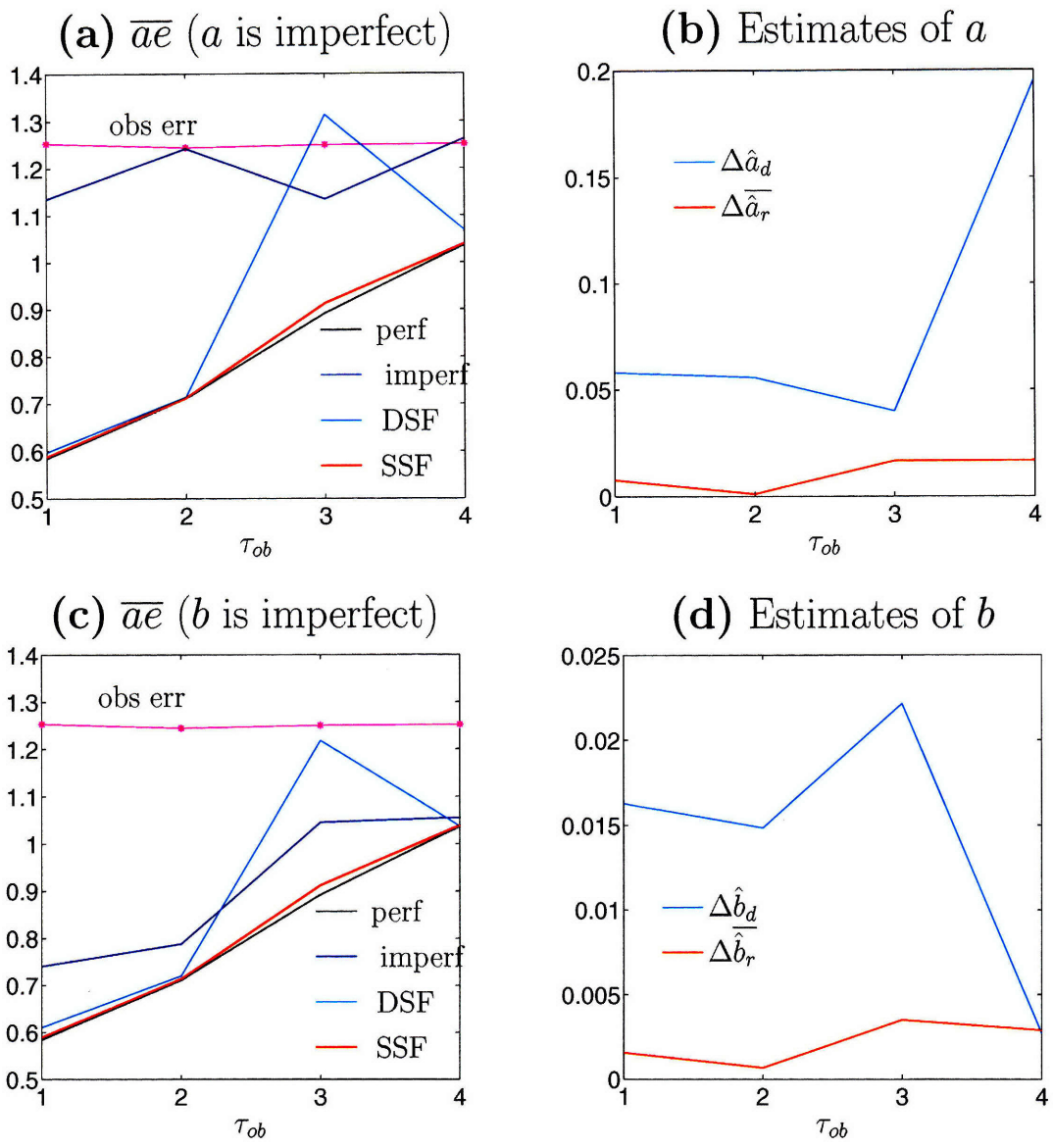


Figure 4-5: Estimation of single parameter. Panels **(a)** and **(b)** are for estimation of a. Panels **(c)** and **(d)** are for estimation of b.
(a) The imperfect model (blue curve) is defined by $\Delta a = 0.5$.
(b) Deviation of estimates of a from the perfect value of a.
(c) The imperfect model is defined by $\Delta b = 0.1$.
(d) Deviation of estimates of b from the perfect value of b.

It is clear from the results presented in this section (4.2.1) that parameter estimation can reduce the model error effectively in the case of one imperfect parameter. SSF clearly outperforms DSF in both $\overline{ae}$ and $\overline{fe}$. Typically, in real models, multiple parameters are imperfect. In section 4.2.2 the reduction of PME when more than one parameter is imperfect is treated. The basic framework is the same as described in this section except that the state is augmented by multiple parameters rather than by a single parameter.

4.2.2 Estimation of multiple parameters

The Ikeda model has three parameters, namely, a, b and μ. In this section the experiments to simultaneously estimate different combinations of a, b and μ are described.

To begin with the simultaneous estimation of a and b is considered. The experiment is as follows. The imperfect model is defined by imperfect values of a and b. Parameter μ is perfect, that is, $\mu = 0.9$. The imperfect model has value of a and b given by a_e and b_e where $a_e = a + \Delta a$ and $b_e = b + \Delta b$. $\Delta a = 0.1$ and $\Delta b = 0.1$ where Δa and Δb are in % of perfect values of a and b, respectively.

The filter used for the estimation of a and b has the following specifications. The state is augmented by $\boldsymbol{\alpha} = [a \;\; b]$. Therefore, the dimension of the augmented state is $n_a = n + n_{\boldsymbol{\alpha}} = 2 + 2 = 4$. $\mathbf{E}_{\boldsymbol{\alpha}}$ has dimension of $n_{\boldsymbol{\alpha}} \times N$, where $n_{\boldsymbol{\alpha}} = 2$. $\mathbf{E}_{\boldsymbol{\alpha}}$ is constructed as follows. The perfect value of $\boldsymbol{\alpha} = [0.4 \;\; 6.0]$ is displaced by 1% of its value. This displaced value is taken as the mean of $\mathbf{E}_{\boldsymbol{\alpha}}$. The members of $\mathbf{E}_{\boldsymbol{\alpha}}$ are constructed by adding perturbations to the mean of $\mathbf{E}_{\boldsymbol{\alpha}}$. For the ensemble in a, the perturbations are drawn from $N(0, \sigma_a)$, where σ_a is chosen to be $\sigma_a = 1\%$ of $a = 0.4$. For the ensemble in b, the perturbations are drawn from $N(0, \sigma_b)$, where σ_b is chosen to be $\sigma_b = 1\%$ of $b = 6.0$. For all the experiments in this section the ensemble size $N = 100$.

The matrix $\mathbf{H}(t)$ is given by,

$$\mathbf{H}(\mathrm{t}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{4.10}$$

Neither of the parameters a and μ is observed. The estimation of parameters is carried out using the DSF and SSF filters. The step forward equations for $\boldsymbol{\alpha}$ for the DSF are given by

$$a_{s+1} = a_s \tag{4.11a}$$

$$b_{s+1} = b_s \tag{4.11b}$$

The step forward equations for $\boldsymbol{\alpha}$ for SSF are given by

$$a_{s+1} = a_s + N(0, \sigma_r^a) \tag{4.12a}$$

$$b_{s+1} = b_s + N(0, \sigma_r^b) \tag{4.12b}$$

Four filters are run in parallel and the same data are assimilated in all these filters. Each of these filters use the EnKF equations given by 4.3 (page 93). These four filters correspond to the perfect model, the imperfect model, DSF filter and SSF filter. The DSF and SSF use the

state ensembles augmented by $\boldsymbol{\alpha}$. The mean of $\mathbf{E}_{\boldsymbol{\alpha}}$ is noted after every analysis. This mean is a vector denoted by $[\hat{a}_d \;\; \hat{b}_d]$ for the DSF filter and $[\hat{a}_r \;\; \hat{b}_r]$ for the SSF filter. The average analysis errors for these filters as a function of τ_{ob} is shown in figure 4-6(a). The red curve corresponds to the SSF filter. Since the value of σ_r^a and σ_r^b that would give the minimum $\overline{ae}$ is not known a priori, experiments for SSF filter are run with different values of σ_r^a and σ_r^b. The red curve in figure 4-6(a) corresponds to the minimum $\overline{ae}$ among the different σ_r^a and σ_r^b. It is clear from figure 4-6(a) that the DSF gives lower analysis errors than the imperfect filter. The SSF improves upon the performance of the DSF filter.

Figure 4-6(b) shows $[\hat{a}_d \;\; \hat{b}_d]$ and $[\overline{\hat{a}_r} \;\; \overline{\hat{b}_r}]$ as % deviations from the perfect values, that is, [0.4 6.0]. $[\hat{a}_d \;\; \hat{b}_d]$ are convergent values given by DSF for a given τ_{ob}. $[\overline{\hat{a}_r} \;\; \overline{\hat{b}_r}]$ are the mean of $[\hat{a}_r \;\; \hat{b}_r]$ time series for a given τ_{ob}. The deviations shown in figure 4-6(b) are the absolute values of the % deviations. Consider the estimate of a for DSF and SSF, that is, the solid curves. The solid red curve lies closer to the zero level than does the solid cyan curve. This means that the SSF is better at retrieving the value of a, on average, than DSF for all τ_{ob}. For $\tau_{ob} = 4$ the deviation for DSF is larger than that for SSF by at least an order of magnitude. Consider the estimates for b, that is, the dashed curves in figure 4-6(b). Similar to the result for a, the SSF is better at retrieving the perfect value of b (on average) than is the DSF. SSF is better at retrieving the values of both a and b. This leads to lower analysis errors for SSF than those for DSF as shown in figure 4-6(a). This result is similar to the corresponding result in the case of single parameter estimation discussed in section 4.2.1.

Consider the estimates of a and b as given by DSF (solid cyan and dashed cyan curves) in figure 4-6(b). The error in the estimate of a and b is quantified by the deviations $\Delta\hat{a}_d$ and $\Delta\hat{b}_d$. For example, for $\tau_{ob} = 4$, $\Delta\hat{a}_d = 0.8$ and $\Delta\hat{b}_d = 0.12$. That is, estimates of a and b deviate significantly from their perfect values and yet the DSF is able to give lower average analysis error (by $\approx 10\%$ for $\tau_{ob} = 4$) compared to the imperfect filter as shown in figure 4-6(a). This is because the errors in the estimate of a and b compensate for each other. The perfect values of a and b need not be retrieved. Parameter estimation could result in significantly imperfect values for a and b in such a manner so as to give lower $\overline{ae}$. The mutual compensation for errors in parameter estimates also happens in the case of SSF.

Next, the experiment for the estimation of b and μ is described. The set-up of experiment is similar to that in the case of estimation of a and b, except that in this case $\boldsymbol{\alpha} = [b \;\; \mu]$. The imperfect model is defined by $\Delta b = 0.1$ and $\Delta\mu = -0.04$. The mean of $\mathbf{E}_{\boldsymbol{\alpha}}$ is obtained by displacing $\boldsymbol{\alpha} = [b \;\; \mu]$ by 1% from its perfect value of $\boldsymbol{\alpha} = [6.0 \;\; 0.9]$. $\mathbf{E}_{\boldsymbol{\alpha}}$ is constructed by adding to the mean, perturbations drawn from $[\sigma_b \;\; \sigma_\mu]$ where $\sigma_b = 1\%$ of b=6.0 and $\sigma_\mu = 1\%$ of μ=0.9.

