# Adaptive error estimation in linearized ocean general circulation models
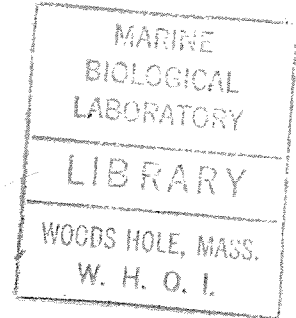
by

## Michael Y. Chechelnitsky

M.Sc. Massachusetts Institute of Technology
Woods Hole Oceanographic Institution
(1996)
B.Sc. Upsala College
(1993)

Submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

June 1999

© Michael Y. Chechelnitsky, 1999

Signature of Author......................................................
Joint Program in Physical Oceanography
Massachusetts Institute of Technology
Woods Hole Oceanographic Institution
February 8, 1999

Certified by.................................................
Carl Wunsch
Cecil and Ida Green Professor of Physical Oceanography
Massachusetts Institute of Technology
Thesis Supervisor

Accepted by.................................................
W. Brechner Owens
Chairman, Joint Committee for Physical Oceanography
Massachusetts Institute of Technology
Woods Hole Oceanographic Institution

# Adaptive error estimation in linearized ocean general circulation models

by

Michael Y. Chechelnitsky

## Abstract

Data assimilation methods, such as the Kalman filter, are routinely used in oceanography. The statistics of the model and measurement errors need to be specified *a priori*. In this study we address the problem of estimating model and measurement error statistics from observations. We start by testing the Myers and Tapley (1976, MT) method of adaptive error estimation with low-dimensional models. We then apply the MT method in the North Pacific ($5°-60°$ N, $132°-252°$ E) to TOPEX/POSEIDON sea level anomaly data, acoustic tomography data from the ATOC project, and the MIT General Circulation Model (GCM). A reduced state linear model that describes large scale internal (baroclinic) error dynamics is used. The MT method, closely related to the maximum-likelihood methods of Belanger (1974) and Dee (1995), is shown to be sensitive to the initial guess for the error statistics and the type of observations. It does not provide information about the uncertainty of the estimates nor does it provide information about which structures of the error statistics can be estimated and which cannot.

A new off-line approach is developed, the covariance matching approach (CMA), where covariance matrices of model-data residuals are "matched" to their theoretical expectations using familiar least squares methods. This method uses observations directly instead of the innovations sequence and is shown to be related to the MT method and the method of Fu et al. (1993). The CMA is both a powerful diagnostic tool for addressing theoretical questions and an efficient estimator for real data assimilation studies. It can be extended to estimate other statistics of the errors, trends, annual cycles, etc.

Twin experiments using the same linearized MIT GCM suggest that altimetric data are ill-suited to the estimation of internal GCM errors, but that such estimates can in theory be obtained using acoustic data. After removal of trends and annual cycles, the low frequency/wavenumber (periods > 2 months, wavelengths > $16°$) TOPEX/POSEIDON sea level anomaly is of the order 6 cm$^2$. The GCM explains about 40% of that variance. By covariance matching, it is estimated that 60% of the GCM-TOPEX/POSEIDON residual variance is consistent with the reduced state linear model.

The CMA is then applied to TOPEX/POSEIDON sea level anomaly data and a linearization of a global GFDL GCM. The linearization, done in Fukumori et al.(1999), uses two vertical mode, the barotropic and the first baroclinic modes. We show that

the CMA method can be used with a global model and a global data set, and that the estimates of the error statistics are robust. We show that the fraction of the GCM-TOPEX/POSEIDON residual variance explained by the model error is larger than that derived in Fukumori et al.(1999) with the method of Fu et al.(1993). Most of the model error is explained by the barotropic mode. However, we find that impact of the change in the error statistics on the data assimilation estimates is very small. This is explained by the large representation error, i.e. the dominance of the mesoscale eddies in the T/P signal, which are not part of the $2°$ by $1°$ GCM. Therefore, the impact of the observations on the assimilation is very small even after the adjustment of the error statistics.

This work demonstrates that simultaneous estimation of the model and measurement error statistics for data assimilation with global ocean data sets and linearized GCMs is possible. However, the error covariance estimation problem is in general highly underdetermined, much more so than the state estimation problem. In other words there exist a very large number of statistical models that can be made consistent with the available data. Therefore, methods for obtaining quantitative error estimates, powerful though they may be, cannot replace physical insight. Used in the right context, as a tool for guiding the choice of a small number of model error parameters, covariance matching can be a useful addition to the repertory of tools available to oceanographers.

Thesis Supervisor: Carl Wunsch,
Cecil and Ida Green Professor of Physical Oceanography,
Department of Earth, Atmospheric, and Planetary Sciences,
Massachusetts Institute of Technology

# Acknowledgments

I would like to thank my thesis advisor Carl Wunsch for pointing out the subject of the thesis and giving me the freedom to pursue it independently. I also very much appreciate several earlier research projects suggested by Carl, which although seemingly unrelated, all proved beneficial in thinking about the thesis problem. Learning to think in physical rather than mathematical terms has proven very difficult, and I constantly referred to Carl's work for guidance. In addition to my advisor, Dimitris Menemenlis has been exceedingly generous with his time, guidance and advice. Dimitris interest in both theoretical and practical aspects has helped me to focus my attention on real-world systems; his comments on many parts of the thesis and the code provided the encouragement when I needed it most. His questions kept me pondering late at night, and I regret that I have not been able to answer so many of them. I would also like to thank Ichiro Fukumori for giving me access to a early draft of the paper, allowing me to use his data assimilation code, and helping me with Fortran. I am indebted to Alexey Kaplan of Lamont for very lengthy discussions of the most technical aspects of this work. My thesis committee, Nelson Hogg, Paola Malanotte-Rizzoli, Jochem Marotzke, and Breck Owens, helped me to stay on track, and I would like to thank them their advice, criticisms, and comments.

Several people made it possible to complete extensive calculations required in the thesis. Chris Hill patiently answered my numerous questions on UNIX and Fortran. Parry Husbands, the author of remarkable MITMatlab, provided unparalleled online support. Charmaine King and Linda Meinke were very helpful with data and systems support.

François Primeau, always interested in learning and talking about any new subject, has been a very special friend. I cannot thank him enough for his efforts and time spent reading several drafts of this manuscript. Dmitri Ivanov helped me with theoretical proofs in Chapter 3. My fellow Joint Program students, especially Misha Solovev, Richard Wardle, Markus Jochum, Alex Ganachaud, Natalia Beliakova, and Vikas Bhushan, made life at MIT/WHOI more enjoyable. I thank Gerard and Gavin, my two formidable tennis partners, for careful reading of the early drafts even when they "did not understand an iota".

John Snygg taught me many important things, among them that "you can make a right decision but for a wrong reason". Shura and my friends, Sergey Cherkis, Kostik Evdokimov, Theo Korzukhin, Tonia Lifshits, Asya Lyass, Lada Mamedova, Anya and Lenya Mirny, Nikita Nekrasov, Lenya Peshkin, Boria Reizis, Lyuba and Misha Roitman, shared their thoughts in many different places over various kinds of drinks. My parents, my brother, and my parents-in-law, whom I have seen too little, gave me boundless love. And Natasha, my dear wife and best friend, has made this all worthwhile.

# Contents

9

# List of Figures

11

12

13

# List of Tables

# Chapter 1

# Introduction

## 1.1   Data Assimilation in Oceanography

In recent years it has become clear that to understand human induced climate change we first need to understand the natural variability of the world climate. The world ocean is one of the parts of the climate system which we understand least. The spatial scales of the large scale ocean circulation are grand, and the intrinsic time scales are very long. To date, the dynamics of this enormous physical system have been grossly undersampled. Observations in the ocean are very difficult and very expensive to make. Laboratory experiments are useful, but limited to idealized problems. General circulation ocean models (GCMs) provide numerical solutions to the physically relevant set of partial differential equations (PDEs). They are routinely used to study ocean dynamics. However, GCMs are very complicated and often have to be run at very coarse spatial and temporal resolution. The models are imperfect as the equations are discretized, the forcing fields are noisy, the parameterization of sub-grid scale physics is poorly known, etc. Therefore, to study the dynamics one needs to combine models and observations, in what is known as data assimilation, or inverse modeling. The subject of inverse modeling deals with various techniques for solving under-determined problems, and is well established in many fields, e.g. solid-earth geophysics. Wunsch (1996) provides a general treatment of the

inverse theory applicable to the oceanographic problems.

The process of data assimilation can be viewed from two different perspectives. On the one hand, it filters the data by retaining only that part which is consistent with a chosen physical model. This is a "filter" in the sense of more familiar frequency filters, e.g. low-pass filters which eliminate high frequency oscillations. On the other hand, it constrains the model by requiring that the state of the model is in agreement with the observations. That is, we use the data as constraints for the models and then use the model to provide information about the regions, or fields, for which we have no observations.

In oceanography data assimilation has three main objectives, as described in detail in an overview of Malanotte-Rizzoli and Tziperman (1996). Using data assimilation for dynamical interpolation/extrapolation to propagate information to regions and times which are void of data has been one the primary goals of inverse modeling in oceanography. For example, the TOPEX/POSEIDON altimeter measures the height of the sea surface relative to the geoid. Using data assimilation, altimetric data can be used to constrain an ocean GCM, and then the output of the GCM provides information about the dynamics of the ocean interior, e.g. Stammer and Wunsch (1996). By using a model to extrapolate sea surface height measurements into temperature one can estimate meridional heat transport across various latitudes, see Figure 1.1 reproduced from Stammer et al. (1997). Traditionally, one would require sending a ship measuring temperature and density profiles across the ocean at all those locations, e.g. Macdonald and Wunsch (1996). Macdonald and Wunsch (1996) used hydrographic data to obtain such estimates at several latitudes where zonal sections were available, shown as open circles on Figure 1.1a. By combining dynamical models with observations one can obtain a global time-dependent picture of the ocean circulation.

In contrast to altimetric measurements which provide information about the sea surface, acoustic tomography samples the ocean interior by transmitting sound pulses from a source to a receiver along multiple paths, Munk et al. (1995). To first order, the sound
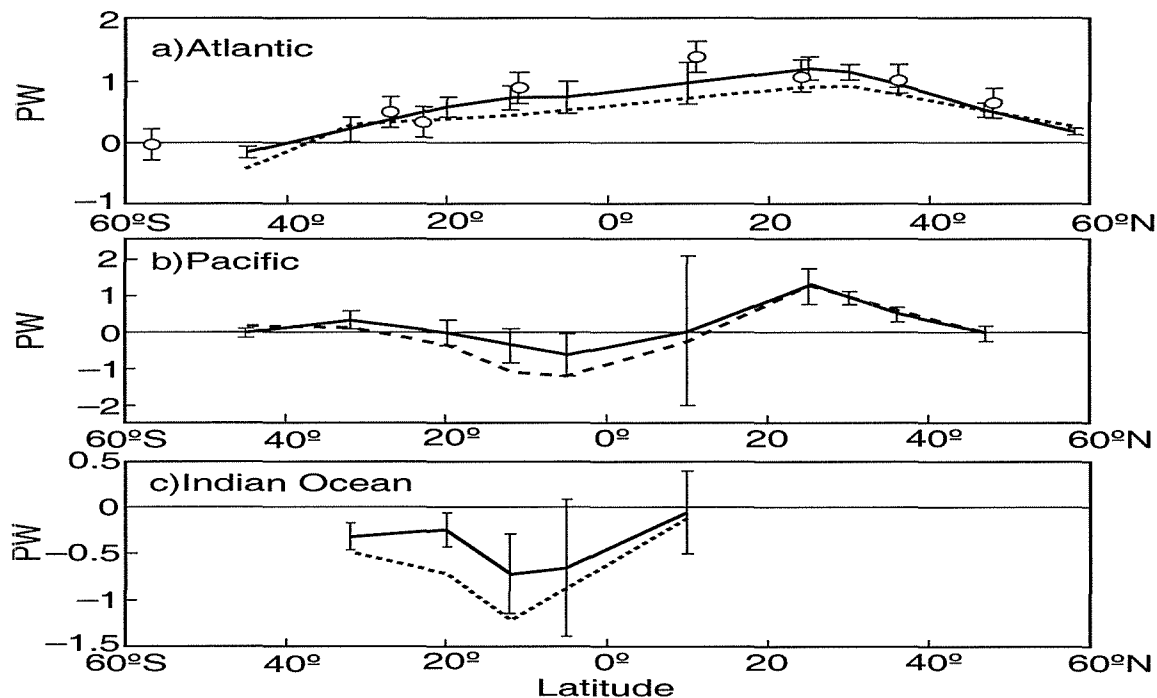
Figure 1.1: Meridional heat transport (in $10^{15}$W) for July-December 1993 estimated from global data assimilation with the MIT model and T/P data (solid lines). The unconstrained model for the Atlantic, the Pacific and the Indian Ocean respectively, are shown with the dashed lines. Bars on the solid lines show RMS variability of the transport estimated over individual 10-day periods (reproduced from Stammer et al. 1997).

speed depends on temperature along the path of the acoustic ray. This temperature information can be then inverted to obtain velocities, displacements, or other physical quantities which can also be estimated from the model output, e.g. Menemenlis et al. (1997b). An example of this is shown on Figure 1.2, where an estimate of depth-averaged temperature and horizontal velocity at a particular vertical level were obtained for the western Mediterranean during beginning of spring, summer and autumn of 1994. Thus, in principle a GCM can be used to extrapolate tomographic data to other areas of the global ocean.

Thirdly, data assimilation can also be used to study dynamical processes by way of improving our understanding of ocean models. Even the most complex ocean GCMs cannot resolve all the dynamically important physical processes, and some processes
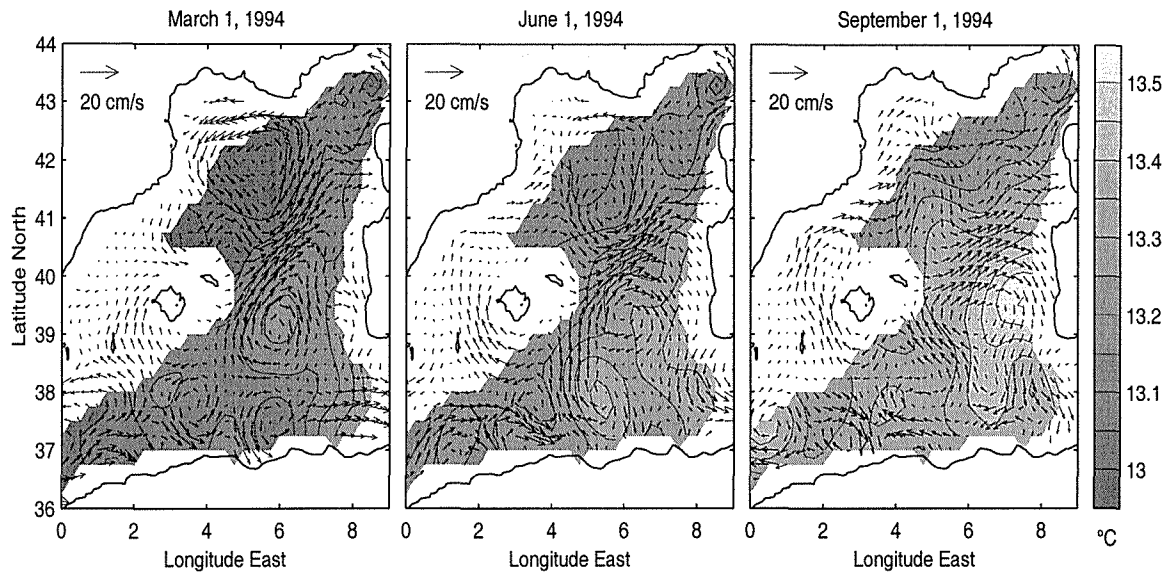
Figure 1.2: Estimates of the western Mediterranean circulation, on March 1, June 1, and September 1, 1994 obtained by combining tomographic and altimetric observations with a GCM. The grey scale shading indicates depth-averaged (0-2000m) potential temperature and the arrows indicate horizontal velocity at 40m (reproduced from Menemenlis et al. 1997).

have to be parameterized. These parameterizations can be simple or quite complex, but are always uncertain in both form and in value of parameters. A few examples of such parameterizations are small scale vertical mixing schemes in the boundary layer, Large et al. (1994), parameterizations of mesoscale eddies in coarse ocean models, Boning et al., (1995), and deep water formation, Visbeck et al. (1996). Observations of most of the above unknown parameters are not available, and many of these parameters cannot be measured directly. They instead can be estimated by using other available data through data assimilation.

Data assimilation can be used for prediction by providing the best estimate of the initial conditions, which are then propagated forward by a model. Forecasting the oceanic fields has become much more important in recent years, and is now done for ENSO on a routine basis. The importance of good initialization fields is clear from a much publicized failure of the Zebiak and Cane model (1987), in predicting the El Niño event of 1997,

## LAMONT MODEL FORECAST OF NINO3
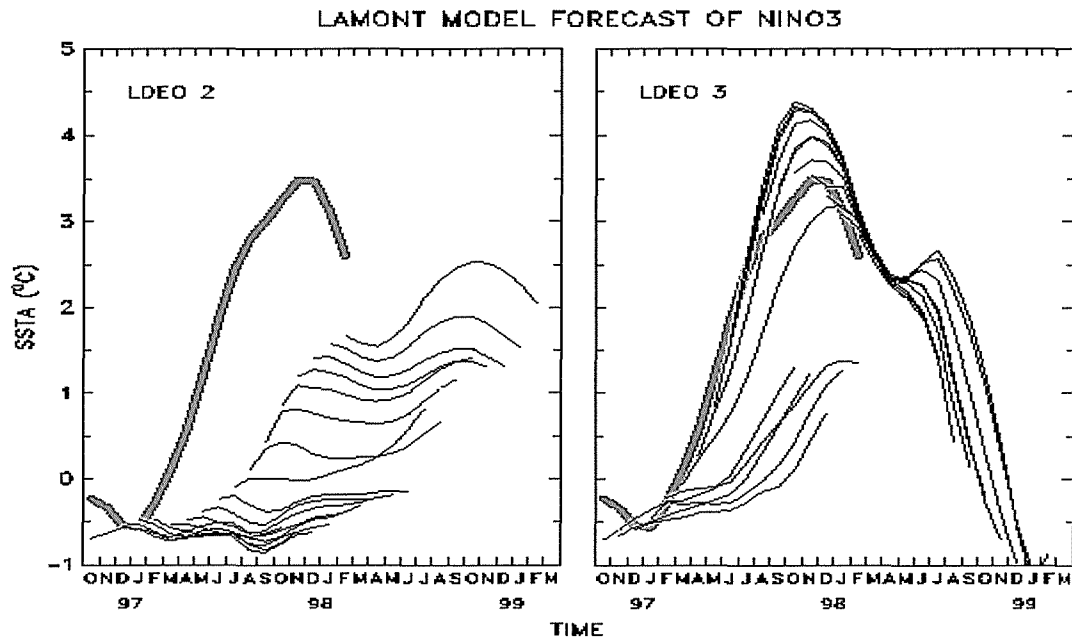
Figure 1.3: Lamont model forecasts of the 1997/1998 El Niño with (right) or without (left) sea level data assimilation. The thick curve is observed NINO3 SST anomaly. Each curve is the trajectory of a 12 month forecast starting from the middle of each month.(reproduced from Chen et al. 1998)

e.g. Kerr (1998). Recent analysis showed that with a different data assimilation scheme, and accordingly a different initial field, the model did a much better job of predicting that year's ENSO event, Chen et al. (1998). Figure 1.3 shows striking difference between two groups, LDEO2 and LDEO3, of Zebiak and Cane model, 12 month forecasts. The positive impact of the sea level data assimilation used in LDEO3 but not in LDEO2 is obvious.

While the theory of inverse modeling, including error estimation, has been well developed by the control engineering community, applying it to oceanographic problems is still a challenge. There are at least three major obstacles: computational burden, memory requirements, and lack of required information. That is, problems encountered in control applications typically have a small number of degrees of freedom, $O(10)$, and long time

20

series measurements of most, if not all, degrees of freedom. The problems encountered in oceanography tend to be of the opposite nature, i.e. very large number of degrees of freedom, and very short and spatially sparse time series of observations. This difference in the problem size and the amount of available data makes "adoption" of engineering methods to oceanography a difficult exercise. Many properties of the estimation algorithms change when observations become sparse and a few, e.g. adaptive Kalman filter algorithms fail to converge, see Chapter 2.

On the other hand, numerical weather prediction, which is based on data assimilation, has served as a role model for oceanographic data. Although there are clear differences between the ocean and atmosphere, most often applications are similar enough to allow use of similar assimilation methods, and the methods and the literature are common. The reason why the application to the ocean trails has been the lack of urgent demand for forecasting and the lack of appropriate synoptic data sets. However, both of these reasons have recently changed. Oceanic forecasting is being done, and synoptic satellite data sets have become available. As shown in this work, although one needs to be aware of the rich meteorologic data assimilation literature, not all methods can be applied to oceanography, and one needs to develop new techniques more directly relevant to oceanographic problems.

Depending on the temporal and spatial scales of interest, and the computational resources available (CPU and memory), one may choose different data assimilation methods, e.g. nudging, adjoint, or Kalman filtering. Typically, for data assimilation with high resolution models one uses the so-called "optimal interpolation" or nudging techniques. The main reason for this is that they are relatively inexpensive from the computational viewpoint. Their main disadvantage is that the choice of weights for blending data and model estimates is chosen in some ad hoc fashion. In a least-squares sense, the optimal solution is given by the Kalman filter (KF, hereafter), Kalman (1960), and a smoother, Rauch et al. (1965). The computational cost of the KF, where one propagates state error covariances at every time step and uses them for estimating the blending weights is great.

Because of this the KF is currently limited to coarse GCMs and problems where one can significantly reduce number of degrees of freedom.

The KF provides a sequential estimate of the state of the system using only prior information. The estimate is obtained by forming a linear combination of the model forecast and observations, weighted by their corresponding uncertainty. For linear models, the KF, with the companion smoother, is equivalent to the adjoint method. The original KF can be extended to non-linear models, in a so-called extended KF.

The Kalman filter propagates the error covariance matrix, i.e. it provides the accuracy of the estimate at every time step. Because of this, the KF is very computationally expensive. For a system with $n$ degrees of freedom (DOF, hereafter), $2n + 1$ integrations of the numerical model are required for a single time step of the KF algorithm. For systems with a large number of DOF (grid points times number of prognostic variables), at least $O(10^5)$ for oceanographic applications, this becomes prohibitively expensive even with largest supercomputers. In addition, the size of the covariance matrices is $O(n^2)$, so that their storage becomes prohibitive as well. To reduce computational and memory costs, many suboptimal schemes have been developed. One can reduce the computational burden by computing approximate error covariances: either by using approximate dynamics, Dee (1991); computing asymptotically constant error covariances, Fukumori et al. (1993); or propagating only the leading eigenvectors of the error covariances (empirical orthogonal functions, EOFs) as in error subspace state estimation, Lermuisaux (1997).

An alternative way of reducing the computational load is to reduce the dimensionality of the problem. One can reduce the dimension of the model by linearizing a GCM onto a coarser horizontal grid or a small set of horizontal EOFs, and reducing the number of points in the vertical by projecting the model onto a small set of vertical basis functions (EOFs or barotropic and baroclinic modes). For example, Fukumori and Malanotte-Rizzoli (1995) used vertical dynamic modes and horizontal EOFs for their coarse model. A more detailed explanation of this approach is given in Section 2.1. In a different method, one computes EOFs of the whole model, and defines new state variables as

coefficients for each individual EOF, e.g. Cane et al. (1996), Malanotte-Rizzoli and Levin (1999). Although the analysis of Cane et al. (1996) deals with the small domain in the tropical Pacific and only several tide gauges are used, the results suggest that reduced space assimilation is in general not inferior to full Kalman filter for short times, even though the dimension of the model is reduced by several orders of magnitude.

In this work, questions of a more theoretical nature are addressed, namely how to do data assimilation when some of the required input information is absent, e.g. statistics of the errors, as explained below. Therefore, we will concentrate on the basic setup and treat only linearized models. Because for linear models the Kalman filter is the most general data assimilation method, for the purposes of this discussion many other data assimilation methods can be viewed as its special cases, and are not considered.