The comparison for $\overline{ae}$ is shown in figure 4-6(c). Clearly, SSF improves upon the imperfect

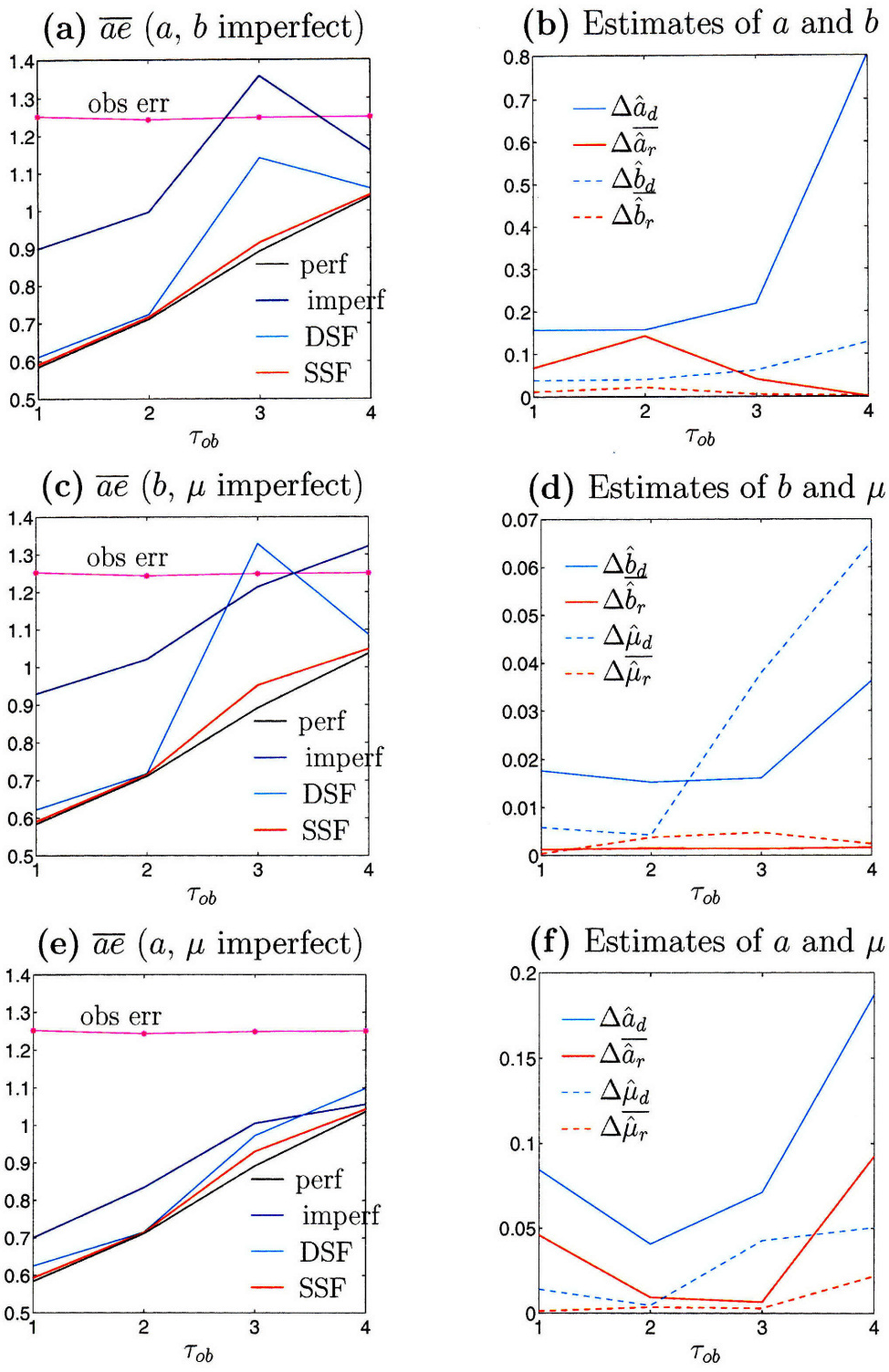


Figure 4-6: Simultaneous estimation of two parameters.
(a) Imperfect model (blue curve) is defined by $[\Delta a \;\; \Delta b] = [0.1 \;\; 0.1]$.
(c) Imperfect model (blue curve) is defined by $[\Delta b \;\; \Delta \mu] = [0.1 \;\; -0.04]$.
(e) Imperfect model (blue curve) is defined by $[\Delta a \;\; \Delta \mu] = [0.1 \;\; -0.04]$.

filter by a big margin. The estimates of b and μ are shown as % deviation from their perfect values in figure 4-6(d). For both the parameters, the SSF gives a better estimate of the parameters than does the DSF. The DSF gives lower $\overline{ae}$ than the imperfect model except for $\tau_{ob}=3$. The SSF gives lower $\overline{ae}$ than either the imperfect model or the DSF.

The results for the estimation of a and μ are shown in panels figure 4-6(e) and figure 4-6(f). The set-up of this experiment is similar to the experiment for estimation of b and μ. The imperfect model is defined by $\Delta a = 0.1$ and $\Delta\mu = -0.04$. For the DSF and SSF filters the state is augmented by $\boldsymbol{\alpha}=[a \;\; \mu]$. $\mathbf{E}_{\boldsymbol{\alpha}}$ is constructed by displacing $\boldsymbol{\alpha}=[a \;\; \mu]$ by 1% of its value and then adding to this value perturbations drawn from $[\sigma_b \;\; \sigma_\mu]=[1\% \;\; 1\%]$ of $[0.4 \;\; 0.9]$. The DSF gives lower $\overline{ae}$ than the imperfect model for all τ_{ob} except $\tau_{ob}=4$. The SSF gives lower $\overline{ae}$ than either the imperfect model or the DSF. This is clearly because the SSF gives a better estimate of the parameters than does the DSF as shown in figure 4-6(f).

To summarize the results of simultaneous estimation of two parameters, the SSF results in lower average analysis errors than either the imperfect model or the DSF for all τ_{ob}. The errors in the estimation of the two parameters could compensate for each other so that lower $\overline{ae}$ could result without the retrieval of the perfect value of the parameters.

Next, the experiment for the simultaneous estimation of three parameters is discussed. The imperfect model is defined by $[\Delta a \;\; \Delta b \;\; \Delta\mu] = [0.1 \;\; 0.1 \;\; -0.04]$. For the DSF and SSF filters the state is augmented by $\boldsymbol{\alpha}=[a \;\; b \;\; \mu]$. Data are assimilated using the EnKF equations 4.3. The dimension of $\mathbf{E}_{\boldsymbol{\alpha}}$ is $n_{\boldsymbol{\alpha}} \times N$ where $n_{\boldsymbol{\alpha}} = 3$. The dimension of the augmented state is $n_a = n + n_{\boldsymbol{\alpha}} = 2 + 3 = 5$. The matrix $\mathbf{H}(t)$ is given by

$$\mathbf{H}(\mathrm{t})=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{4.13}$$

None of the three parameters is observed. The specifications for the filter used to estimate the parameters is as follows. The mean of $\mathbf{E}_{\boldsymbol{\alpha}}$ is obtained by displacing the perfect value of $[a \;\; b \;\; \mu] = [0.4 \;\; 6.0 \;\; 0.9]$ by 1%. $\mathbf{E}_{\boldsymbol{\alpha}}$ is constructed by adding (to this mean) perturbations drawn from $[\sigma_a \;\; \sigma_b \;\; \sigma_\mu] = [1\% \;\; 1\% \;\; 1\%]$ of $[0.4 \;\; 6.0 \;\; 0.9]$. The step forward equations for $\boldsymbol{\alpha}$ for DSF are given by

$a_{s+1} = a_s$

$b_{s+1} = b_s$

$\mu_{s+1} = \mu_s$

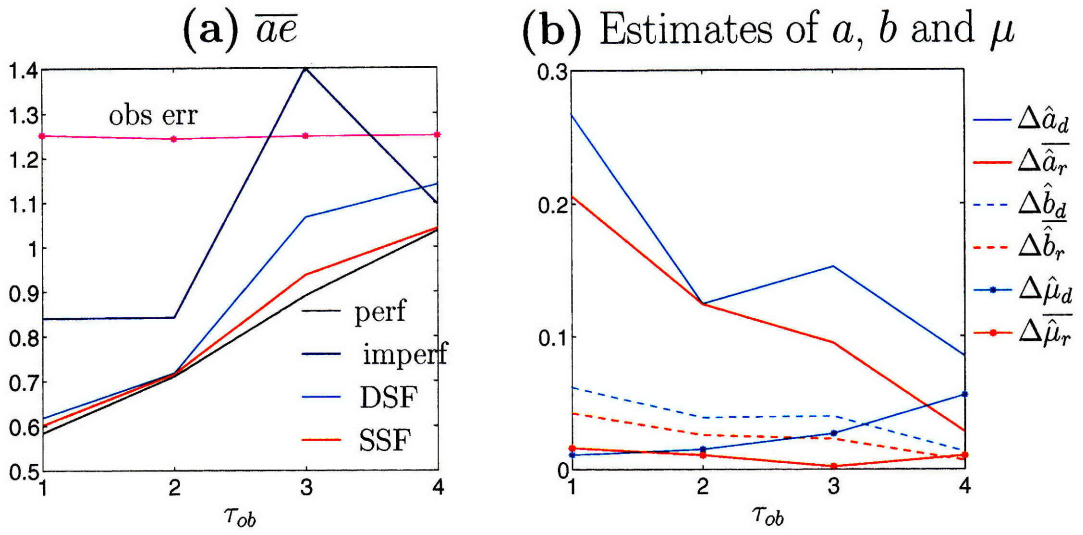


Figure 4-7: Simultaneous estimation of a, b and μ. Ensemble size $N=100$.
(a) Imperfect model (blue) is defined by $[\Delta a \;\; \Delta b \;\; \Delta\mu]=[0.1 \;\; 0.1 \;\; -0.04]$.
(b) Estimates of a, b and μ as % deviations from their perfect values. For a given parameter, the SSF estimate (red) is closer to 0 than the DSF estimate(cyan). An exception occurs for the estimate of μ at $\tau_{ob}=1$.

The step forward equations for $\boldsymbol{\alpha}$ for SSF are given by

$$a_{s+1} = a_s + N(0, \sigma_r^a) \tag{4.15a}$$

$$b_{s+1} = b_s + N(0, \sigma_r^b) \tag{4.15b}$$

$$\mu_{s+1} = \mu_s + N(0, \sigma_r^\mu) \tag{4.15c}$$

The comparison of average analysis errors in shown in figure 4-7(a). As in the case of estimation of two parameters, SSF gives lower $\overline{ae}$ than either the DSF or the imperfect model. The values of σ_r^a, σ_r^b and σ_r^μ in equations 4.15 that would give lowest $\overline{ae}$ is not known *a priori*. Therefore, experiments are run for various values of σ_r^a, σ_r^b and σ_r^μ. The red curve shown in figure 4-7(a) and the curves shown in figure 4-7(b) corresponds to values of σ_r^a, σ_r^b and σ_r^μ that give least $\overline{ae}$. As shown in figure 4-7(b) the estimates of parameters as given by SSF are closer to the perfect values than those given by DSF. Consider the DSF estimates of the parameters shown in cyan for $\tau_{ob}=1$. The error in the estimate of a is as high as 0.26. Even so, the DSF improves over the imperfect model as shown in figure 4-7(a). This is because the error in the estimate of a could be compensated by errors in estimates of b and μ.

In the current section (4.2) the estimation of single and multiple parameters is discussed. It is found that the SSF gives a better estimate of the parameters than does the DSF for estimation of single and multiple parameters. Consequently, SSF results in better average errors than the DSF. In general, DSF gives lower average error than the parametrically imperfect model. In the case of simultaneous estimation of multiple parameters the fact that estimation results

in lower average errors need not mean that the perfect parameter values are retrieved. It is possible that error in the estimate of one parameter are compensated by those in other parameters.

The current section (4.2) assumed that the model is parametrically imperfect and structurally perfect. Almost all real atmospheric models are structurally imperfect apart from being parametrically imperfect. In general, structural model error (SIM) is more realistic than PIM. The next section (4.3) discusses parameter estimation in the presence of structural model error.

4.3 Parameter tuning in the Ikeda model

In this section the Ikeda model with only SME is considered. The perfect Ikeda model is defined by the following equations.

$$x_{s+1} = 1 + \mu(x_s \cos\theta - y_s \sin\theta) \tag{4.16a}$$

$$y_{s+1} = \mu(x_s \sin\theta + y_s \cos\theta) \tag{4.16b}$$

where

$$\theta = a - \frac{b}{(x_s^2 + y_s^2 + 1)}$$

and $a=0.4$, $b=6$, $\mu=0.9$. Perfect Ikeda model means the model is both parametrically and structurally perfect. Consider the following parameterically perfect, but structurally imperfect Ikeda model,

$$x_{s+1} = 1 + \mu(x_s tcos\theta - y_s tsin\theta) \tag{4.17a}$$

$$y_{s+1} = \mu(x_s tsin\theta + y_s tcos\theta) \tag{4.17b}$$

$$\text{where } \theta = a - \frac{b}{(x_s^2 + y_s^2 + 1)} \tag{4.17c}$$

$a=0.4$, $b=6$, $\mu=0.9$.