## 1.2 Adaptive Error Estimation

Apart from the difficulties associated with the large dimensionality of the problem, it is critical for the statistics of the model errors and the measurement noise to be known for the KF estimates to be optimal. Although in some cases of oceanographic interest the errors in observations may be relatively well-determined, e.g. Hogg (1996), this is not typically true because the measurement errors, as defined in the data assimilation context, include the missing model physics, or representation error. This is due to the fact that the model cannot distinguish the processes which are missing from the model, e.g. scales smaller than the model grid size, from the errors in the observations, see the discussion in Section 2.2. The model errors, or system noise, are usually poorly known. Therefore, the resulting estimate of the assimilated ocean state is far from optimal.

Figure 1.4 taken from Dee (1995) shows an example of the effect of incorrectly specified model error statistics on the performance of the KF. The plot shows the time evolution of the root-mean-square (RMS) energy errors for two data assimilation experiments, performed with the two-dimensional linear shallow water model of Cohn and Parrish (1991).

23
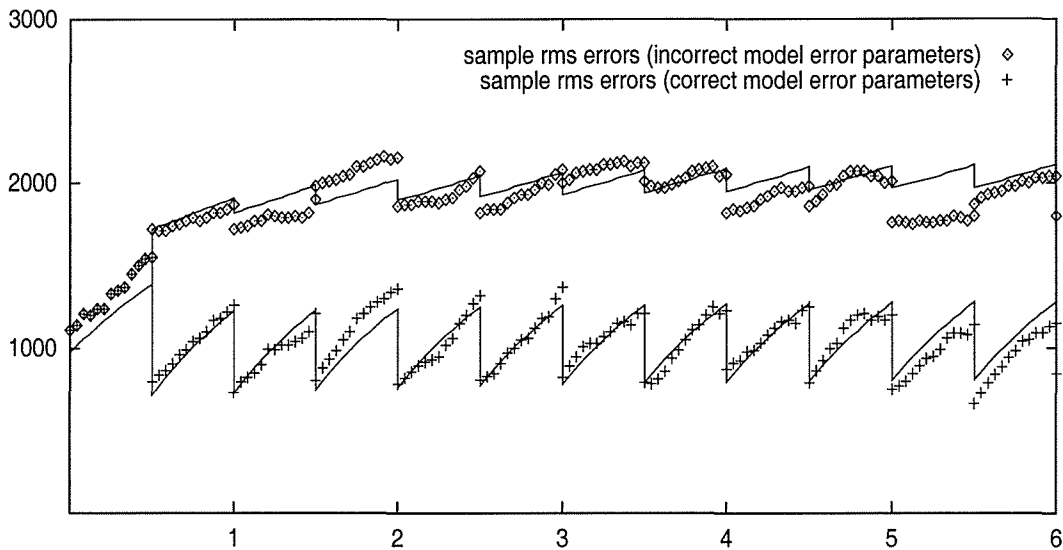
Figure 1.4: RMS energy error evolution for SKF with incorrectly specified (upper part) and correctly specified (lower part) model error statistics. Marks indicate actual RMS errors, curves correspond to their statistical expectations, the time is in days (reproduced from Dee 1995) .

The curves correspond to statistical expectations of forecast and analysis errors obtained from KF theory while the marks denote the errors that actually occurred. In each experiment the same 12-hourly batches of synthetic radiosonde observations were assimilated into a forecast model by means of a simplified KF (SKF), Dee (1991). The two experiments differ only in the way that the model error covariance is specified. The RMS-error curve and marks in the lower part of the figure result from a correct specification of model error covariance, while those in the upper part result from a misspecification of the model error covariance. The disastrous effect of erroneous model error information is clearly visible. The average analysis error level more than doubles after a few days of assimilation. In some cases the assimilation of observations actually has a negative impact on the analysis (e.g. days 0.5, 1.5, and 5.5).

In Figure 1.5 we demonstrate the effect of error covariances on the estimates of the ocean state. The figure shows a comparison of the sea surface height anomaly for one particular cycle of the T/P altimeter (January 1-10, 1994). The top plot shows the

T/P measurements, and the two lower ones show corresponding estimates obtained using the T/P measurements with an approximate Kalman Filter (Chapter 5)[1]. It has to be noted that these data assimilation experiments were done with carefully chosen, and not very different, error covariances. The small-scale signal seen in the T/P measurements is missing in the KF estimates because the assimilation was done with a coarse grid (reduced-state) model, i.e the KF serves as a filter to remove the small scales. Because the error covariances were similar, the two fields obtained with the KF are similar overall. However, there are important differences. For example, the second assimilation (Figure 1.5c) has a strong positive anomaly in the West Equatorial Atlantic which is completely missing in the first assimilation. Unlike the twin experiment example presented above, in this case we do not know the true state. Although there are some tests which allow to check consistency of the estimates, it has been shown that application of such tests to global data assimilation with realistic data distribution is problematic (Daley, 1993). Unless one has independent data, it is difficult to decide which of the two estimates is "better", i.e. closer to the true field. Only careful choice of the a priori error covariances can make the estimates of the state credible.

To make matters worse, when we have wrong estimates of the error covariances, the estimates of the state uncertainty are also wrong, i.e. both estimates of the state and its uncertainty depend critically on the covariances of the model and measurement errors. In addition, the state itself depends on the first-order moments, or bias, of the errors.

Error covariances are also very useful for analyzing the performance of a GCM. They provide information on which geographic locations and what spatial scales the GCM's performance is good or bad. In addition, they set a metric for comparing different GCMs. A quantitative tool which would allow one to perform such comparisons would be highly desirable, for an example of the difficulties one faces when attempting to evaluate the performance of different atmospheric GCMs see Gates (1995).

---

[1] The data assimilation estimates have used 1 year of the T/P measurements starting from January 1, 1993.

a) [−100.,−0.23,109.9]

b) [−25.9,0.209,20.68]

c) [−30.7,0.227,26.9 ]
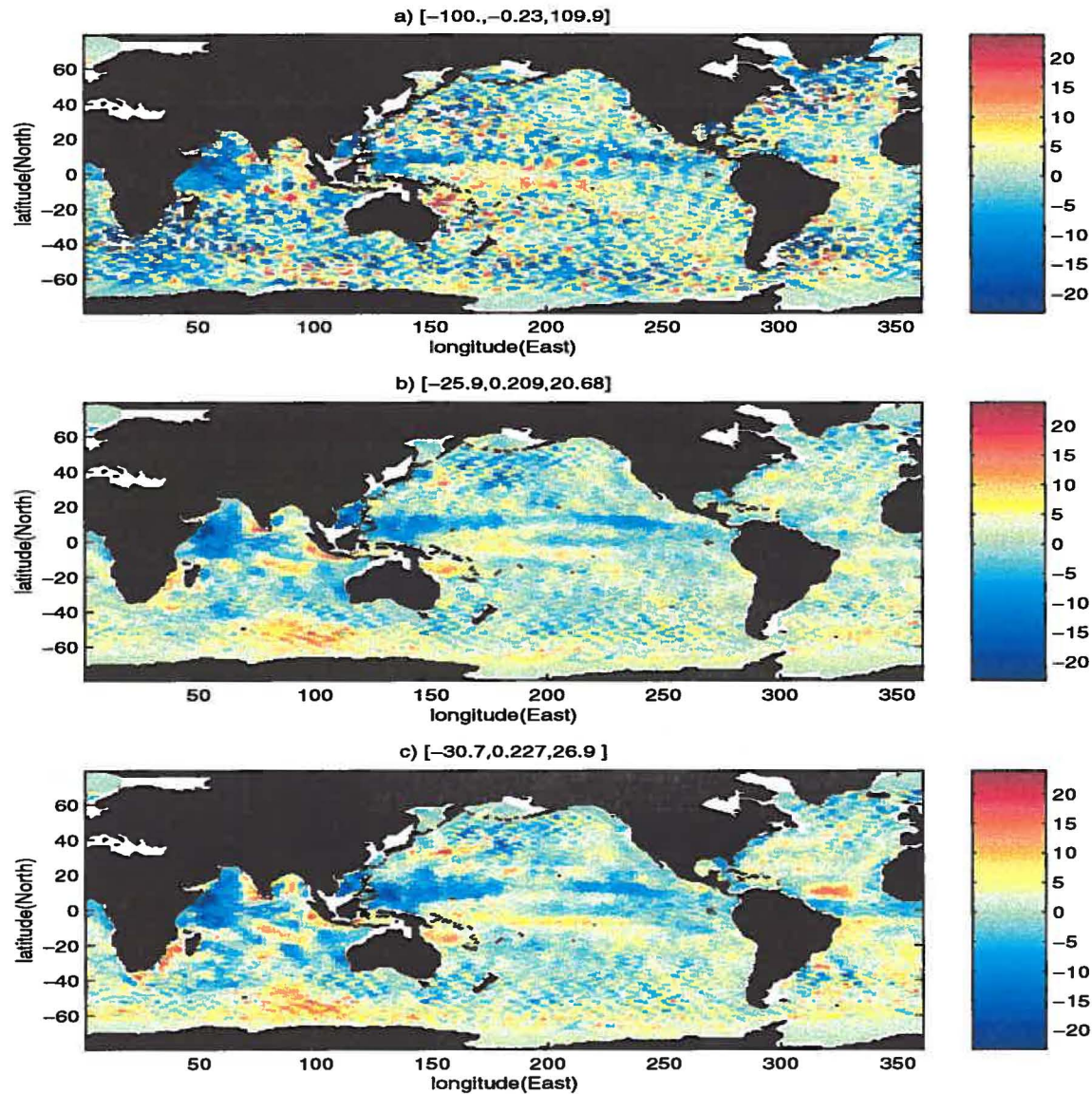
Figure 1.5: Sea surface height anomalies (in cm) for the a) T/P data set, b) and c) data assimilation estimates using the same T/P observations for two different choices of the error covariances. The error covariances for the data assimilation were chosen adaptively (Chapter 5), and were not very different. Overall, the two assimilation estimates are similar, but there are significant differences, e.g. in the equatorial Atlantic. The assimilation runs are described in Chapter 5.

# 1.3    Present Work

The problem of estimating and understanding the error statistics is the subject of this work. It has only recently received attention in the oceanographic community. For a long time the problem did not attract significant attention for several reasons. First, use of the error statistics requires a very significant increase in computational resources, and a significant reduction in the number of degrees of freedom. The methods of state reduction, see Chapter 2, have not been tested and applied to large GCMs until recently. In addition, one needs to use approximate data assimilation schemes, and they also have only recently been developed. Thirdly, the fact that the estimation is sensitive to the specification of the statistics of the observational, and especially model errors, needed to be established, e.g. Wergen (1992). It is worth noting that operational centers still routinely use a "perfect model" assumption, even though there is a strong consensus that present-day GCMs have great difficulty in simulating the real ocean and atmosphere.

The following is a sequential account of the work which can be taken by someone who faces the problem of estimating error statistics for large-scale data assimilation and who wants to understand why the covariance matching algorithm (CMA, hereafter) was developed and what are its advantages. Alternative ways of reading the manuscript are discussed below in Section 1.3.1.

We start addressing the error estimation problem by presenting existing adaptive data assimilation methods. In the present context the term "adaptive" means that we are using data for the simultaneous estimation of the error statistics and the ocean's state. Such adaptive methods were applied to an oceanographic problem in a paper by Blanchet et al. (1997), (BFC97, hereafter). BFC97 ran twin experiments with three different adaptive data assimilation methods, developed and improved upon by a number of different authors in the control and meteorological literature. BFC97 used simulated tide-gauge data and a reduced-state model in the tropical Pacific. In Chapter 2 we test these adaptive methods by trying to obtain quantitative estimates of the large scale internal errors in a GCM using simulated T/P altimeter data as well as acoustic tomography

27

data.

Following the discussion in BFC97 we single out the adaptive method of Myers and Tapley (1976, MT) as the representative of these adaptive methods. Firstly, we present the analysis of the MT method, with low dimensional models. We show that while in principle this method can provide estimates of the model error statistics it has several major drawbacks. When we have sparse observations, the estimates of the error statistics may be sensitive to the initial guess of the model error covariances. The method requires running the Kalman filter, and it takes many iterations for the method to converge. This makes it computationally expensive. Estimation of both model and measurement statistics is unstable, and can lead to wrong estimates. There is no information about uncertainties of the derived error covariances, on how much data is required, and on which parameters can be estimated and which cannot.

We use a twin experiment approach, described in Section 2.8, to show that with the linearized MIT GCM, the MT method is sensitive to the initial choice of the error statistics, and to the kind of observations used in the assimilation. In Section 2.9 we show that similar results are obtained with a maximum likelihood method. The conclusion is that neither of the adaptive data assimilation methods is suitable for quantifying large scale internal ocean model errors with the available altimetric or acoustic tomography observations.

In a different approach, Fu et al. (1993) and Fukumori et al. (1999) estimated the measurement error covariance by comparing the observations with the model forecast without any data assimilation. This method is closely related to the new approach, which we develop and call the Covariance Matching Approach (CMA). It is described in Chapter 3. Although related to the previous methods, the new approach relaxes some of the restrictive assumptions of the method used by Fu et al. (1993). It makes use of information in a more efficient way, allows one to investigate which combination of parameters can be estimated and which cannot, and to estimate the uncertainty of the resulting estimates.

In Chapter 4 we apply the CMA to the same linearized version of the MIT GCM with ATOC and T/P data. Through a series of twin experiments, which use synthetic acoustic thermometry and T/P data, we show that the covariance matching approach is much better suited than the innovation–based approaches for the problem of estimating internal large scale ocean model error statistics with acoustic measurements, but not with altimetric measurements. Because the method uses observations directly instead of the innovations, it allows concurrent estimation of measurement and model error statistics. This is not possible with the adaptive methods based on innovations (Moghaddamjoo and Kirlin 1993).

We then test the CMA with the real TOPEX/POSEIDON altimetry and the ATOC acoustic tomography data. We show that for this model most of the model–data misfit variance is explained by the model error. The CMA can also be extended to estimate other error statistics. It is used to derive estimates of the trends, annual cycles and phases of the errors. After removal of trends and annual cycles, the low frequency/wavenumber (periods > 2 months, wavelengths > 16°) TOPEX/POSEIDON sea level anomaly is order 6 cm$^2$. The GCM explains about 40% of that variance, and the CMA suggests that 60% of the GCM-TOPEX/POSEIDON residual variance is consistent with the reduced state dynamical model. The remaining residual variance is attributed to measurement noise and to barotropic and salinity GCM errors which are not represented in the reduced state model. The ATOC array measures significant GCM temperature errors in the 100–1000 m depth range with a maximum of 0.3° at 300 m.

In Chapter 5, we apply the CMA to a second problem, one which involves estimating global ocean error statistics for a linearized GFDL GCM, with only the barotropic and first baroclinic internal modes[2]. The obtained estimates of error statistics are significantly different from those used in the study of Fukumori et al. (1999), where the

---

[2]Although the vertical modes can only be defined for the linear ocean model, they can be used as a set of vertical basis functions. Fukumori et al. (1998) show that the barotropic and first baroclinic mode explain most the variability of the T/P sea level anomaly. A linearized model based on these two modes is satisfactory for data assimilation needs.

linearized model is used for global data assimilation. The CMA estimate of the model error covariance based on the error model of Fukumori et al. (1999) on average explains forty percent of the model–data residual variance. Most of the model error variance is explained by the barotropic mode, and that the model error corresponding to baroclinic velocities has a negligible contribution.

The CMA estimates of the error covariances are then used with a global data assimilation scheme. Based on analysis the statistics of the innovations, we show that the quality of the data assimilation estimates is improved very little. As pointed out in Chapter 3 the problem of error statistics estimation is very under-determined. Therefore, to obtain statistically significant estimates of the error statistics it is crucial to have a good physical understanding of the model shortcomings. The covariances used in Fukumori et al.(1999), already tuned to the model-data residuals, use the error structures which prove to be quite robust. Comparison of several data assimilation experiments which differ only by the choice of the error covariances demonstrate that data assimilation estimates are not very sensitive to a particular parametrization of the adaptively tuned error statistics.

The summary of the thesis and perspectives for future research are given in Chapter 6.

## 1.3.1   Outline of the Thesis

With a complete description of the work given above, we give advice on how to read the thesis. A complete summary of the notation (with a reference to the original equations) and acronyms is given in tables A.1 and A.2.

The reader who is primarily interested in the results can start directly with the examples. An application of the CMA to a linearized version of the MIT GCM with the TOPEX/POSEIDON altimetry and the ATOC acoustic tomography data is presented in Chapter 4. Second application of the method and an example of data assimilation with adaptively tuned error covariances are presented in Chapter 5. The model used in this chapter is a linearization of the global GFDL GCM and the data consist of the T/P measurements of sea level anomaly.

For a more detailed description of the CMA the reader should consult Chapter 3. The basic algorithm is presented in Section 3.2. For a discussion of practical issues which are important in realistic applications and for the extensions of the method to other statistics one should consult Sections 3.3 and 3.5.

For a reader familiar with, or interested in, innovation based adaptive methods, a comparison of innovation based methods and the CMA is presented in Section 3.6. To get a deeper understanding of an innovation-based approach due to Myers and Tapley (1976), one can consult Section 2.6, where an analytical representation of the method with a scalar model is discussed, and Section 2.7, where a numerical implementation of the method with a multivariate (2 DOF) model is given. In Section 2.8 we demonstrate that this method fails with a linearized version of the MIT GCM, while the CMA can be successfully used in the same setup presented, Section 4.3.

The reader who is willing to take the time and travel the long road can read the work sequentially as described above.

# Chapter 2

# Methods of Adaptive Error Estimation

We start this chapter by setting the problem up and providing a mathematical description. We then discuss available methods of adaptive error estimation. To illustrate the methods we restrict our attention to the following question: "For a linear model with four vertical modes, can we estimate the mean variance of model error for each mode based on the two kinds of available measurements: altimetric measurements of the sea surface height and acoustic tomography measurements of sound speed converted into temperature anomalies?" We use a linearized GCM of the North Pacific, where more than a year of high quality acoustic data are available in addition to the altimetric data. We use the GCM of Marshall et al. (1997a, 1997b) and the reduced state linearization described in Menemenlis and Wunsch (1997). We concentrate on the adaptive method of Myers and Tapley (1976) (MT, hereafter), and in addition consider the maximum likelihood approach of Dee (1991, 1995).

After we describe the model and the methods, we investigate properties of the MT algorithm with low-dimensional systems in order to gain better understanding of the algorithm. We start with a model with one degree of freedom, and then extend the results to a model with two degrees of freedom. The analysis with low-dimensional models illus-

trates the non-linear character and complexity of the adaptive error estimation problem. It provides guidelines for the applicability of the MT algorithm.

Next, we present results of twin experiments with a linearized GCM, i.e. experiments for which synthetic data are used. We start with a series of the experiments in which we compute a single posterior estimate of both model and measurement uncertainties. These runs allow us to sample the parameter space and to develop intuition for the particular linear model appropriate to our experiment and for the two kinds of measurement, altimetric and tomographic. We then present a series of fully adaptive twin experiments. Based on the twin experiments we show that the performance of the adaptive filter depends on the type of observations. The adaptive method of MT cannot estimate the correct uncertainty structure with synthetic altimetric observations, but can do so once a significant number of synthetic tomographic rays are included in the assimilation. However, it fails with the tomographic measurements available at the time this analysis was carried out.

Based on these results, and the fact that the method is sensitive to the initial guess of the error covariances, and, moreover, provides no information on the uncertainty of the derived estimates, we conclude that the estimates we would obtain with real data could not be trusted. In addition, we show why the maximum likelihood method of Dee (1995), provides similar negative results (Section 2.9.1).

The chapter is organized as follows. In the next Section we present the model and the observational networks used in this chapter. In Sections 2.4-2.5 we review the basic and adaptive Kalman filter algorithm. For analysis of the MT method with a one degree of freedom (DOF) model for which analytical representation is obtained see Section 2.6. In Section 2.7 we turn to the analysis of the MT adaptive filter with a 2 DOF model, which allows to consider the important case of incomplete data coverage. In Section 2.8 we present the results of the twin experiments with simulated altimetric and acoustic tomography data. We draw conclusions about the performance and limitations of these adaptive techniques in Section 2.10.

## 2.1 Model

Dynamical models describe how information, or physical properties, propagate in space and time. Ocean models describe how the physical quantities of the ocean (e.g. fluid velocities, temperature, and pressure) change in time and space. Given boundary and initial conditions, we can use the model to obtain information about the state of the ocean in a particular region of the ocean, or at some later time. For reasons outlined in the introduction we will be concerned only with linear, or linearized, models. Below we provide a short description of how a linearized ocean model can be obtained. Complete descriptions of the two different linearized models used in this work are given by Menemenlis and Wunsch (1997), for the linearized MIT GCM, and by Fukumori et al. (1999), for the linearized GFDL GCM.

The models are discretizations of the incompressible Navier-Stokes (NS) equations together with an equation of state. The MIT GCM, developed by Marshall et al. (1997a, 1997b), the linearization of which is used in this chapter, solves the NS equations in spherical geometry with height as a vertical coordinate and with arbitrary basin geometry. It is integrated in hydrostatic mode for the Pacific Ocean with realistic topography and coast lines, and insulating bottom and side walls. A no-slip wall side condition and a free-slip bottom condition are used. The model domain extends from 30°S to 61°N meridionally, and from 123°E to 292°E zonally, with a uniform horizontal grid spacing of 1°. There are 20 vertical levels, with a maximum depth of 5302m. The model time step is 1 hour.

The model is relaxed to climatological values of temperature and salinity at the surface with a time scale of 25 days. Because the model is restricted to the northern part of the Pacific, at the southern boundary the model is relaxed over a 500 km zone with a time scale of 5 days at the boundary increasing linearly to 100 days at the edge of the 500 km.

To obtain a model for large scale ocean climate estimation studies, we need to linearize

the model and to reduce its dimension. First, the GCM is represented algebraically as

$$\boldsymbol{\zeta}_{\mathrm{GCM}}(t + \delta t) = \mathcal{M}[\boldsymbol{\zeta}_{\mathrm{GCM}}(t), \mathbf{w}_{\mathrm{GCM}}(t)], \tag{2.1}$$

where $\mathbf{w}_{\mathrm{GCM}}(t)$ represents boundary conditions and model parameters at time $t$. Column vectors are written as bold lower case characters, and matrices as bold upper case characters. A complete summary of the mathematical notation is presented in Table A.1. The state $\boldsymbol{\zeta}_{\mathrm{GCM}}(t)$ consists of all prognostic variables used in the GCM, and as such has dimension of 1,335,582 for the configuration used in computing the linearized model. We make a fundamental assumption that for large scales, the difference between the true state on the model grid and the model state,

$$\mathbf{p}(t) = \mathbf{B}^*[\hat{\boldsymbol{\zeta}}_{\mathrm{GCM}}(t) - \boldsymbol{\zeta}_{\mathrm{GCM}}(t)], \tag{2.2}$$

is governed by linear dynamics:

$$\mathbf{p}(t + 1) = \mathbf{A}(t)\mathbf{p}(t) + \boldsymbol{\Gamma}(t)\mathbf{u}(t), \tag{2.3}$$

where $\mathbf{p}(t)$ is the coarse state (large scale) error vector, $\mathbf{B}^*$ defines a mapping from the fine (GCM grid) to the coarse (large scale) grid; $\mathbf{A}(t)$ is the coarse state transition matrix; $\mathbf{u}(t)$ is the model error vector, and $\boldsymbol{\Gamma}(t)$ projects the large scale model error $\mathbf{u}(t)$ onto the coarse state. Note that we distinguish the GCM error in the reduced (coarse) space, $\mathbf{p}(t)$, from the stochastic noise driving the GCM error, $\mathbf{u}(t)$, denoted as model error for consistency with the standard KF notation. The true state is denoted by the circumflex. The time step of the reduced state model has been taken to be unity, and in practice is considerably longer than the time step of the GCM. We use a 30 days time step for the linearized model (LM, hereafter) and one-hour time step for the MIT GCM.

It is important to realize that the linear model is not a linearization of the model around its mean state, such as a commonly used linearization of the non-linear quasi-geostrophic model (Pedlosky, 1987 , p. 499). The linear model, equation (2.3), provides an approximate description of how the large scale differences of the GCM estimate and the true ocean on the GCM grid propagate in time. This assumption is based on a

fundamental requirement that the large, slow scales of oceanic variability are separated from meso-scale and other short-term variability, and that the smaller scales effect on the large scale differences can be modeled as a white noise process; see Menemenlis and Wunsch (1997).