Note that the perfect parameter values are used, but $\sin\theta$ and $\cos\theta$ in the perfect model equations are replaced by $tsin\theta$ and $tcos\theta$, which represent the truncated sin series and truncated cos series respectively. The following equations gives $tcos\theta$ and $tsin\theta$ truncated upto the first

k terms, that is terms $> k$ are ignored.

$$tcos\theta = \sum_{n=0}^{k} \frac{(-1)^n \theta^{2n}}{2n!} \tag{4.18a}$$

$$tsin\theta = \sum_{n=0}^{k} \frac{(-1)^n \theta^{2n+1}}{(2n+1)!} \tag{4.18b}$$

The model denoted by SIM1 (structurally imperfect model 1) has $k=7$, that is, the *tsin* and *tcos* series include only the first 7 terms. Including only the first $k=7$ terms is defined to be a low level of structural model error. SIM1 is Ikeda model with low structural model error. The model denoted by SIM2 has $k=6$. SIM2 is defined to be Ikeda model with high structural model error in comparison with SIM1.

The aim of parameter estimation is to reduce the structural model error in SIM1 and SIM2. The reduction of model error by the estimation of parameters in the presence of structural model error is termed *parameter tuning*. The term *parameter estimation* is reserved for the case when it is known that the model is parametrically imperfect but structurally perfect. The term *parameter tuning* implies that parameter(s) are being estimated in order to offset structural model error. If the DSF is used for parameter tuning, then it is called *deterministic parameter tuning* (DPT). If the SSF is used for parameter tuning then it is called *stochastic parameter tuning* (SPT). Section 4.3.1 describes DPT and SPT for SIM1 and presents the intercomparison results. Section 4.3.2 presents intercomparison results for SIM2. For experiments described in sections 4.3.1 and 4.3.2 the ensemble size is $N=100$.

4.3.1 Low structural model error (SIM1)

Consider the tuning of parameter μ in the presence of low structural model error. The EnKF described by equations (4.3a) through (4.3e) is implemented. The state is augmented by the parameter μ. Therefore, $\boldsymbol{\alpha}=\mu$ and $n_{\boldsymbol{\alpha}}=1$. The dispersion within the initial $\mathbf{E}_{\boldsymbol{\alpha}}$ is $\sigma_\mu=1\%$ of 0.9. The DPT and SPT filters are run in parallel using the same data. The DPT uses equation (4.2) and SPT uses equation (4.5) for stepping forward the ensemble members of $\mathbf{E}_{\boldsymbol{\alpha}}$. The SPT filter is run with different magnitudes of stochastic noise (σ_r^μ). The results presented correspond to the value of σ_r^μ that yields the lowest analysis errors. The perfect and imperfect filters are also run in parallel with the DPT and SPT filters.

Consider the DPT for $\tau_{ob}=1$. As described in section 4.2.1 (page 92), data assimilation tends to minimize the dispersion within $\mathbf{E}_{\boldsymbol{\alpha}}$. The estimate of μ, namely $\hat{\mu}_d$, is shown as the cyan curve in figure 4-8. The perfect value of $\mu=0.9$ is shown as the black curve. The dispersion

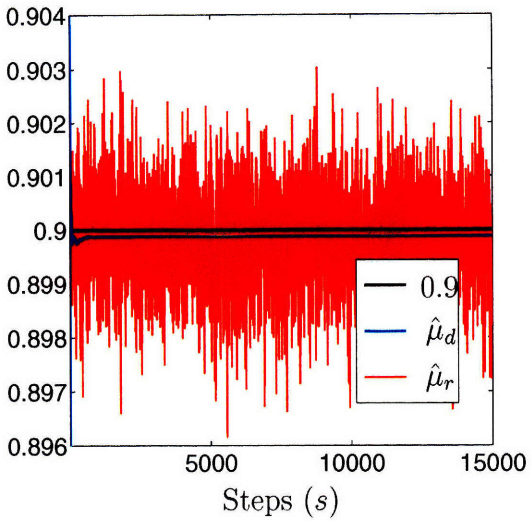


Figure 4-8: Estimates of μ for SIM1

within $\mathbf{E}_{\boldsymbol{\alpha}}$ goes to 0 in about $s = 100$ steps and the value of $\hat{\mu}_d$ converges to ≈ 0.8995. The ability of the tuning of μ to offset the structural model error in SIM1 saturates at $s = 100$. At this point all the ensemble members of $\mathbf{E}_{\boldsymbol{\alpha}}$ collapse to the same value. This problem of filter divergence is solved by using SPT. Consider the SPT for $\tau_{ob} = 1$. The estimate $\hat{\mu}_r$ is shown in red in figure 4-8. The value of $\hat{\mu}_r$ varies a lot and does not settle down to a particular value. This is of course due to the stochastic step forward equation used for propagating the members of $\mathbf{E}_{\boldsymbol{\alpha}}$. The variation in $\hat{\mu}_r$ is such as to offset the structural model error in SIM1. This statement is supported by the comparison between average analysis errors shown

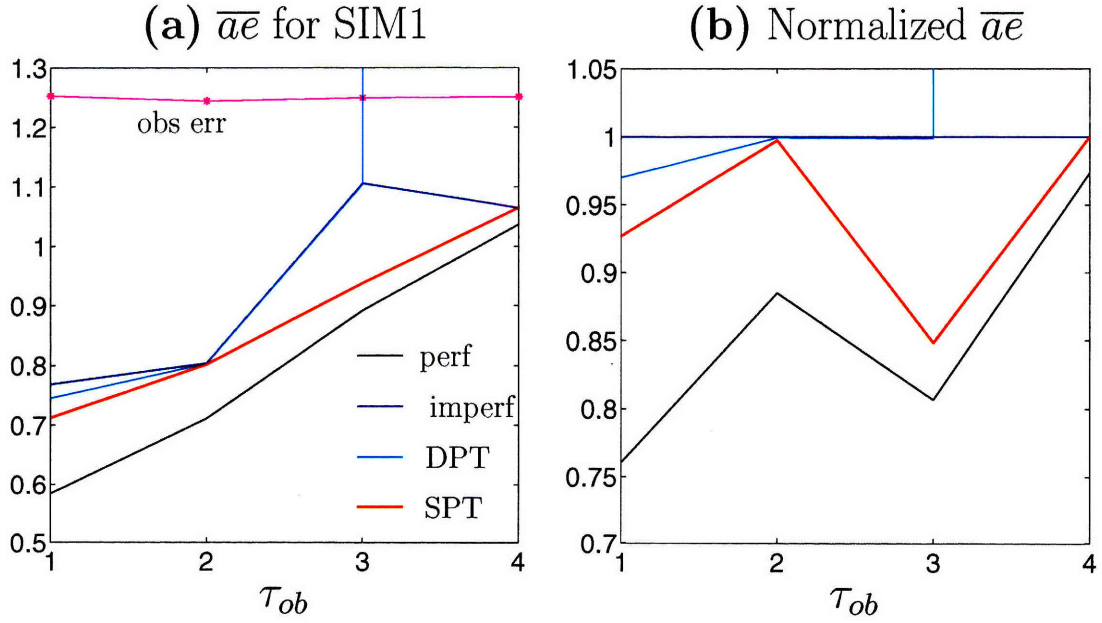


Figure 4-9: Analysis errors for tuning of μ for SIM1.(a) Average analysis error expressed in % attractor size. (b) $\overline{ae}$ in panel(a) normalized by the $\overline{ae}$ of imperfect filter.

in figure 4-9(a). It is seen that SPT gives lower $\overline{ae}$ than the DPT (The red curve lies below the cyan curve.) The perfect (black curve) and imperfect (blue curve) are shown for comparison. The advantage of SPT over DPT is quantified by normalizing each curve in panel (a) by the imperfect curve. This normalized $\overline{ae}$ is shown in panel (b). The SPT offers an advantage of about 1-15% over DPT depending on τ_{ob}. The least advantage is realized for $\tau_{ob} = 2$. The highest advantage is realized for $\tau_{ob}=4$, when the DPT results in the distortion of the Ikeda attractor. The issue of distortion of the attractor is discussed in more detail in section 4.3.2. It should be noted that for $\tau_{ob}=4$ the SPT result is almost the same as that of the imperfect model.

It is important to understand that the DPT and SPT do not try to recover the value of $\mu=0.9$, as in the case of parameter estimation described in section 4.2. In section 4.2 the model was parameterically imperfect but structurally perfect. In the context of PIM the reduction of model error involves the data driving the parameter estimate towards the perfect value of $\mu=0.9$. In the context of SIM, the reduction of model error involves tuning the parameter in order to offset structural error. The aim of parameter tuning is not to drive the value of the tuned parameter to the perfect parameter value. This is evidenced in the values of $\hat{\mu}_r$ and $\hat{\mu}_d$. These are tabulated as % deviation from the perfect value of $\mu=0.9$ in table 4.1.

τ_{ob}	1	2	3	4
$\hat{\mu}_d$ (10^{-2})	1.30	2.03	7.56	5.08
$\overline{\hat{\mu}_r}$ (10^{-2})	1.86	3.39	2.67	3.21

Table 4.1: Tuned values of μ for SIM1

For $\tau_{ob}=1,2$, $\hat{\mu}_d$ is closer to $\mu=0.9$ than $\hat{\mu}_r$ is on average. But $\overline{\hat{\mu}_r}$ need not be close to $\mu=0.9$ to be able to compensate for the structural model error in SIM. In fact, at a particular time step s, $\hat{\mu}_r$ has the value that seeks to offset the structural model error at that particular step. This is the reason why $\hat{\mu}_r$ varies with time s in figure 4-8. The DPT also tries to compensate for structural model error and therefore it is not desirable for $\hat{\mu}_d$ to converge to $\mu=0.9$. $\hat{\mu}_d$ converges to some other value in its attempt to compensate for structural model error. Its ability to do so is limited by the filter divergence in the parameter space. The stochastic noise in SPT prevents filter divergence, thus allowing the tuned parameter value to compensate for the structural error in a state-dependent manner. For $\tau_{ob}=3,4$, $\hat{\mu}_r$ is closer to $\mu=0.9$ than $\hat{\mu}_d$ is on average. Irrespective of how close $\hat{\mu}_r$ is to $\mu=0.9$, SPT gives lower average errors than does DPT as shown in figure 4-9(a).

The Ikeda model has three parameters. The choice of tuning μ is arbitrary. Any of the other parameters, namely a and b could be tuned to compensate for the structural error. It is also possible to tune more than one parameter. Experiments are run to intercompare the efficacy

of tuning various parameter combinations to offset the structural error.

The filters DPT and SPT for a are run exactly in the same manner, except that $\mathbf{E}_{\boldsymbol{\alpha}}$ contains the ensemble of a rather than of μ. The initial dispersion within $\mathbf{E}_{\boldsymbol{\alpha}}$ is chosen to be $\sigma_a = 1\%$ of the perfect value of $a = 0.4$. The filters for a are designated DPT(a) and SPT(a). The filters for b are run similarly with $\mathbf{E}_{\boldsymbol{\alpha}}$ containing the ensemble of b. $\sigma_b = 1\%$ of the perfect value of $b = 6.0$. The filters for b are designated DPT(b) and SPT(b).

Next, the experiments for various combinations of multiple parameters are run. For example for the combination of a and b, $\mathbf{E}_{\boldsymbol{\alpha}}$ has dimension of $n_{\boldsymbol{\alpha}} \times N$, where $n_{\boldsymbol{\alpha}} = 2$. For $n_{\boldsymbol{\alpha}} = 2$, the matrix $\mathbf{H}(t)$ is given by equation 4.10 (page 101). Neither of the parameters a and b is observed. The filters for the combination of a and b are designated DPT(ab) and SPT(ab). Similar experiments are run for other combinations of parameters, namely, $b\mu$, $a\mu$ and $ab\mu$. For the case of $ab\mu$, $n_{\boldsymbol{\alpha}} = 3$ and consequently $\mathbf{E}_{\boldsymbol{\alpha}}$ has dimension of $3 \times N$. The filters are designated DPT($ab\mu$) and SPT($ab\mu$). The matrix $\mathbf{H}(t)$ for DPT($ab\mu$) and SPT($ab\mu$) is given by equation 4.13 (page 104). Of course, the same data are used in all these tuning experiments so that their performance can be intercompared. In all these experiments, $\sigma_{\boldsymbol{\alpha}} = 1\%$ of the perfect value of $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is the vector containing the corresponding parameters.