The state reduction operator $\mathbf{B}^*$ projects the difference between the hypothetical true state and the model state onto some truncated basis set. In practice, the true state $\hat{\zeta}_{\mathrm{GCM}}(t)$ is approximated by some reference state $\bar{\zeta}$, and the linearization is effectively done around that state; see Fukumori and Malanotte-Rizzoli (1995). The reduction operator may be thought of as a filter which attenuates small scale noise in order to capture the relevant ocean-climate signal. A pseudo-inverse operator, $\mathbf{B}^*$, which maps back from the coarse (reduced) state to the fine (GCM) state can be defined such that

$$\mathbf{B}^*\mathbf{B} = \mathbf{I}, \mathbf{B}\mathbf{B}^* \neq \mathbf{I}, \tag{2.4}$$

and $\mathbf{I}$ is the identity matrix. Therefore, it is possible to write

$$\mathbf{B}\mathbf{p}(t) + \boldsymbol{\epsilon}(t) = [\hat{\boldsymbol{\zeta}}_{\mathrm{GCM}}(t) - \boldsymbol{\zeta}_{\mathrm{GCM}}(t)], \tag{2.5}$$

where $\boldsymbol{\epsilon}(t)$ represents the high frequency/wavenumber components that lie in the null space of the transformation $\mathbf{B}^*$,

$$\mathbf{B}^*\boldsymbol{\epsilon}(t) = \mathbf{O}, \tag{2.6}$$

$\mathbf{O}$ is the matrix of zeroes. These operators are represented schematically in Figure 2.1. The linearized model implies that the large-scale perturbations described by the linearized state vector $\mathbf{p}(t)$ are approximately dynamically decoupled from the null space, i.e.

$$\mathbf{B}^*\mathcal{M}(\boldsymbol{\epsilon}(t)) \approx \mathbf{0}. \tag{2.7}$$

The validity of this assumption needs to be tested with each particular model. We refer the reader to Sections 4 and 6 of Menemenlis and Wunsch (1997), for a demonstration of its validity with the MIT GCM.

36
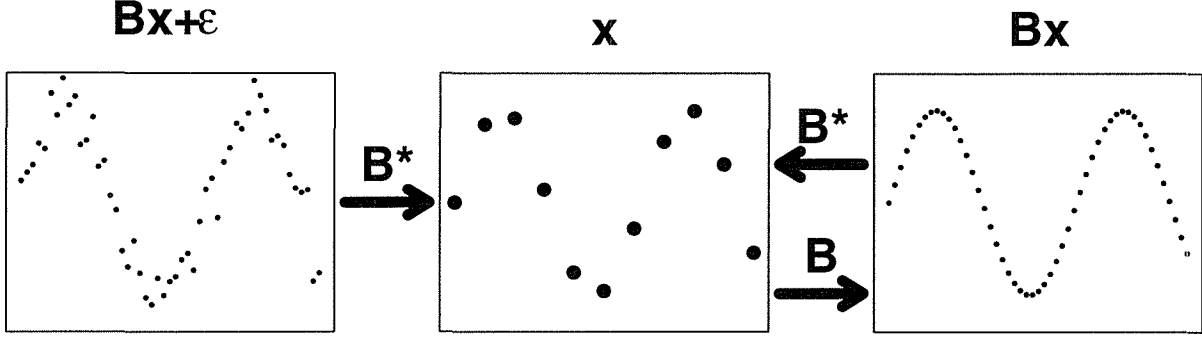
Figure 2.1: Schematic representation of the interpolation and state reduction operators.

The primary purpose of the state reduction operator, $\mathbf{B}^*$ in (2.2), is to reduce the problem size while preserving sufficient resolution to characterize the important physical processes under study. The choice of $\mathbf{B}^*$ needs also to be guided by sampling requirements so as to avoid aliasing. One may wish to define $\mathbf{B}^*$ as a combination of horizontal, vertical, and time reduction operators:

$$\mathbf{B}^* = \mathbf{B}_h^* \mathbf{B}_v^* \mathbf{B}_t^*. \tag{2.8}$$

Corresponding pseudo-inverse operators can be defined and

$$\mathbf{B} = \mathbf{B}_t \mathbf{B}_v \mathbf{B}_h. \tag{2.9}$$

In practice, each pseudo-inverse operator can be defined as

$$\mathbf{B} = \mathbf{B}^{*\mathrm{T}} \left( \mathbf{B}^* \mathbf{B}^{*\mathrm{T}} \right)^{-1}, \tag{2.10}$$

where superscript T denotes the transpose. Non-singularity of $\mathbf{B}^* \mathbf{B}^{*\mathrm{T}}$ is satisfied for any but the most unfortunate choice of $\mathbf{B}^*$ since the number of rows of $\mathbf{B}^*$ is much less than number of columns. For the linearization of the MIT GCM, the vertical state reduction operator $\mathbf{B}_v^*$ maps perturbations onto four vertical temperature EOFs, computed from the difference between GCM output and measured temperature profiles; see Menemenlis et al. (1997a). The EOFs are displayed on Figure 2.2. Horizontal filtering is done using a two-dimensional Fast Fourier Transform algorithm, and setting coefficients corresponding

37
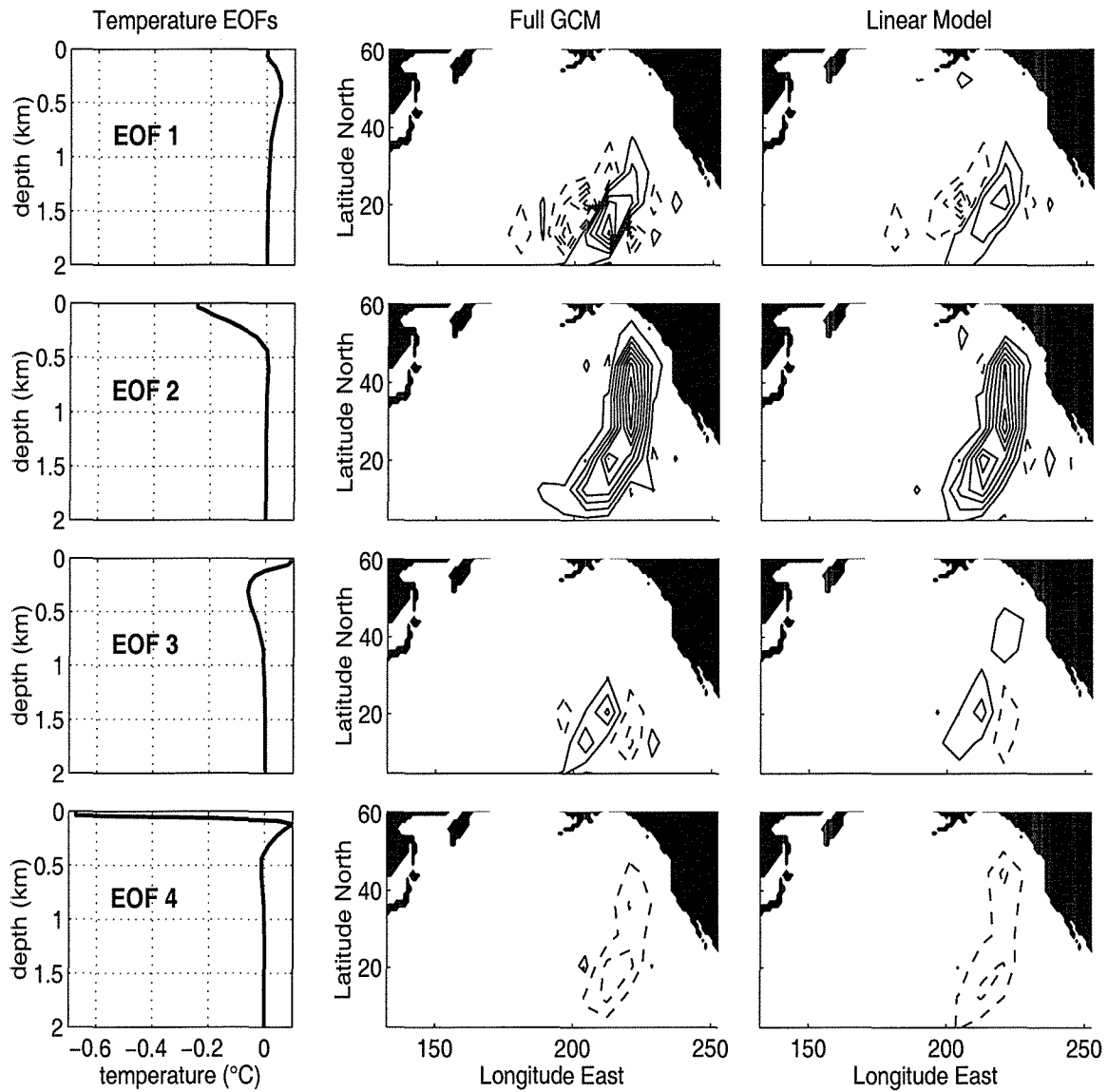
Figure 2.2: Response of the Marshall et al. (1997a, 1997b), Ocean General Circulation Model (OGCM) to a large scale meridional temperature perturbation: the first column displays four vertical temperature Empirical Orthogonal Functions (EOFs) used for state reduction in this study; the second column displays the exact OGCM response projected onto these four EOFs; and the third column is the response of a time-invariant reduced-state linear model to the same perturbation. The perturbation response shown here follows the initial temperature anomaly by a six-month interval and displays a characteristic Rossby-wave-like pattern, with the information propagating westward at an increasingly faster rate as one approaches the equator.

to wavelengths shorter than 16° to zero. The resulting fields are then subsampled at 8° intervals, both zonally and meridionally. This particular choice of $\mathbf{B}_v^*$ and $\mathbf{B}_h^*$ is one of convenience and suffices for the present study. No explicit time filtering is required as the the fields have red frequency spectra, and horizontal filtering makes the time filtering unnecessary.

For the MIT GCM the perturbations were obtained in the following manner. The model was initialized from climatological annual mean temperature and salinity obtained from Levitus (1982), and a resting flow field. It was then integrated for 17 years with annual mean temperature, salinity, and surface wind forcing. From year 18 onwards it was integrated with monthly mean temperatures and seasonal salinities and monthly winds from Trenberth et al. (1989), all linearly interpolated to 24 hour intervals. From year 29, surface heat and freshwater fluxes from Oberhuber (1988) were introduced in the surface layer, while continuing to relax to climatological temperature and salinity. This run of the model adequately reproduces the large scale wind driven circulation, but fails to properly represent the small scale processes. It is however adequate for the present work which aims to quantify the large scale error structure. Figure 2.3 shows a particular monthly-mean sea surface elevation and horizontal velocity produced by the GCM.

The MIT GCM was then integrated for 2 more years with monthly forcing starting from the spun-up state, obtained above, to produce a reference state, $\bar{\zeta}$, for the perturbation analysis. To generate perturbations relative to this reference state (used instead of the true $\hat{\zeta}_{\text{GCM}}(t)$ in equation (2.2)), temperature anomalies are introduced, and the model is integrated with the same boundary conditions and model parameters as for the reference state, following the Green's function approach of Stammer and Wunsch (1996). The anomalies are computed by applying $\mathbf{B}_v\mathbf{B}_h\delta_{\mathbf{p}}$ temperature perturbations to the MIT GCM, where $\delta_{\mathbf{p}}$ represents a delta vector, with all zeroes except for only one element, of the same size as the reduced state vector $\mathbf{p}$. The resulting perturbations relative to the reference state are projected back onto the coarse grid, and form columns of the coarse state transition matrix $\mathbf{A}$. The time-invariant linear model $\mathbf{A}$ is able to satisfactorily re-

Marshall et al. GCM: Surface elevation and horizontal velocity at 37.5 for July 1996
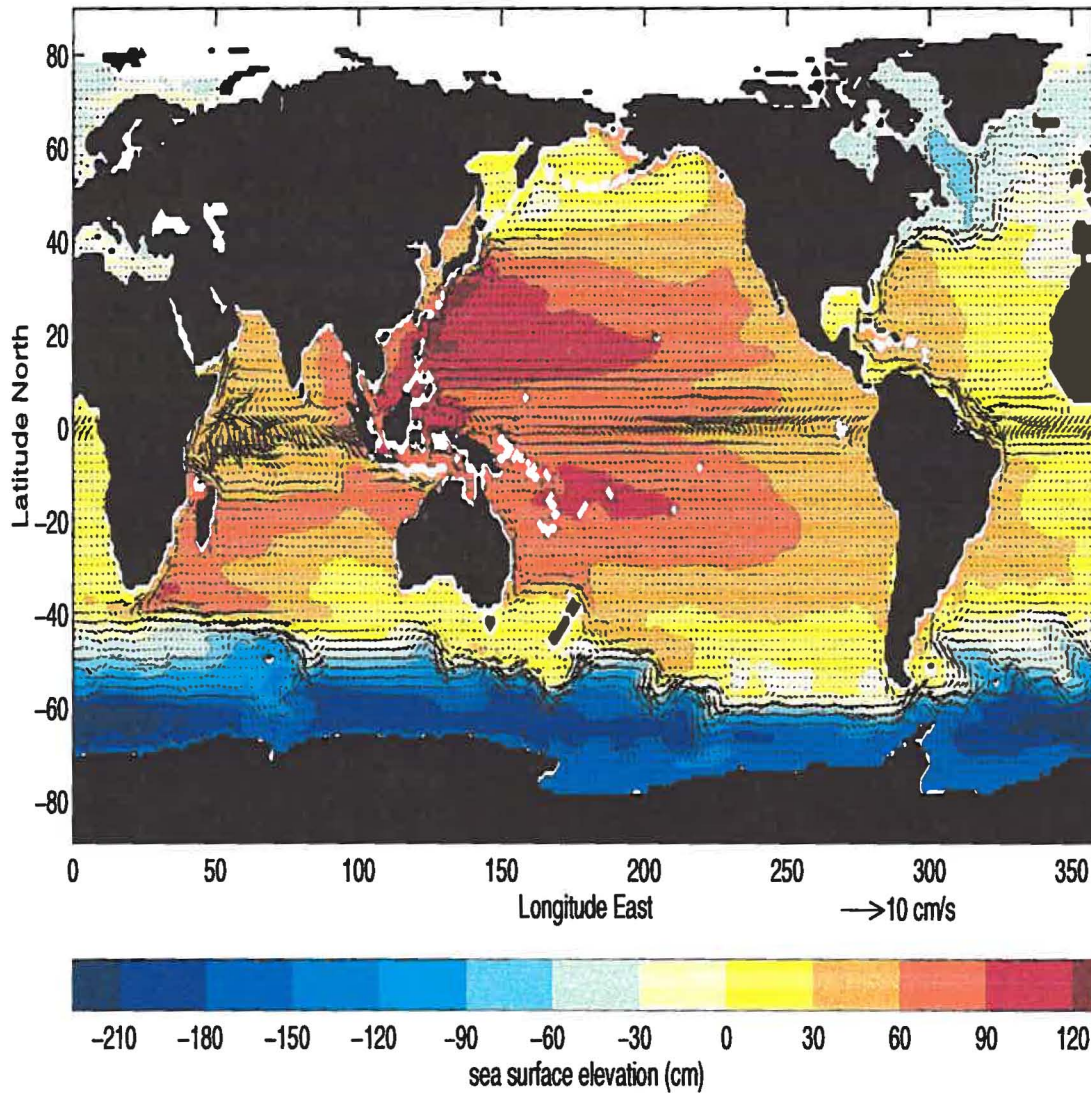


Figure 2.3: July monthly-mean sea surface elevation and horizontal velocity in the 2nd layer produced by the MIT GCM run.

produce the large scale response to a temperature perturbation of fully non-linear GCM (Figure 2.2).

There are other methods available for state reduction. For example, one can choose to start from the coarse to fine transformation $\mathbf{B}$ instead of $\mathbf{B}^*$ as above; see Fukumori and Malanotte-Rizzoli (1995). In addition, one can choose the coarse state entirely of EOF coefficients, as in Cane et al. (1996). Alternatively, instead of using the Green's functions one can use principal oscillation patterns and compute the response of the GCM to random initial fields.

## 2.2 Data

The first dataset used in this study consists of over four years (October 1992 – February 1997) of TOPEX/POSEIDON satellite altimeter sea-surface anomaly. Altimetric observations provide a dynamical surface boundary condition for the ocean general circulation, Stammer et al. (1996). In contrast acoustic tomography samples the interior ocean by transmitting sound pulses from sources to receivers along many paths, Munk et al. (1995). Variations in acoustic travel times are, to first order, a measure of temperature anomalies. To a much lesser degree they are also related to variations in current velocity and salinity. Figure 2.4 displays the estimation domain and ATOC acoustic paths used in the present analysis, superimposed on a map of rms sea-surface variability from the TOPEX/POSEIDON altimeter.

These observations measure properties of the real ocean and can be described symbolically as

$$\boldsymbol{\eta}_{\text{ocean}}(t) = \boldsymbol{E}_{\text{ocean}}(t)\boldsymbol{\zeta}_{\text{ocean}}(t) + \boldsymbol{\nu}_{\text{ocean}}(t) \tag{2.11}$$

where $\boldsymbol{\zeta}_{\text{ocean}}(t)$ represents the state of the real ocean (an infinite dimensional vector), $\boldsymbol{E}_{\text{ocean}}$ represents the measurements' sampling operator, and $\boldsymbol{\nu}_{\text{ocean}}(t)$ denotes the instrument noise. The true state on the model grid $\hat{\boldsymbol{\zeta}}_{\text{GCM}}(t)$ is assumed to be related to the real

41
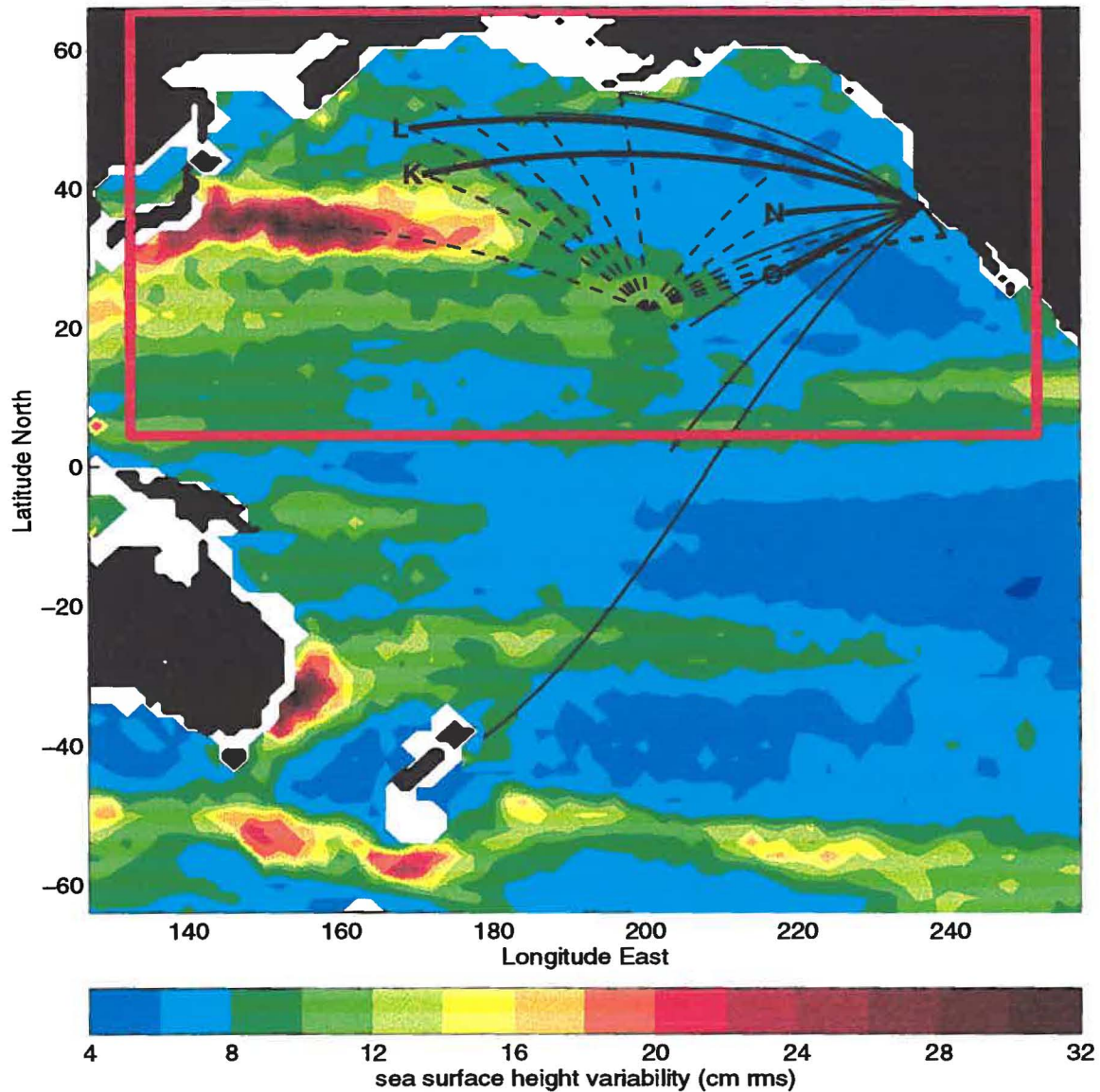
Figure 2.4: Sea surface height variability, in cm rms, measured by the TOPEX-POSEIDON altimeter for the period November 21, 1992 – November 17, 1995. The solid lines indicate the present coverage of the Acoustic Thermometry of Ocean Climate (ATOC) array using a single acoustic source near the California coast in operation since January 1996. The paths shown in dashed lines represent the increased coverage that will result from the installation of a second source near Hawaii in early 1997. The present study is of the region enclosed by the red rectangle and is based on a preliminary analysis of acoustic data from paths K, L, N, and O (bold lines).

ocean through some operator $\mathcal{N}$:

$$\hat{\boldsymbol{\zeta}}_{\mathrm{GCM}}(t) = \mathcal{N}\left(\boldsymbol{\zeta}_{\mathrm{ocean}}(t)\right) \tag{2.12}$$

To relate the observations $\boldsymbol{\eta}_{\mathrm{ocean}}(t)$ to the model state we rewrite equation (2.11) as

$$\boldsymbol{\eta}_{\mathrm{ocean}}(t) = \mathbf{E}(t)\hat{\boldsymbol{\zeta}}_{\mathrm{GCM}}(t) + \left(\boldsymbol{E}_{\mathrm{ocean}}(t)\boldsymbol{\zeta}_{\mathrm{ocean}}(t) - \mathbf{E}(t)\mathcal{N}\left(\boldsymbol{\zeta}_{\mathrm{ocean}}(t)\right)\right) + \boldsymbol{\nu}_{\mathrm{ocean}}(t) \tag{2.13}$$

where projection operator $\mathbf{E}(t)$ has been redefined as the model and the real ocean states are defined on different grids. Furthermore, the relation between observations and the GCM state is assumed to be linear. Typically, matrix $\mathbf{E}(t)$ is sparse with only a few non-zero elements.

The second term on the RHS of equation (2.13) describes the difference between the real ocean and the finite dimensional model, and is termed "representation error"; see Fukumori et al. (1998). It corresponds to processes which affect observations but that are missing from the model, and typically correspond to scales smaller than the model grid size. As far as the model is concerned it is indistinguishable from the instrument error and the two can to be lumped together into, for a lack of a better term, measurement error $\boldsymbol{\nu}(t)$:

$$\boldsymbol{\nu}(t) = \left[\boldsymbol{E}_{\mathrm{ocean}}(t)\left(\boldsymbol{\zeta}_{\mathrm{ocean}}(t)\right) - \mathbf{E}(t)\mathcal{N}\boldsymbol{\zeta}_{\mathrm{ocean}}(t)\right] + \boldsymbol{\nu}_{\mathrm{ocean}}(t). \tag{2.14}$$

To summarize, the measurements can be represented as some linear combination of the state vector $\hat{\boldsymbol{\zeta}}_{\mathrm{GCM}}(t)$ plus noise $\boldsymbol{\nu}(t)$:

$$\boldsymbol{\eta}_{\mathrm{ocean}}(t) = \mathbf{E}(t)\hat{\boldsymbol{\zeta}}_{\mathrm{GCM}}(t) + \boldsymbol{\nu}(t). \tag{2.15}$$

It is convenient to define the observed difference between the measurements and the GCM prediction:

$$\begin{aligned} \mathbf{y}(t) &= \boldsymbol{\eta}_{\mathrm{ocean}}(t) - \mathbf{E}(t)\boldsymbol{\zeta}_{\mathrm{GCM}}(t) \\ &= \mathbf{E}(t)\mathbf{B}\mathbf{p}(t) + \mathbf{r}(t). \end{aligned} \tag{2.16}$$

The observed difference is now expressed in terms of the reduced state vector $\mathbf{p}(t)$. The observational noise term includes two contributions, one due to measurement error $\boldsymbol{\nu}(t)$, which includes unresolved scales and the missing physics of the GCM, and another due to high frequency, small-scale variability present in the GCM but not in the reduced state model, $\boldsymbol{\epsilon}(t)$:

$$\mathbf{r}(t) = \mathbf{E}(t)\boldsymbol{\epsilon}(t) + \boldsymbol{\nu}(t). \tag{2.17}$$

Our goal is to quantify errors in the large scale baroclinic variability of the GCM in the North Pacific relative to the variability measured by the altimeter and the acoustic tomography array.