The performance of these filters in decreasing the average analysis errors is compared in figure 4-10 for different τ_{ob}. The curves show the average analysis errors normalized by the corresponding value for the imperfect model. The curve for the imperfect model, which has a value of 1 for all τ_{ob}, is not shown. The perfect model curve, which is not shown, lies at around 0.5.

In general, SSF outperforms DSF for a given combination of parameters. The exception occurs at $\tau_{ob} = 2$ for the combination $b\mu$ (blue), $a\mu$ (black) and $ab\mu$ (yellow). In general, SPT for any combination is better than DPT for any combination. The best performance is given by SPT(a) for all τ_{ob} among all combinations.

The comparison within SPT for single parameters shows that SPT(a) is the best, followed by SPT(b) followed by SPT(μ).

The comparison within SPT for various combinations of parameters is interesting. Consider the dashed lines in figure 4-10. Among the combination of two parameters, the combination of ab gives the lowest error for all τ_{ob}. This combination also outperforms the $ab\mu$ combination for all τ_{ob}. Thus the lowest analysis errors need not result from tuning all the parameters. For $\tau_{ob} = 1$, SPT($ab\mu$) is outperformed by SPT(a), SPT(ab), and SPT($b\mu$). For $\tau_{ob} = 2$, all other combinations outperform SPT($ab\mu$) except for SPT(μ). For $\tau_{ob} = 3$, SPT(a), SPT(b) and SPT(ab) outperform SPT($ab\mu$).

The comparison within DPT shows that in general it is better to tune multiple parameters than single parameters. However, tuning multiple parameters requires a high amount of data. Therefore, in practice, the number of tunable parameters is limited by the amount of data available.

Also it is to be noted that for $\tau_{ob}=4$, neither the DPT nor the SPT give considerable advantage over the imperfect model. The reduction of model error even with SPT may not be effective for all τ_{ob}.

Next, the forecast errors are considered. Figure 4-11(a) shows the $\overline{fe}$ for all filters normalized by the $\overline{fe}$ for the imperfect filter, for $\tau_{ob}=1$. The curves for the perfect and imperfect filter are not shown. In general parameter tuning gives lower forecast errors than the imperfect model. For a given parameter combination, SPT filters perform better than the DPT filters. In general, SPT($ab\mu$) gives the lowest $\overline{fe}$ except for $\tau=1$, when SPT(a) gives the lowest $\overline{fe}$. The SPT(a) filter gives the lowest $\overline{ae}$ as seen from figure 4-10 for $\tau_{ob}=1$. But SPT(a) does not give the lowest $\overline{fe}$ for all τ for $\tau_{ob}=1$. Thus the reduction of model error which might be effective in obtaining lowest analysis errors might not result in the lowest forecast errors. Among the two parameter combinations, SPT(ab) gave the lowest analysis error for $\tau_{ob}=1$. In the $\overline{fe}$ too, SPT(ab) gives the lowest error for all τ among all the two parameter combinations.

Figure 4-11(b) shows the normalized $\overline{fe}$ for $\tau_{ob}=2$. The SPT filters need not perform better than DPT filters, for a particular parameter combination. For example, DPT($b\mu$) and

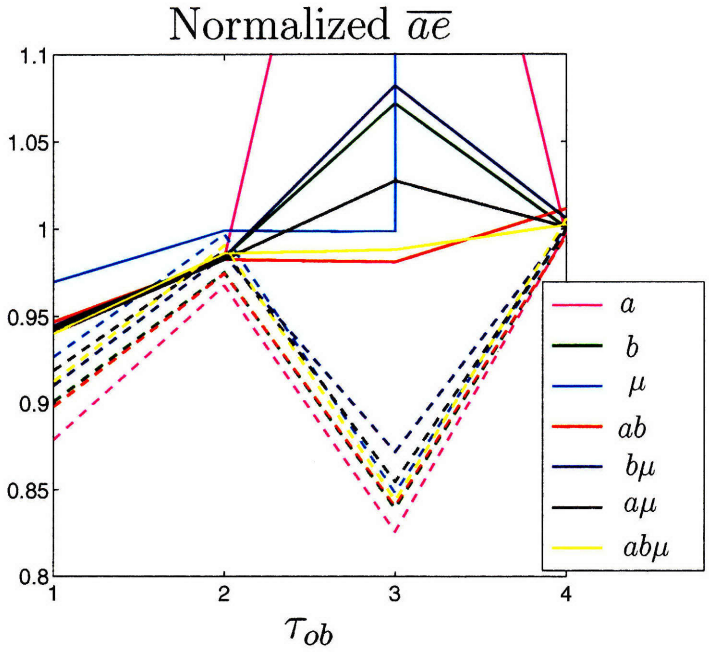


Figure 4-10: Intercomparison between various DPT and SPT filters for SIM1. Solid lines are for DPT and dashed lines for SPT. The different colors show the different parameters combinations that are tuned.

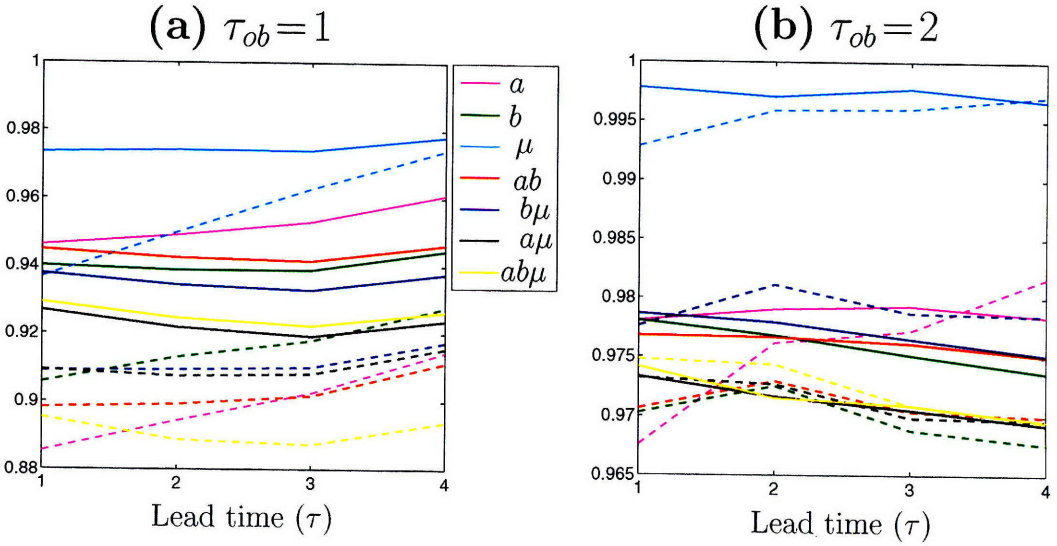


Figure 4-11: Normalized $\overline{fe}$ for SIM1. Solid lines are for DPT and dashed lines for SPT. The legend applies to both the panels.

DPT($ab\mu$) outperform the corresponding SPT filters for almost all τ. In general, different combinations give the least $\overline{ae}$ for different τ. In general, the least $\overline{ae}$ at a particular τ is given by a SPT filter.

In general, it appears that SPT is more effective in reducing SME than DPT, at least for the level of SME considered in this section.

This section (4.3.1) used SIM1 as the imperfect model. The next section (4.3.2) presents results for SIM2 which has more SME that SIM1. The experiments for SIM2 are run exactly in the same manner as those for SIM1.

4.3.2 High structural model error (SIM2)

The objective of this section is to ascertain whether some of features of parameter identified in the last section hold true in the case of high SME. The section 4.3.1 used SIM1 which had a low level of structural model error. The experiments in this section are similar to those described in section 4.3.1 except that in this section SIM2 is used instead of SIM1. SIM2 has a higher level of structural model error compared to SIM1, as defined in section 4.3 (page 106).

Figure 4-12 shows the $\overline{ae}$ for the perfect, imperfect, DPT(μ) and SPT(μ) filters. The observation error is 1.26. The curve for SPT(μ) shown as dashed red curve lies very close to the observational error. The imperfect filter error for $\tau_{ob} = 1$ is 10.67 and that for DPT(μ) is 10.34. The imperfect model results in very high analysis error, especially for $\tau_{ob} > 1$. This is because the Ikeda model attractor for this high level of SME is completely distorted. Simi-

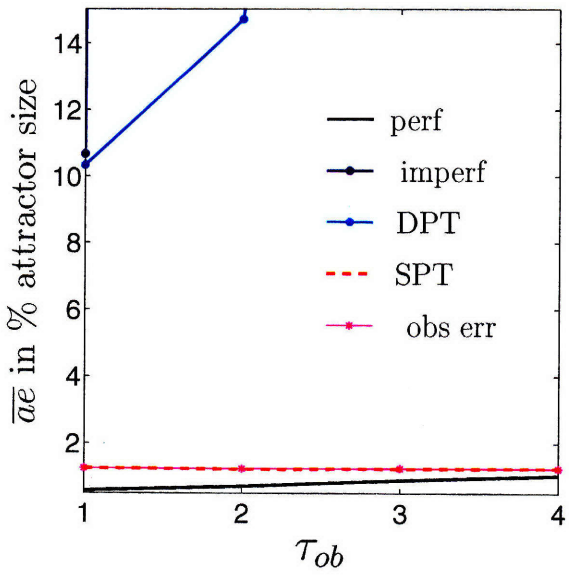


Figure 4-12: μ is tuned for SIM2.

larly, for $\tau_{ob} > 2$, the DPT filter gives a high error because the Ikeda attractor is distorted. Figure 4-13(a) shows the distorted attractor as defined by the analysis of DPT(μ) for $\tau_{ob}=3$. As described in section 4.3.1, the parameter tuning tries to compensate for the structural model error by changing the value of the parameter. In DPT the value of the parameter may change dramatically in order to compensate for the structural error. If, at a particular time step s, the value of $\hat{\mu}_d$ carries the analysis point to a point outside the basin of attraction, then subsequently the value of $\hat{\mu}_d$ may not change enough to bring the analysis back into the basin of attraction. Figure 4-13(b) shows $\hat{\mu}_d$ in cyan color. $\hat{\mu}_d$ varies in order to offset the structural error. Because the structural error is high, $\hat{\mu}_d$ may have to change dramatically to compensate for the model error. In the course of the variation of $\hat{\mu}_d$, a *bad* value of $\hat{\mu}_d$ carries the analysis to a fixed point, outside the basin of attraction. Thereafter it is not able to change much to return to the attractor. In DPT, the parameter value is trying to compensate for SME globally. The DPT is seeking a *single* value, that is, a state-independent value that would compensate for SME at all regions in phase space.

The red curve in figure 4-13(b) shows $\hat{\mu}_r$ (given by SPT). The analysis defined by $\hat{\mu}_r$ has an in-built variation in it owing to the stochastic noise in the step forward equation. Therefore, $\hat{\mu}_r$ has the ability to return the analysis to the basin of attraction in case it happens to leave the basin of attraction. SPT tries to obtain a value of $\hat{\mu}_r$ that is state-dependent. Therefore, $\hat{\mu}_r$ varies with time. SPT results in different values for $\hat{\mu}_r$ at different local regions in the phase space so that SME is offset locally. This is the reason why SPT gives far lower analysis error than DPT, as shown in figure 4-12.

The $\overline{ae}$ in figure 4-12 is for tuning μ. What about tuning other parameter(s)? Experiments

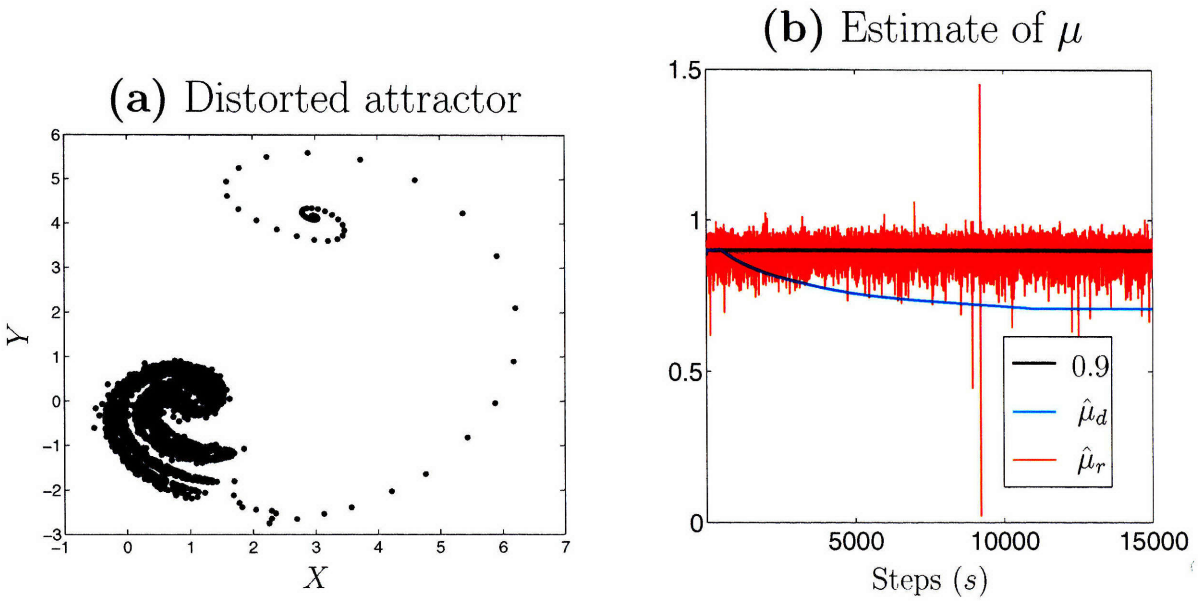


Figure 4-13: **(a)** The attractor defined by $\hat{\mu}_d$. **(b)** The cyan curve shows the $\hat{\mu}_d$ that resulted in the distorted attractor shown in (a).

are carried out in which different combinations of parameters are tuned. The average $\overline{ae}$ for $\tau_{ob} = 1$ are shown in table 4.2 for different parameter combinations. The $\overline{ae}$ for the perfect and imperfect filters are $\overline{ae} = 0.58$ and $\overline{ae} = 10.67$, respectively. The observation error is 1.26. It is clearly seen that SPT outperforms DPT for all combinations of parameters. Within the DPT filters, lower errors are obtained by tuning multiple parameters rather than a single parameter. Among the single parameter filters, tuning b gives the best result. Among all the parameter combinations, SPT(ab) gives the best result.

The results for $\tau_{ob} = 2, 3, 4$, which are not shown, bear out the statement that SPT filters outperform DPT filters in giving lower $\overline{ae}$.

Next, the forecast errors are considered. Figure 4-14 show the forecast errors for SPT filters for $\tau_{ob} = 1, 2$. (The curves for $\tau_{ob} = 2$ are scaled up so that they do not overlap with curves for $\tau_{ob} = 1$.) The curves for the imperfect and DPT filters (which are not shown) lie way above the SPT curves. The curves for the perfect filter (not shown) lie below the SPT filters. For $\tau_{ob} = 1$ and $\tau = 1$, the best performance is by SPT(ab), closely followed by SPT(a). For $\tau_{ob} = 1$, $\tau = 2$ the best performance is by SPT(μ), closely followed by SPT(a). For $\tau_{ob} = 2$, the lowest fer is for SPT(μ) for $\tau = 1, 2$. Among the multiple parameters, the lowest fer is for SPT(ab)

Parameter	a	b	μ	ab	$b\mu$	$a\mu$	$ab\mu$
DPT	719.30	715.86	10.34	10.86	11.95	12.32	10.26
SPT	1.25	0.99	1.26	0.98	1.09	1.22	1.07

Table 4.2: $\overline{ae}$ in % attractor size for SIM2 for $\tau_{ob} = 1$

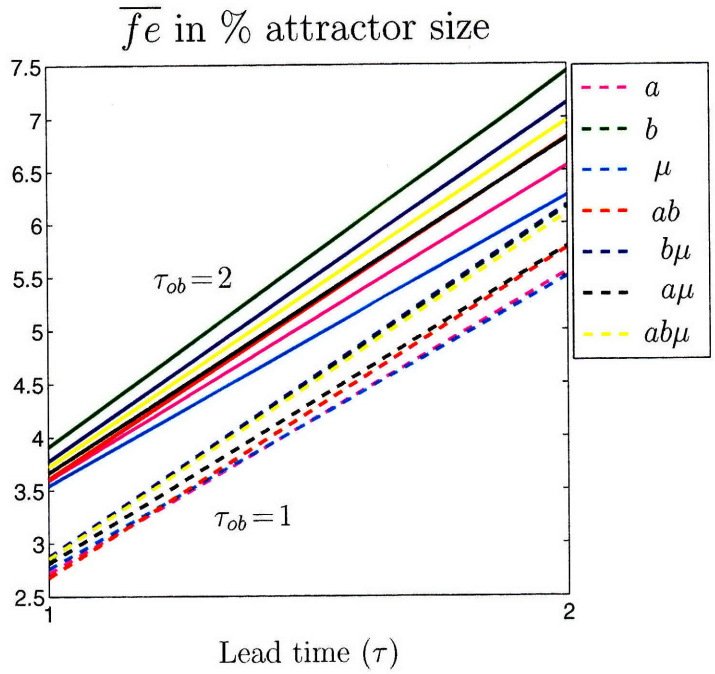


Figure 4-14: Forecast errors for SIM2 for $\tau_{ob} = 1, 2$. Each line type is for a particular τ_{ob}.

for $\tau = 1, 2$. The trends shown for $\tau = 1, 2$ in general continue for $\tau = 3, 4$, which are not shown.

For the high level of SME considered in this section, SPT outperforms DPT by a big margin, both in average analysis and forecast errors. For the low level of SME considered in section 4.3.1, SPT is marginally better than DPT filter. In the next section, the lessons learned from the parameter estimation and tuning experiments presented in sections 4.2 and 4.3 are highlighted and discussed.

4.4 Lessons from model error reduction in Ikeda

The reduction of model error in the Ikeda model was explored in sections 4.2 and 4.3. In section 4.2, the parametrically imperfect, but structurally perfect Ikeda model was considered while in section 4.3 the structurally imperfect but parametrically perfect model was considered. The important lessons learned from the Ikeda model experiments are discussed below.

When the model is parametrically imperfect, parameter estimation can reduce the model error to a large extent thus lowering the average analysis and forecast errors compared to the imperfect model. Especially, SSF performs better than DSF. This is because SSF prevents filter divergence in the parameter space. The SSF reduces the average errors in case of single parameter and multiple parameter estimation. In the case of multiple parameter estimation,

the values of parameter estimates should be interpreted carefully. This is because the error in the estimate of one parameter can be compensated by errors in the estimates of other parameters in such a way that lower average analysis errors are realized. The parameter estimates need not be close to their perfect values to give lower average analysis errors.

When the model is structurally imperfect, in general, better forecast and analysis errors result from SPT than DPT. For lower level of SME DPT performs quite well, but for higher level of SME the DPT is not able to considerably lower the analysis and forecast errors. For higher levels of SME one has to resort to SPT.

In case of lower SME, tuning the parameter a gives the lowest analysis errors followed by ab. But if the SME is higher, then tuning the parameters ab gives the lowest analysis errors. Tuning parameter b comes in a close second. For lower SME, the parameter combination that gives lowest forecast errors depends on τ_{ob} and the lead time τ. In general, tuning the combination ab gives the best forecast error.

For higher SME, tuning the parameter μ in general gives the lowest forecast errors. Tuning the parameter combination ab does well, though not the best. In general, it seems that for the SIM, irrespective of whether the SME is high or lower, tuning the parameter combination ab yields better average analysis and forecast errors. This can be explained by considering equations (4.17a), (4.17b), (4.17c) (page 106) and equation (4.18) that give the form of structural model error. The structural model error is due to the truncated series of sin and cos of θ. θ is a function of a and b as can be seen from equation (4.17c) (page 106). Therefore, tuning the combination ab is most effective in offsetting the structural model error. This also shows that the most effective parameters informs us about the way in which the model is structurally imperfect.

In general, if one has to employ DPT rather than SPT, then tuning multiple parameters is better than tuning single parameters.

There is a basic difference between the reduction of model error for PIM and SIM. For PIM, the data are used to alter the imperfect model so as to recover the perfect model. But for SIM, one knows that tuning the parameters will not alter the imperfect model to recover the perfect model. One is only trying to compensate for the SIM by tuning the parameters. In the case of the structurally imperfect Ikeda, one knows that parameter tuning is not going to recover the Ikeda model, which contains all the terms in the sine and cosine series.

Even for a low-dimensional system like Ikeda, reducing the SME poses a challenge. It is difficult to identify a priori how many and which parameters should be tuned to obtain the best average errors. Given the parameters that should be tuned, it is not possible to a priori decide the magnitude of stochastic noise to be used in the step forward equation for the

parameters.

It is to be noted that DPT can be interpreted as a global parameter estimation technique. If, say, one parameter is tuned DPT seeks a single estimate of the this parameter that offsets SME. This estimate is global in the sense that it remains the same irrespective of which region of the phase space is being visited. On the other hand, SPT can be interpreted as local or state-dependent parameter estimation. SPT allows the parameter estimate to vary with time, that is, it gives a state-dependent estimate of the parameter that offsets the SME. SPT it more successful than DPT because SPT tunes the parameter in a state-dependent manner, that is, it allows different values of the parameter depending on the region of the attractor. On the other hand, it is very difficult to determine a single value of the parameter that can successfully compensate for SME in all regions of the phase space. In DPT, the attractor could distort, giving rise to high analysis and forecast errors.

The results obtained from SIM are more relevant for application to real atmospheric models than those obtained from PIM. This is because almost all atmospheric models are both structurally and parametrically imperfect. In general, SIM overwhelms PIM.

In real atmospheric models, it is safer to assume that the SME is on the higher side rather than on the lower side. Therefore, the lessons learned from the SIM2 case are the most relevant. There are two lessons that are useful from the point of view of application to real atmospheric models. First, SPT is able to successfully compensate for high levels of model error. Second, it suffices to tune only a few number of parameters to achieve reduction in SME. As seen in experiments with SIM2, tuning a greater number of parameters need not translate into lower average errors compared to tuning a smaller number of parameters. This result is particularly important for application to real atmospheric models. In general, atmospheric models have a number of parameters. The tuning of parameters is computationally expensive. Therefore, the result that tuning only a few parameters compared to the total number of parameters can efficiently offset SME is of significance. Of course, one has to tune the most effective parameters. It is not possible to a priori choose the most effective parameters.

Similar results in regards of using SPT to offset structural model error have been obtained in the context of the Lorenz model by Hansen and Penland (2007). The implementation of SPT in an atmospheric model can be termed *stochastic parameterization.* In the traditional sense, parametrizations are used to account for unresolved sub-grid scale phenomena. But in the context of SPT, stochastic parameterization is a way to compensate for model error arising from, but not limited to unresolved sub-grid scale phenomena.

The SPT is implemented in the Navy Operational Global Atmospheric Prediction System (NOGAPS). The description of the experiments and results are presented in the section 4.5.

4.5 Parameter tuning in the NOGAPS

Real atmospheric models use a number of parametrizations to represent different subgrid scale processes. NOGAPS uses the Emanuel scheme (Emanuel, 1991; Emanuel and Zivkovic-Rothman, 1999) for parametrization of convection. NOGAPS employs the the K-theory (Louis, 1979) for vertical flux parametrization. The details of the NOGAPS model are given in Hogan and Rosmond (1991). A particular "version" of the NOGAPS model is defined to be the perfect model. Structural model error is introduced by changing a parameter in the vertical flux parametrization. The parameters in the Emanuel scheme are tuned to offset the SME.