## 2.3   Mathematical Formulation

With the reduced state model described above, we present adaptive estimation algorithms in a more general setup of a linear prediction model, written as

$$\mathbf{p}(t+1) \;\; = \;\; \mathbf{A}(t)\mathbf{p}(t) + \mathbf{G}(t)\mathbf{w}(t) + \boldsymbol{\Gamma}(t)\mathbf{u}(t), \tag{2.18}$$

$\mathbf{p}(t)$ denotes the state space vector, $\mathbf{w}(t)$ denotes the known forcing, $\mathbf{u}(t)$ denotes the model error, or system noise. The time step is taken to be $\delta t = 1$ for simplicity. The state space vector $\mathbf{p}(t)$ includes physical quantities necessary to describe the system at time $t$ (CGM error on the coarse grid for the reduced state model (2.3)). The vector $\mathbf{p}(t)$ has length $N$. Forcing $\mathbf{w}(t)$ includes boundary conditions, and is externally prescribed to the model. $\mathbf{G}(t)$ maps the forcing onto the state. For the reduced state model, equation (2.3), the forcing $\mathbf{w}(t)$ is assumed to be identically zero. The model prediction at time step $t + 1$ depends on $\mathbf{A}(t)$, the "state transition matrix" which represents an approximation of the model dynamics, the estimate of the state at the previous time step $t$, and the forcing.

We complete the description of the model by providing initial conditions, with the

corresponding error covariance

$$\mathbf{\Pi}(0) = < [\hat{\mathbf{p}}(0) - \mathbf{p}(0)][\hat{\mathbf{p}}(0) - \mathbf{p}(0)]^T > . \tag{2.19}$$

In the same way we define the state error covariance at time $t$

$$\mathbf{\Pi}(t) = < [\hat{\mathbf{p}}(t) - \mathbf{p}(t)][\hat{\mathbf{p}}(t) - \mathbf{p}(t)]^T > . \tag{2.20}$$

We rarely have good estimates of the uncertainty covariance matrix $\mathbf{\Pi}(0)$ for the initial condition. However, for stable models results of data assimilation are insensitive to the choice of $\mathbf{\Pi}(0)$ and a matrix with very large diagonal elements is chosen. In that case $\mathbf{\Pi}(t)$ rapidly decreases with time and reaches a steady value if the system is controllable and observable, see Anderson and Moore (1979).

We relate observations $\mathbf{y}(t)$ to the state space vector $\hat{\mathbf{p}}(t)$ in the same way as for the reduced state model (linearized GCM) in equation (2.16):

$$\mathbf{y}(t) = \mathbf{H}(t)\mathbf{p}(t) + \mathbf{r}(t), \tag{2.21}$$

where $\mathbf{r}(t)$ stands for the observational, or measurement, noise. $\mathbf{H}$ is the "observation matrix". The length of vector $\mathbf{y}(t)$ is equal to $M$ and is typically smaller than $N$ (for altimetric measurements on the LM grid $M = 128$). The observations are available over $T$ time steps.

## 2.3.1  Errors

The model error $\mathbf{u}(t)$ accounts for imperfect knowledge of the forcing, the linearization error, the discretization error, the truncation error, and external forces that are not represented by the forcing term $\mathbf{w}(t)$, etc. The model errors $\mathbf{u}(t)$ are typically assumed to be white and stationary in time, and normally distributed with the mean and covariance given by:

$$< \mathbf{u}(t) >= \bar{\mathbf{u}}; \; < [\mathbf{u}(t') - \bar{\mathbf{u}}][\mathbf{u}(t) - \bar{\mathbf{u}}]^T >= \hat{\mathbf{Q}}\delta_{t,t'}, \tag{2.22}$$

where $< \cdot >$ stands for the expectation operator. In principle, the assumption that the model errors are uncorrelated in time can be relaxed; see Gelb (1979), but then one needs to provide the correlation structure for the errors. The mean $\bar{u}$ is conventionally set to zero, i.e. the model error is assumed to be unbiased. In principle, the algorithms can be readjusted for the biased case; see Blanchet (1997), and Dee and da Silva (1997). Equations (2.18) to (2.20) give a full description of the model and we can integrate the initial conditions forward. However, in the absence of any data assimilation uncertainty of the estimate, $\mathbf{\Pi}(t)$, will grow with time and very soon the estimate will become useless. We rescue the situation by assimilating data when it becomes available. This prevents the uncertainty of the estimate from growing linearly with time.

The observational noise $\mathbf{r}(t)$ has two contributions: 1) error inherent in any real physical observation, i.e. instrument error, and 2) any physics which is in the data and not in the model; see equation (2.17). The noise is assumed to be stationary, white in time, Gaussian with zero mean and covariance matrix $\hat{\mathbf{R}}$:

$$< \mathbf{r}(t) >= 0; < \mathbf{r}(t)\mathbf{r}(t')^{\mathrm{T}} >= \hat{\mathbf{R}}\delta_{t,t'}. \tag{2.23}$$

The terms model and observational errors, adapted from the control theory, are perhaps confusing. Namely, the model error term, $\mathbf{u}(t)$ does not include all the shortcomings of the model, and some of the model flaws, the representation errors, are included in the term $\mathbf{r}(t)$, called observational errors. Moreover, this division is data-dependent, i.e. the split will differ from one data set to another. Only when the measurements are available on the grid of the linear model, the observational error will consist of the inaccuracies of the observations alone. This complicates the interpretation of the results, as small model errors do not necessarily imply that the model is very good, since the part of the model flaws termed representation errors, e.g. inaccurate representation of the eddies in coarse grid models, is large (see Chapter 5 for additional discussion).

Another example is that of observations which include effects of the internal waves and a model which completely neglects the internal waves. For consistent assimilation of such data into this model one would need to remove the internal waves from the

observations. This may be difficult to do, and an alternative is to model the internal waves as errors in the observations, as the so-called representation errors.

## 2.4  Kalman Filter

With the model and measurement models given above, we can now define the estimation problem. The objective of inverse theory is to obtain a best possible, here in the least squares sense, estimate of the state of the model $\mathbf{p}(t)$ from observations $\mathbf{y}(t)$. Mathematically, the goal is to minimize a suitably defined cost function. The cost function can be written as

$$J_{\text{total}} = (\mathbf{p}(0) - \mathbf{p}_0)^{\mathrm{T}} \mathbf{\Pi}(0)^{-1} (\mathbf{p}(0) - \mathbf{p}_0) + \sum_{t=1}^{T} J(\mathbf{u}(t), \mathbf{r}(t), t|t) \qquad (2.24)$$

$$J(\mathbf{u}(t), \mathbf{r}(t), t|t) = \mathbf{r}(t)^{\mathrm{T}} \hat{\mathbf{R}}^{-1} \mathbf{r}(t) + \mathbf{u}(t)^{\mathrm{T}} \hat{\mathbf{Q}}^{-1} \mathbf{u}(t), \qquad (2.25)$$

subject to the models

$$\mathbf{r}(t) = \mathbf{y}(t) - \mathbf{H}(t)\mathbf{p}(t),$$

$$\mathbf{u}(t-1) = \mathbf{p}(t) - \mathbf{A}(t-1)\mathbf{p}(t-1) - \mathbf{G}(t-1)\mathbf{w}(t-1).$$

The notation $J(\mathbf{u}(t), \mathbf{r}(t), t|t)$ means that only observations prior and including time $t$ are used. The cost function seeks the state vector $\mathbf{p}(t), 0 \leq t \leq T$, and the model error, or control, vector, $\mathbf{u}(t), 0 \leq t \leq T - 1$, that satisfy the model equation (2.18) and that agree with the observations and the initial conditions to an extent determined by the weight matrices, namely the covariance matrices $\mathbf{\Pi}(0), \hat{\mathbf{R}}, \hat{\mathbf{Q}}$. Accordingly, the first term on the RHS of equation (2.25) penalizes the misfit between the observations and the model estimate, and the second term acknowledges the fact that driving the model with arbitrarily large controls is not acceptable.

The Kalman filter sequential algorithm yields a minimum of the cost function when we use real-time observations (the individual terms given by $J(\mathbf{u}(t), \mathbf{r}(t), t|t)$), as is typical in engineering applications. When observations are stored and available all at once (the individual terms change to $J(\mathbf{u}(t), \mathbf{r}(t), t|T)$), as is typical of oceanographic studies, the

47

Kalman filter needs to be supplemented by a smoother; see Wunsch (1996). In this case of batch observations and with the assumptions made above, the sequential algorithm provides an answer identical to an adjoint solution, which tries to minimize the global sum (2.24) as one huge optimization problem.

The important point is that error covariance matrices serve as weights for the cost function. Therefore, they are central to data assimilation, independent of a particular algorithm (such as Kalman filter, adjoint, or an approximation such as a nudging algorithm). When we have a poor estimate of error covariances the solution minimizes a wrong cost function, and is therefore far from optimal; see Todling and Cohn (1995). To make this distinction explicit, we introduce new notation for the prior error covariance matrices, $\mathbf{Q}$ and $\mathbf{R}$, that is we drop the hats over the error covariance matrices.

Before we turn our attention to adaptive algorithms we present the Kalman filter algorithm. For a complete treatment one should consult Anderson and Moore (1979). One of the many equivalent Kalman filter formulations taking the state from time $t$ to time $t+1$ is

$$\tilde{\mathbf{p}}(t+1|t) = \mathbf{A}\tilde{\mathbf{p}}(t|t) + \mathbf{G}\mathbf{w}(t), \tag{2.26}$$

$$\mathbf{\Pi}(t+1|t) = \mathbf{A}\mathbf{\Pi}(t|t)\mathbf{A}^{\mathrm{T}} + \mathbf{\Gamma}\mathbf{Q}\mathbf{\Gamma}^{\mathrm{T}}, \tag{2.27}$$

$$\mathbf{K}(t+1) = \mathbf{\Pi}(t+1|t)\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{\Pi}(t+1|t)\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{2.28}$$

$$\tilde{\mathbf{p}}(t+1|t+1) = \tilde{\mathbf{p}}(t+1|t) + \mathbf{K}(t+1)\left(\mathbf{y}(t+1) - \mathbf{H}\tilde{\mathbf{p}}(t+1|t)\right), \tag{2.29}$$

$$\mathbf{\Pi}(t+1|t+1) = \mathbf{\Pi}(t+1|t) - \mathbf{K}(t+1)\mathbf{H}\mathbf{\Pi}(t+1|t). \tag{2.30}$$

Estimates obtained using data assimilation are denoted by a tilde. We assume that $\mathbf{A}$, $\mathbf{G}$, $\mathbf{\Gamma}$, $\mathbf{H}$, $\mathbf{R}$ and $\mathbf{Q}$ are time independent. This assumption can be relaxed without any change in the algorithm, and is used only to simplify the discussion. We use the notation where $(t+1|t)$ represents estimates after the model simulation, or forecast, from time $t$ to $t+1$, and $(t+1|t+1)$ represents estimates after the Kalman filter assimilation, or analysis. It is shown schematically in Figure 2.5.