In this work, NOGAPS is used as *Single column model* (SCM). SCM is a tool used to develop and test new methodologies, parametrizations, etc. SCM is popular because it is computationally far less expensive than a 3-dimensional model. NOGAPS has been used in the SCM mode by other researchers (Peng et al., 2004). The SCM in this work is located at 17.23 °N 280 °E. Among the many new features of the Emanuel convection scheme or parametrization, its treatment of cloud base mass flux is very important. In the next section (4.5.1), this treatment is briefly reviewed. The parameters in the Emanuel parametrization are tuned in this work.

4.5.1 The Emanuel convection scheme

In the Emanuel convection scheme, the cloud base mass flux depends on the buoyancy of the parcel at the cloud base and in the sub-cloud layer. The step forward equation for the cloud base mass flux over time $\Delta t = t_2 - t_1$ is given by

$$\mathbf{M}(t_2) = \mathbf{M}(t_1) + \Delta\mathbf{M} \tag{4.19}$$

where $\mathbf{M}$ =Cloud-base mass flux. The change in the mass flux in time Δt is given by

$$\Delta\mathbf{M} = \alpha(\Delta T_{LCL} + \Delta T_{sub} + \Delta T_k) - \lambda M(t_1) \tag{4.20}$$

where

ΔT_{LCL} = Virtual temperature difference between the parcel and the environment at the LCL.

ΔT_{sub} = Mean virtual temperature difference between the updraft and the environment in the subcloud layer.

ΔT_k = Perturbation temperature for sub-grid scale disturbance.

α = Relaxation factor.
λ = Damping factor.

Equation (4.20) relaxes $\mathbf{M}$ towards its *subcloud-layer quasi-equilibrium value.* The subcloud layer quasi-equilibrium hypothesis was proposed by Raymond (1995). This hypothesis states that, typically, the subcloud-layer tends to be in a state of balance between two competing mechanisms that increase and decrease the equivalent potential temperature of the subcloud-layer. These mechanisms are the surface fluxes and the convection, respectively. It is possible that the parcel rises above its LCL even though the buoyancy at the LCL is negative. This is because of turbulent processes near the LCL. This ability of the parcel to overcome negative buoyancy is quantified by ΔT_k. The time scale required by $\mathbf{M}$ to relax to its equilibrium value is decided by λ.

A parameter tuning technique has been used by Emanuel and Zivkovic-Rothman (1999)(hereafter EZ) to estimate the values of α and ΔT_k in a SCM. EZ used TOGA-COARE data to minimize the root-mean-square relative humidity error. The adjoint of of the linear tangent of the SCM was used to achieve this. The values estimated by EZ are $\alpha = 0.02\ kg\ m^{-2}\ {}^{\circ}\mathrm{K}^{-1}\ s^{-1}$ and $\Delta T_k = 0.65\,{}^{\circ}\mathrm{K}$. The minimization was global in nature and hence single values of α and ΔT_k were obtained.

In this work, state-dependent values of α and ΔT_k are obtained by carrying out local minimization. This distribution of values is obtained for the perfect and imperfect models. Section 4.5.2 presents the methodology and results of the parameter tuning experiment.

4.5.2 Tuning experiments and results

The NOGAPS SCM for a particular value of k_m is defined to be the perfect model. Structural model error (SME) is introduced in the vertical flux parameterization by changing the diffusion coefficient for momentum in the vertical direction (k_m).

The horizontal forcings for the SCM are provided by the output of the 3-dimensional NOGAPS model. The physics in the SCM does not feedback into these forcings. At a given time step, the SCM outputs the meteorological variable denoted by $\mathbf{p}$. $\mathbf{p}$ is a vector of variables, namely, zonal wind speed (u), meridional wind speed (v), temperature (T) and specific humidity(q). The observations of these variables for each time step are available. These observations are denoted by $\mathbf{p}_i^{obs}$. The subscript $i = 1, \ldots, 4$ denotes the component of the $\mathbf{p}$.

The cost function (CF) quantifies the mismatch between the variables and their observations

(over all model levels). The cost function(CF) is defined by

$$CF = \sum_{l=1}^{30} \sum_{i=1}^{4} \frac{(\mathbf{p}_i - \mathbf{p}_i^{obs})^2}{\sigma_{\mathbf{p}_i}^2} \tag{4.21}$$

$\sigma_{\mathbf{p}_i}^2$ is the variability in each of the variables, $i=1,\ldots,4$. l is an index that denotes the model level.

CF is minimized with respect to ΔT_k and α for every time step. The minimization is mathematically given by

$$\partial(CF)/\partial\alpha = 0 \tag{4.22a}$$

$$\partial(CF)/\partial\Delta T_k = 0 \tag{4.22b}$$

The values of α and ΔT_k that minimize CF for the perfect model case are plotted in figure 4-15(a). A distribution of values is obtained rather than a single value because the minimization is local rather than global. k_m is decreased by 50% and the state-dependent minimization is carried out. Figure 4-15(b) shows the distribution of the tuned parameters so obtained. α and ΔT_k adjust their values so as to offset the SME introduced by decreasing k_m. The values of ΔT_k tend to move towards the higher side. The mean of ΔT_k is 1.14 which is higher than that for the perfect model case. This result can be understood intuitively. The SME tends to decrease the vertical momentum flux. The value of ΔT_k then adjusts to offset this error.

Figure 4-15(c) shows the distribution when k_m is increased by 10%. The distribution shifts slightly to the lower end compared to the perfect model. Figure 4-15(d) shows the distribution when k_m is increased by 50%. The distribution clearly moves towards the lower value of ΔT_k. The mean of ΔT_k is 0.76, which is lower than that for the perfect model case. The SME tends to increase the vertical momentum flux; the values of ΔT_k tend to compensate for this increase.

The important point to note is that tuned parameters react to the level of SME. One obtains a different distribution by introducing different levels of SME on the positive and negative side of the perfect model. The result obtained in the context of Ikeda system presented in section 4.3 seems to hold in the case of a real atmospheric model.

The distribution of α and ΔT_k could be used in the forecasting phase of the NOGAPS model. This distribution can be treated as a probability distribution function from which to sample values for parameters for ensemble forecasting. The distribution corresponding to a particular imperfect model tends to offset the structural model error in the model and therefore sampling from this distribution is more useful in lowering the forecast errors than using the single fixed

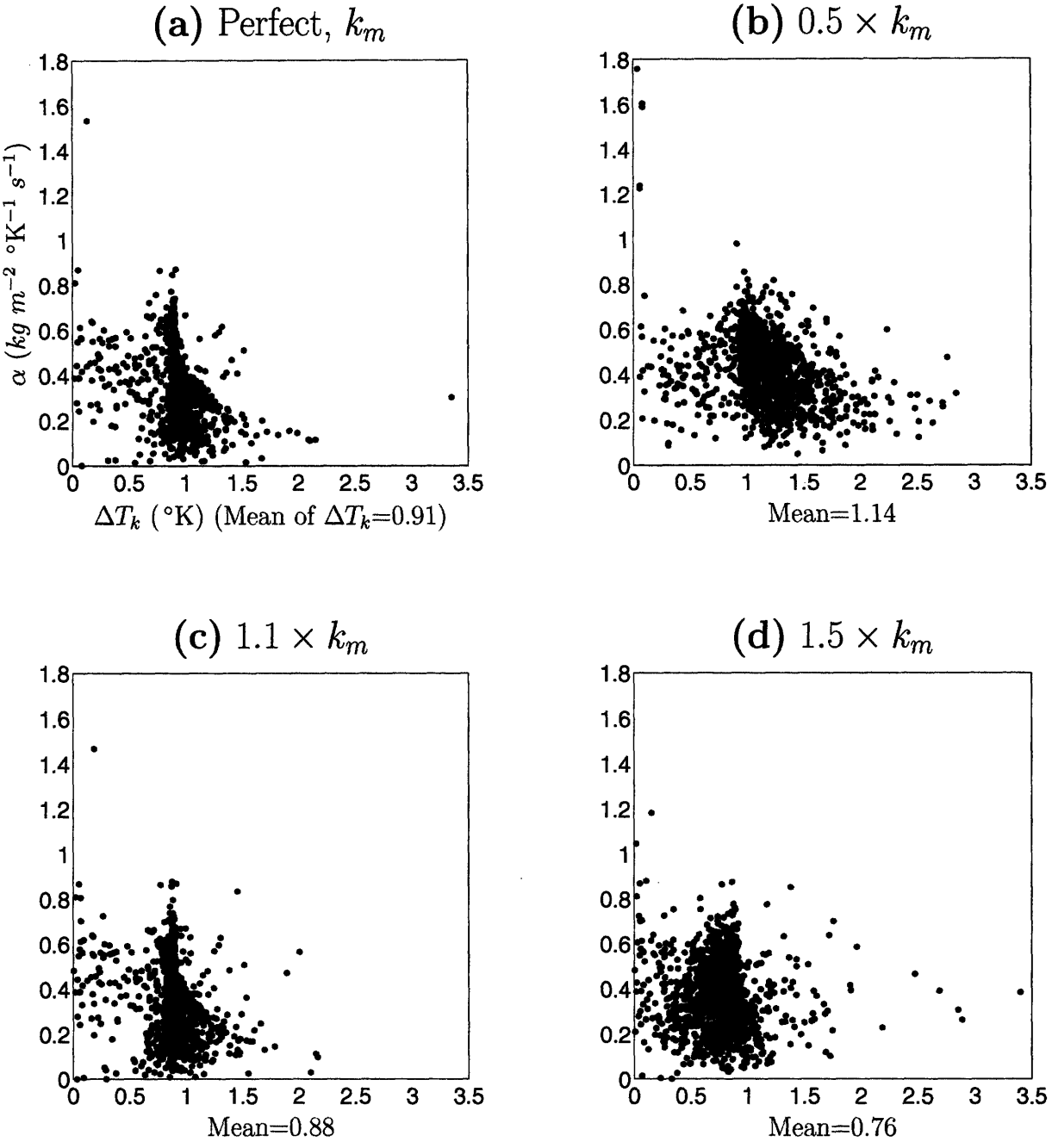


Figure 4-15: Distribution of tuned values of α and ΔT_k.

(a) Perfect model. Mean of $\Delta T_k = 0.91$.
(b) 50% decrease in k_m. Mean of $\Delta T_k > 0.91$.
(c) 10% increase in k_m. Mean of $\Delta T_k < 0.91$.
(d) 50% increase in k_m. Mean of $\Delta T_k < 0.91$.

value for each of the parameters. This work of using the parameter distribution in the forecast phase of NOGAPS is underway. This technique of state-dependent parameter tuning has been used in the context of boundary layer cloud parametrization by Golaz et al. (2007).

The next section discusses the conclusion of the reduction of model error approach as applied to the Ikeda and the NOGAPS models.

4.6 Conclusions

In this chapter, the utility of the reduction of the model error approach has been demonstrated. In the reduction of model error approach parameter(s) are treated as variable(s) and data are used to estimate or tune the parameter(s).

The experiments with low-dimensional Ikeda model are used to illuminate features of parameter estimation or tuning as a tool to reduce model error. The lessons learned from the Ikeda experiment are as follows. If the model is parametrically imperfect then the stochastic stepping forward of the parameter is more effective in reducing PME than the deterministic stepping forward. This is because the stochastic stepping forward prevents filter divergence in the parameter space. If the model is structurally imperfect then the stochastic tuning of parameters is far more effective in reducing SME than the deterministic tuning. This is because the stochastic tuning tends to find the parameter value which locally compensates for the SME. The deterministic tuning tends to find value(s) of parameters(s) that compensates for the SME globally. The stochastic tuning is more effective because it is a state-dependent tuning while deterministic tuning is state-independent tuning. The stochastic tuning gives time varying values of parameters. The parameter(s) which are a strong function of the SME should be stochastically tuned because they are most effective at reducing the SME. But it might not be possible to a priori identify these most effective parameters. The most effective parameters can be identified post-priori, that is, after different combinations of parameters are tuned and the corresponding analysis errors are determined. The most effective parameters inform us about the component of the model that has a large SME.

Parameter tuning to reduce SME is not a silver bullet that eliminates SME completely. In particular, it is difficult to retrieve the perfect model by tuning parameters in the presence of SME. The stochastic parameter tuning only seeks to compensate for SME in a state-dependent way.

This technique of stochastic parameter tuning is implemented in the NOGAPS SCM. A partic-

ular version of the NOGAPS SCM is defined as the perfect model. A structural imperfection is introduced in this model by changing the coefficient of vertical momentum flux in the parametrization of momentum flux. The parameters in the Emanuel convection scheme are tuned and it is found that the distribution of these parameters tends to offset the structural model error in the parametrization of the momentum flux. The distribution of the tuned parameter values are consistent with the physical mechanism involved in convection. If model error is introduced by decreasing k_m which is the diffusion coefficient for momentum in the vertical direction then the cloud-base mass flux decreases. The tuned values of ΔT_k increase compared to the perfect model thus increasing the cloud-base mass flux. Thus the tuned values of ΔT_k respond to the introduced model error in a way that is consistent with the physical mechanism of convection.

Chapter 5

Concluding remarks

In this chapter the work presented in chapters 2, 3 and 4 is summarized, put in perspective and consolidated. Also, possible future work is anticipated.

The broad theme of this thesis is to apply concepts from chaos theory and linear algebra to ensemble-based filtering and forecasting. The thesis focuses more on addressing the model error problem than the initial condition problem, though chapter 2 uncovers interesting aspects of the initial condition problem. This thesis does not address the problem of uncertainty in boundary conditions. The investigation in chapter 2 is from the point of view of understanding predictability in the light of the geometric structure of chaotic attractors. Chapters 3 and 4 address the handling of model error in ensemble-based filters. Chapter 2 makes the point that given a perfect model, predictability depends on the sensitivity of the local attractor to initial conditions. Chapter 3 makes the point that parametric model error structure depends on the sensitivity of the local attractor structure to the parameters.

Chapter 3 and 4 investigate two different approaches to deal with the model error problem. These two approaches are the *accounting for model error* and *reduction of model error*. Chapter 3 explores the accounting approach in the EnKF by combining the uncertainty due to the initial condition error and that due to parametric model error. The work in chapter 3 assumes that the model is parametrically imperfect and structurally perfect. The Ikeda and L63 models are used for the work in chapter 3. The MMM and PVM methods introduced in chapter 3 provide a state-dependent estimate of the model error direction. Chapter 4 investigates the the reduction of model error approach. The reduction of model error approach involves the correction of the model with the help of the data. The reduction of model error approach is applied separately to parametrically and structurally imperfect Ikeda models. The ultimate aim of this work is to improve the state estimation in real atmospheric models. The results of the reduction of model error in the Ikeda system are applied to the NOGAPS model. The

state-dependent estimation of the Emanuel parameters is able to partially offset the structural model error in the NOGAPS model.

Section 5.1 highlights the conclusions of this thesis. Section 5.2 lays out some work that could be undertaken in the future.

5.1 Conclusions

The quantification of the predictability of a given system is a problem of fundamental importance.

The Lyapunov number gives a state-independent quantification of predictability assuming a linear evolution of error. But predictability, in general, is state-dependent. Given a particular state the predictability depends on the lead time or the forecast time. The plain singular value (σ) quantifies the state-dependent, finite lead time predictability. The plain singular value assumes a linear evolution of error. σ gives the growth factor of an error aligned along the fastest growing direction at a particular state in the phase space. The initial and final time singular vectors give fastest growing initial direction and the direction this initial direction evolves to, at final time, respectively. The singular vector decomposition of the linear propagator gives the plain singular values and vectors. The variation in predictability over the model attractor as quantified by σ is studied for the 2-dimensional Ikeda system. σ shows a lot of variation with state on the Ikeda attractor. It is possible that $\sigma < 1$ for a state on the attractor implying that some regions on the attractor could have excellent predictability. The plain singular value defines predictability as the *fastest divergence of nearby trajectories.* It assumes that the initial uncertainty distribution is isotropic. The structure of this uncertainty distribution depends on the local attractor structure. This uncertainty structure, quantified by a covariance matrix, is non-isotropic, that is, the magnitude of uncertainty is different in different directions. If the non-isotropic nature of the initial uncertainty is taken into consideration, that is, the covariance is factored into the singular vector decomposition then one obtains the *normed or relevant singular values*, σ^r. The normed singular value gives the growth rate along the initial direction that evolves into the direction of maximum error at final time. σ^r is relevant to predictability in the sense that it takes into account the fact that uncertainty at initial time has different magnitudes in different directions. Therefore, the initial direction of maximum growth rate need not evolve into the direction of maximum error at final time. The predictability as given by σ^r might be more useful in certain applications, like targeted observations, in which one is typically interested in knowing the initial

directions that grow into directions of maximum forecast error at final time. The picture of predictability as given by σ^r in the case of the Ikeda attractor is drastically different from that given by σ. The picture given by σ^r shows many more regions in the Ikeda attractor where errors shrink than those given by σ. It should be noted that σ and σ^r provide two different definitions of predictability. σ implies the *fastest divergence of nearby trajectories* as the definition of predictability. σ^r implies the *divergence of nearby trajectories that results in maximum separation between those trajectories* at final time as the definition of predictability. The comparison between pictures of predictability as given by σ and σ^r is done assuming that the Ikeda model is perfect.

In reality models are imperfect. The picture of predictability computed using the imperfect model need not be the correct one. Therefore, the predictability as given by σ^r is computed for the imperfect Ikeda model. First, the parametrically imperfect Ikeda model is considered. In general, the picture of predictability for the imperfect model is different than that for the perfect model. This is because attractors corresponding to the perfect model and the imperfect models are in general different. Next, the structurally imperfect Ikeda model is considered. Structural model error is far more serious than parametric model error in rendering the model attractor different from the perfect model attractor. Consequently, the picture of predictability for the structurally imperfect model is drastically different than that for the perfect model.

The different pictures of predictability going from the Lyapunov number to σ^r for the structurally imperfect model give a cascade of predictability pictures going from idealistic to realistic scenarios. The Lyapunov number assumes that error growth rate is the same at all states on the attractor. The plain singular value is more realistic than the Lyapunov number in that it gives the state-dependent predictability. But σ assumes that the initial uncertainty distribution is isotropic. The picture given by σ^r is even more realistic because it takes into consideration the non-isotropic structure of the initial uncertainty. The most realistic picture is given by the σ^r calculated for the imperfect model.

The final time error in prediction is due to both initial condition error and model error. The fraction of error in the total final time error due to initial condition error and parametric model error is calculated in the Ikeda system. The square of the magnitude of the final time error is equal to the sum of the squares of the magnitudes of the state error, the model error and the interaction error. The interaction error term contains products of state and model error. The variation of these fractions over the Ikeda attractor is studied. It is found that the model error could be as important in rendering the forecast imperfect as the initial condition error.

The problem of model error is addressed in ensemble-based filtering and forecasting with two different approaches. In the *accounting approach*, it is assumed that the uncertainty due to

initial condition error can be linearly combined with that due to model error. This assumption leads to the Additive Error Approximation (AEA). The accounting approach is investigated for the parametrically imperfect Ikeda and L63 models. The state-dependent initial condition uncertainty is quantified by the sensitivity of the local attractor to a perturbation in the initial state. Analogously, the state-dependent uncertainty due to model error is quantified by the sensitivity of the local attractor to the parameter. The MultiModel Method (MMM) estimates the state-dependent model error by employing multiple ensembles corresponding to different parameters. Each of the multimodel ensembles tends to lie on the attractor characterized by the parameter value used by the particular ensemble. The dispersion across the multimodel ensembles is the MMM estimate of the model error. The Parametric Singular Vector Method (PSVM) is a linear approach to obtain the model error estimate given by MMM. Parametric singular vector gives the sensitivity of the state to a perturbation in the parameter(s). The sensitivity of the state to the parameter given by the parametric vector may not be the sensitivity of the attractor to the parameter. This is because the PSVM steps forward the same state using two different values of the parameter assuming that the parametric perturbation is small enough, that is, in the linear limit. However, this perturbation in the parameter, no matter how small, induces a perturbation in the state that evolves with time. This perturbation in state is called the model-induced state error. Consequently, the parametric sensitivity contains signals due to both the parametric error and due to the model-induced state error. For a given state, the model error signal undergoes a transient and simultaneously the model-induced state error evolves. After a particular time called the optimal time, the model error signal dominates the model-induced state error. At this particular time, the parametric vector gives the correct sensitivity, that is, the sensitivity of the attractor to the parameter. This is because the model error manifests itself maximally in the parametric vector. After the optimal time, the model-induced state error overwhelms the model error signal and therefore the parametric vector tends to point in the local attractor direction. The phase after the optimal time is called the post-optimal phase. The phase before the optimal time when the model error undergoes a transient is called the pre-optimal phase. The model error estimate given by the parametric vector in the pre-optimal and post-optimal phase is incorrect in the sense that it does not give the sensitivity of the attractor to the parameter. The parametric vector at the optimal time gives the correct estimate of the model error. The optimal time is state-dependent. The criterion to calculate the state-dependent optimal time is not known. The average optimal time is calculated by comparing the parametric vectors with the multimodel vectors. The average optimal time for the Ikeda system is 3 steps, which is about 1.5 times the error doubling time of the Ikeda system.

The MMM and PVM are two methods to obtain a state-dependent estimate of model error. The SQM (Static **Q** Method) is an (offline) method to obtain a state-independent estimate of model error. The MMM, PVM and SQM are compared within the AEA framework for the

Ikeda model. The MMM gives the least average errors because it provides a state-dependent estimate of model error structure. The PVM closely approximates the performance of MMM. The SQM is worse than MMM and PVM. This is because SQM is a state-independent estimate of the model error while MMM and PVM are state-dependent estimates. Similar results are obtained for the L63 system. The average optimal time for the L63 system is about 2.5 times its error doubling time.

The *reduction* of model error is another approach to deal with the model error problem. In the reduction approach the data are used to alter the imperfect model so that the model error is decreased. The estimation of a parameter is a typical example of reduction of model error. The reduction of model error approach is applied to the Ikeda system. First, the parametrically imperfect Ikeda model is considered. The reduction of model error involves augmenting the state with the parameter. The deterministic step forward (DSF) equation or the stochastic step forward (SSF) could be used to propagate the parameter in time. The observations are of state only, that is, the parameter is not observed. An ensemble in the parameter space is constructed in addition to the ensemble in the state space. The observations are assimilated into the augmented ensemble. The covariance between the state and parameter corrects the parameter. If the DSF is used for the parameter then ensemble members in the parameter space tend to collapse to the same value. That is, there is filter divergence in the parameter space. In this case the estimated value of the parameter does not converge to the perfect parameter value. The filter divergence is prevented by using the SSF. The magnitude of the noise to be used in the SSF is not known a priori. Therefore, experiments with different value of the stochastic noise are performed. It is found that SSF gives lower average analysis and forecast errors than DSF. Both DSF and SSF give lower average error than the imperfect model. The SSF parameter estimate is closer to the perfect parameter value than is the DSF estimate. These results hold true irrespective of which of the three parameters in the Ikeda model are imperfect. In the parametrically imperfect Ikeda model, it is found that parameter estimation is able to reduce the average analysis and forecast errors by about 1-40%, depending on τ_{ob} and the imperfect parameter. It is possible to tune more than one parameters. The experiments of multiple parameter estimation show that SSF gives lower average analysis and forecast errors than DSF. The SSF parameter estimates are closer to the perfect parameter values than are the DSF estimates. It is possible that the error in estimate of one parameter compensates for that in another parameter in such a way that the error in the state is decreased. Therefore, a lower average analysis error may not mean that the estimates of the parameters are close to the perfect parameter values.

The reduction of model error approach is applied to the structurally imperfect Ikeda model. The structural imperfection is introduced in the Ikeda model by replacing sine and cosine functions by their truncated series. The parameter estimation in the presence of the structural

model error is termed *parameter tuning*. Parameter tuning attempts to offset the effect of the structural model error (SME). Parameter tuning done by using the DSF is called the deterministic parameter tuning (DPT) and that done by using SSF is called the stochastic parameter tuning (SPT). The experiments of parameter tuning are performed for a low level of SME and high level of SME. The SPT decreases the average analysis error compared to the structurally imperfect model. For both low and high level of SME, SPT gives lower average analysis errors than the DPT. The DPT tries to find state-independent values of parameters that compensate for the SME. That is, the DPT tries to find global values for parameters that offset the SME. On the other hand, the SPT finds the values of parameters that compensate for the SME locally. SPT is more successful that the DPT because SPT is a state-dependent parameter tuning. In the case of high level of SME the DPT, in its attempt to find a global parameter value to compensate for SME could result in an estimate that distorts the attractor leading to high analysis and forecast errors. On the other hand, SPT does not lead to the distortion of the attractor because it seeks a local, rather than global parameter estimate. The parameter tuning experiments show that out of the three parameters in the Ikeda model, tuning a and b performs best in compensating for the SME. This is because the SME introduced by truncating the sine and cosine series is a strong function of a and b. In the case of the Ikeda model, for this particular SME, a and b are the most effective parameters.

The conclusions drawn from the reduction of model error experiments in the Ikeda model have a great potential for application to real atmospheric models. The results of the reduction of SME are more pertinent than those of reduction of PME. This is because real atmospheric models are structurally imperfect. The lessons learned from parameter tuning in the Ikeda model are applied to the NOGAPS model. The NOGAPS model employs the Emanuel convection scheme. The Emanuel convection scheme relaxes the cloud-base mass flux towards its subcloud-layer quasi-equilibrium value. The time evolution of the subcloud-layer quasi-equilibrium is a function of the relaxation factor (α) and of a perturbation temperature (ΔT_k) representing turbulent processes near the LCL. Structural model error is introduced in the NOGAPS model by changing the vertical momentum flux coefficient (k_m). The Emanuel parameters (α and ΔT_k) are tuned in the NOGAPS model to offset the structural model error. This tuning is state-dependent, so that one obtains a distribution of the Emanuel parameters rather than a fixed value for the parameters. This distribution responds to different magnitudes of structural model error in NOGAPS induced by changing k_m. For the structurally perfect NOGAPS model, the mean of ΔT_k is 0.91 °K. The structural model error induced by decreasing and increasing k_m by 50% shifts the mean of ΔT_k to 1.14 °K and 0.76 °K respectively. The tuned values of ΔT_k are consistent with the physics of the structural model error in the NOGAPS model. The reduction in k_m decreases the cloud-base mass flux. The parameter tuning responds to this model error by increasing ΔT_k, thereby increasing the cloud-base mass flux. On the other hand an increase is k_m decreases ΔT_k thereby decreasing the cloud-base

mass flux. Thus the parameter tuning is able to successfully, even though partially offset structural model error in the NOGAPS model.

It should be noted that parameter tuning only mitigates the SME rather than completely eliminating it. Parameter tuning is not a panacea that can make a structurally imperfect model to become perfect. Rather it is a tool that can be usefully employed to inform us about which components of a given model have large structural errors. Then the appropriate or most effective parameters can be tuned to partially offset the SME. The application of results from Ikeda to NOGAPS is an illustration of how the results from investigating low-dimensional systems can be successfully employed in high-dimensional atmospheric models.

Of course, the work done in this thesis raises many interesting questions. The next section outlines some possible work that could be undertaken in the future.

5.2 Further work

There is a large scope for further work in both the accounting for model error and the reduction of model error approaches.

The most important piece of work that could make the PVM technique applicable to real atmospheric models is the identification of a criterion for the state-dependent optimal time (τ_{om}). In this thesis the average τ_{om} is calculated with the help of the MMM and this average τ_{om} is used to illustrate the PVM. The PVM can be practically useful if τ_{om} can be calculated without using MMM. τ_{om} is basically the *transient time* with respect to the parameter. The *optimal time* can be called the *parametric transient time.* A starting point to identify the criterion for state-dependent optimal time is to identify the criterion for state-dependent transient time. In theory, transient time is infinity because the off-attractor point approaches the attractor asymptotically. In practice, the point can be said to be on the attractor when it comes within the δ neighborhood of the attractor. It appears that formulating a criterion for state dependent transient time would be a fundamental result in chaotic and applied chaotic dynamics. In this respect, workers in the field of ensemble-based filtering are placed at a vantage point because they have access to the local attractor structure.

In the reduction of model error approach, the distribution of Emanuel parameters can be used to sample parameter values for ensemble-based forecasting. The 3-dimensional NOGAPS model could be used for this purpose. This is a major undertaking that could ascertain the significance of the stochastic parametrization in reducing forecast errors in ensemble-based

forecasting.

The accounting and reduction approaches are investigated separately, which may be misleading because these approaches are not mutually exclusive. They have been considered separately for simplicity, so that it would be easier to understand the science and engineering involved in each approach. Hopefully, these approaches could be combined and used simultaneously to mitigate the model error problem to a greater extent than that using either of them in isolation. It might even be possible to combine more than one technique of accounting for model error. For example, model output statistics from multimodel ensembles could be exploited and model error could be accounted for in the EnKF.

Bibliography

Aksoy, A., F. Zhang, and J. Nielsen-Gammon, 2006a: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea breeze model. *Mon. Wea. Rev.*, **134**, 2951–2970.

— 2006b: Ensemble-based simultaneous state and parameter estimation with MM5. *Geophysical Research Letters*, **33**, L12801.

Anderson, J., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2902.

Anderson, J. and S. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2748.

Annan, J. and J. Hargreaves, 2004: Efficient parameter estimation for highly chaotic systems. *Tellus*, **56A**, 520–526.

Annan, J., J. Hargreaves, N. Edwards, and R. Marsh, 2005a: Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modelling*, **8**, 135–154.

Annan, J., D. Lunt, J. Hargreaves, and P. Valdes, 2005b: Parameter estimation in an atmospheric GCM using the ensemble Kalman filter. *Nonlin. Processes Geophys.*, **12**, 363–371.

Barkmeijer, J., M. VanGijzen, and F. Bouttier, 1998: Singular vectors and estimates of the analysis-error covariance metric. *Quart. J. Roy. Meteor. Soc.*, **124**, 1695–1713.

Burgers, G., P. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Derber, J., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437–2446.

Ehrendorfer, M. and J. Tribbia, 1995: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **54**, 286–313.

Emanuel, K., 1991: A scheme for representing cumulus convection in large-scale models. *J. Atmos. Sci.*, **48**, 2313–2335.

Emanuel, K. and M. Zivkovic-Rothman, 1999: Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.*, **56**, 1766–1782.

Epstein, E., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739–759.

Evensen, G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, **97**, 17905–17924.

— 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99(C5)**, 10143–10162.

— 2005: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, **53**, 343–367.

— 2006: *Data assimilation: The Ensemble Kalman Filter*. Springer.

Fisher, R., 1912: On an absolute criterion for fitting frequency curves. *Messenger of Math*, **41**, 155–160.

Gauss, C., 1857: *Theory of the motion of the heavenly bodies about the sun in conic sections*. Dover.

Gelb, A., 1974: *Applied optimal estimation*. MIT press.

Ghil, M. and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Adv. in Geophys.*, **133**, 141–266.

Golaz, J., V. Larson, J. Hansen, S. D., and G. B., 2007: Elucidating model inadequacies in a cloud parameterization by use of an ensemble-based calibration framework. *Mon. Wea. Rev.*, **135**, 4077–4096.

Hamill, T. and J. 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.

Hansen, J., 2002: Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Wea. Rev.*, **130**, 2373–2391.

Hansen, J. and K. Emanuel, 2002: Forecast 4d-Var : Exploiting model output statistics. *Quart. J. Roy. Meteor. Soc.*, **129**, 1255–1267.

Hansen, J. and C. Penland, 2006: Efficient approximate techniques for integrating stochastic differential equations. *Mon. Wea. Rev*, **134**, 3006–3014.

— 2007: On stochastic parameter estimation using data assimilation. *Physica D*, **230**, 88–98.

Hansen, J. and L. Smith, 2001: Probabilistic noise reduction. *Tellus*, **53A**, 585–598.

Hogan, T. and T. Rosmond, 1991: The description of the navy operational global atmospheric prediction system's spectral forecast model. *Mon. Wea. Rev.*, **125**, 1786–1815.

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

— 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289.

Houtekamer, P., H. Mitchell, G. Pellerin, B. M., 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.

Ikeda, K., 1979: Multiple-valued stationary state and the instability of the transmitted light by a ring cavity. *Opt. Commun.*, **30**, 257–261.

Kalman, R., 1960: A new approach to linear filtering and prediction problems. *J. Basic. Eng.*, **82**, 35–45.

Khade, V. and J. Hansen, 2004: State dependent predictability: Impact of uncertainty dynamics, uncertainty structure and model inadequacies. *Nonlin. Processes in Geophys.*, **11**, 351–362.

Kolmogorov, A., 1941: Interpolation and extrapolation von stationaren zulfalligen folgen. *Bull. Acad. Sci. Ser. Math.*, **5**, 3–14.

Krishnamurti, T., C. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, and E. Williford, 2000: Multimodel ensemble forecasts for weather and seaonal climate. *J. of Climate*, **13**, 4196–4216.

Le Dimet, F. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, **38A**, 97–110.

Lea, D., , T. Haine, M. Allen, and J. Hansen, 2002: Sensitivity analysis of the climate of a chaotic ocean circulation model. *Quart. J. Roy. Meteor. Soc.*, **128**, 2587–2605.

Lea, D., M. Allen, and T. Haine, 2000: Sensitivity analysis of the climate of a chaotic system. *Tellus*, **52A**, 523–532.

Leith, C., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Lipschutz, S., 1991: *Linear algebra*. McGraw-Hill.

Lorenz, E., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

— 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321–333.

Louis, J., 1979: A parametric model of vertical eddy fluxes in the atmosphere. *Boundary Layer Meteorology*, **17**, 187–202.

McLaughlin, D., 1995: Recent developments in hydrologic data assimilation. *Rev. of Geophys.*, **33**, 977–984.

— 2002: An integrated approach to hydrologic data assimilation: Interpolation, smoothing and forecasting. *Advances in Water Resources*, **25**, 1275–1286.

Menemenlis, D., I. Fukumori, and T. Lee, 2005: Using Green's functions to calibrate an ocean general circulation model. *Mon. Wea. Rev.*, **133**, 1224–1240.

Molteni, F., R. Buizza, T. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system : Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.

Orrell, D., L. Smith, J. Barkmeijer, and T. Palmer, 2001: Model error in weather forecasting. *Non. Processes in Geophys.*, **8**, 357–371.

Oseledec, T., 1968: A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, **19**, 197–231.

Palmer, T., 1994: Singular vectors and the predictability of weather and climate. *Philosophical Transactions of The Royal Society of London A*, **348**, 459–475.

— 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304.

Palmer, T., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653.

Palmer, T. and R. Hagedorn, 2006: *Predictability of weather and climate.* Cambridge university press, 128-143 pp.

Peng, M., J. Ridout, and T. Hogan, 2004: Recent modifications of the Emanuel convective scheme in the navy operational global atmoshpheric prediction system. *Mon. Wea. Rev.*, **132**, 1254–1268.

Penland, C., 2003: A stochastic approach to nonlinear dynamics : a review. *Bull. Amer. Meteor. Soc.*, **84**, 921–925.

Raymond, D., 1995: Regulation of moist convection over west Pacific warm pool. *J. Atmos. Sci.*, **52**, 3945–3959.

Reichle, R., D. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103–114.

Smith, L., C. Ziehmann, and K. Fraedich, 1999: Uncertainty dynamics and predictabiity in chaotic systems. *Quart. J. Roy. Meteor. Soc.*, **125**, 2855–2886.

Strang, G., 1988: *Linear algebra and its Applications.* Penguin.

Talagrand, O., 1997: Assimilation of observations, an introduction. *J. Meteor. Soc. Japan*, **75**, 191–209.

Thompson, P., 1957: Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, **9**, 275–295.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC : The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Wiener, N., 1949: *The exptrapolation,interpolation and smoothing of stationary time series.* John Wiley and Sons.

Wunsch, C., 1996: *The ocean circulation inverse problem.* Cambridge University Press.

Ziehmann, C., L.A.Smith, and J. Kurths, 1998: The bootstrap and Lyapunov exponents in deterministic chaos. *Physica.D*, **126**, 49–59.