The Kalman filter, with the full algorithm given above, provides an optimal estimate

48
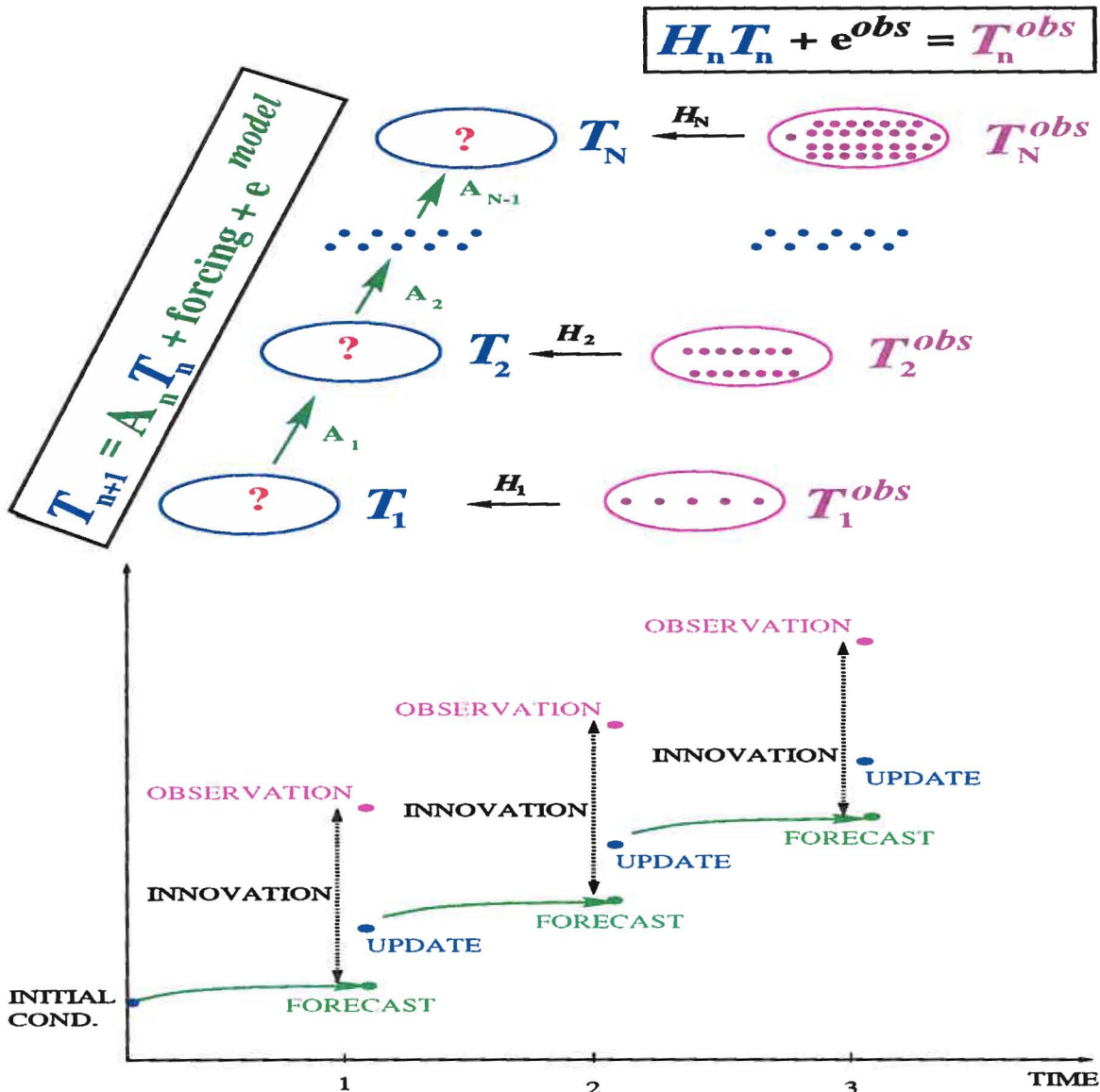
Figure 2.5: Two graphical representations of the Kalman filter. **Top)** $T_n$ represents all model variables, $T_n^{obs}$ represents observations, $A_n$ dynamical model which forecasts $T_{n+1}$ using $T_n$, $H_n$ projection operator which maps model variables onto observations at time $n$. Both model and measurement equations have errors, $e^{model}$ and $e^{obs}$, respectively. The average magnitudes of these errors (given by their covariances) are used as weighting matrices to obtain the best possible estimate at time $n$, denoted by the question mark, which is then propagated forward to obtain the forecast at time $n + 1$. **Bottom)** Alternative view of the Kalman filtering. Starting from the initial condition at time 0, we obtain the forecast at time 1, and then get a weighted average of the forecast and the observations (the difference is called innovation vector) to obtain the update state at time 1. This update is then used as initial condition and the model is propagated again to give a forecast at time 2, and so on.

49

of the state vector in the least squares sense *only* when the correct estimates of $\mathbf{Q}$ and $\mathbf{R}$ are available, i.e. $\mathbf{Q} = \hat{\mathbf{Q}}$ and $\mathbf{R} = \hat{\mathbf{R}}$. In any practical situation, even when all other assumptions are valid, the estimates are suboptimal because we do not know the true covariances. To complicate the matter, the results of the data assimilation are dependent on the estimates $\mathbf{Q}$ and $\mathbf{R}$ in a complex non-linear fashion.

The part of the observational error covariance matrix $\mathbf{R}$ which corresponds to instrument error is typically better known. The other part, which represents the missing physics is model dependent and is therefore poorly known. In some oceanographic studies it is reasonable to assume that $\mathbf{R}$ is dominated by the instrument error.

The model error covariance matrix is most often poorly known. For a system with $N$ state elements, we have to specify $N(N + 1)/2$ elements of $\mathbf{Q}$. But we cannot hope to estimate the full matrix $\mathbf{Q}$. In a typical oceanographic application $N$ is at least on the order of 1000. This enormous informational requirement, rather than the computational cost, is the real obstacle to a successful implementation of the optimal filter. No data assimilation procedure can produce meaningful results if the required information is missing. However, we can try to remedy the problem by applying an adaptive Kalman filter.

## 2.5   Adaptive Kalman Filter

The subject of adaptive estimation is a vast one. Here we present only a summary of the methods which may be useful in the oceanographic context. The sources come from oceanographic, meteorological, and control engineering literature.

The Kalman filter employs a large number of assumptions. In a real system many of them are violated and the filter estimate is suboptimal. Some studies have concentrated on analyzing effects of one particular assumption. For example, Daley (1992a) considered effects of serial correlation in the model error. Here we consider the most fundamental case where all the assumptions are valid but that the error covariance matrices of the

model error, $\hat{\mathbf{Q}}$, and the measurement error, $\hat{\mathbf{R}}$, are not known.

This problem received a lot of attention in control engineering literature in the late nineteen sixties and the beginning of the nineteen seventies. Different methods dealing with a general class of problems have been developed, ranging from estimation of the noise statistics as well as the state transition matrix to methods which were applicable to real time applications with little additional computational cost, e.g. reviews of Sage and Husa (1969), and Mehra (1970). For the problem considered in this work, where the dimension of the state is relatively large, the noise statistics are assumed stationary in time and is the only unknown, only some of these methods are of interest. In a recent paper Blanchet et al. (1997) considered the relevant methods of adaptive error estimation using twin experiments based on a reduced space tropical Pacific ocean model.

## 2.5.1 The Method of Myers and Tapley

The first method, originally due to Myers and Tapley (1976) (MT, hereafter), uses estimates obtained with the Kalman filter to compute approximations of the system noise. That is, the true values for the model error, $\mathbf{u}(t)$, defined in equation (2.18), are replaced by the KF ones, $\tilde{\mathbf{u}}(t)$,

$$\tilde{\mathbf{u}}(t) = \tilde{\mathbf{p}}(t+1|t+1) - \tilde{\mathbf{p}}(t+1|t), \qquad (2.31)$$

where the definition of the Kalman filter forecast, equation (2.26), has been used. To obtain the KF estimates, $\tilde{\mathbf{p}}(t+1|t+1)$ and $\tilde{\mathbf{p}}(t+1|t)$ one needs to run the KF, and thus to provide initial guesses for the error covariances, $\mathbf{Q}_0$ and $\mathbf{R}_0$. The method is empirical, but related to the maximum likelihood methods presented in Abramson (1968), and Maybeck (1982); see BFC97.

We then define an unbiased estimator of the mean of $\tilde{\mathbf{u}}$ and the covariance of the system noise by using sample estimates over the last $S$ steps, and subtracting the expected values,

$$< \tilde{\mathbf{u}}(t) > \quad = \quad \frac{1}{S} \sum_{i=t-S+1}^{t} \tilde{\mathbf{u}}(i),$$

51

$$\tilde{\mathbf{Q}}_{\mathrm{mt}}(t) = \frac{1}{S-1} \sum_{i=t-S+1}^{t} \Big( [\tilde{\mathbf{u}}(i) - <\tilde{\mathbf{u}}(t)>][\tilde{\mathbf{u}}(i) - <\tilde{\mathbf{u}}(t)>]^{\mathrm{T}} \qquad (2.32)$$
$$- \frac{S-1}{S} \big[ \mathbf{A}\mathbf{\Pi}(i-1|i-1)\mathbf{A}^{\mathrm{T}} - \mathbf{\Pi}(i|i) \big] \Big)$$

Note that if we know that the model error has zero mean, we can set $<\tilde{\mathbf{u}}(t)> = 0$. To derive the last term on the RHS of equation (2.32), the bias correction, derived in Myers (1974), the sample estimates, equation (2.31), are assumed to be independent in time[1]. The bias correction is only valid when the prior estimate of the model error covariance is equal to the true one, i.e $\mathbf{Q}_0 = \hat{\mathbf{Q}}$. Still, its use for an incorrect prior estimate is justified if recursive application of the algorithm can be shown to converge to the true estimate. For example, analysis with a scalar model, i.e. a model with 1 degree of freedom, shows that the MT algorithm converges to the correct estimate when the bias correction is applied, (Section 2.6).

A similar estimate can be derived for the measurement noise covariance:

$$\tilde{\mathbf{r}}(t) = \mathbf{y}(t) - \mathbf{H}\tilde{\mathbf{p}}(t|t-1),$$
$$\tilde{\mathbf{R}}(t) = \frac{1}{S-1} \sum_{i=t-S+1}^{t} \big[ \tilde{\mathbf{r}}(i)\tilde{\mathbf{r}}(i)^{\mathrm{T}} \big], \qquad (2.33)$$

and it is assumed that the measurement error has zero mean. After a similar correction for the statistical bias, the MT estimate of the measurement noise covariance becomes

$$\tilde{\mathbf{R}}_{\mathrm{mt}}(t) = \tilde{\mathbf{R}}(t) - \frac{1}{S} \sum_{i=t-S+1}^{t} \mathbf{H}\mathbf{\Pi}(i|i-1)\mathbf{H}^{\mathrm{T}}. \qquad (2.34)$$

To apply the algorithm, we choose initial estimates for $\mathbf{Q}$ and $\mathbf{R}$, allow for initial transients to settle, and then replace $\mathbf{Q}$ and $\mathbf{R}$ by their estimates, using equations (2.32) and (2.34), at every time step. A complete description of the algorithm is given in Section 2.6. There are several potential problems with this algorithm.

Maybeck (1982) states that the existence of an independent and unique solution for his method is subject to question, and furthermore, simultaneous estimation of both $\mathbf{Q}$

---

[1]This assumption is not strictly correct as both $\tilde{\mathbf{u}}(t)$ and $\tilde{\mathbf{u}}(t-1)$ depend on the observations $\mathbf{y}(t-1)$, $\mathbf{y}(t-2)$, etc.

and **R** is not well behaved. This fact was originally recognized by Myers (1974), and later confirmed by Groutage et al. (1987). In the twin experiments below we test the algorithm by estimating the model error covariance only and performing sensitivity study with wrong choices for the measurement error covariance **R**.

The estimates from equations (2.32) and (2.34) can be non-positive semidefinite. This is clearly troublesome as a true covariance matrix must always be positive semidefinite. When a covariance matrix is non-positive semi-definite, the Kalman filter algorithm, see equations (2.26-2.30), becomes numerically unstable. Several ways to deal with this problem have been proposed. Myers and Tapley (1976) reset negative diagonal elements to their absolute value, while BFC97 proposed setting negative eigenvalues to zero. We have used the latter approach.

The method can be unstable if one tries to estimate too many parameters. The method also can be sensitive to the choice of the averaging window length $S$ (see the discussion in BFC97). For the model and data considered in BFC97 the algorithm gave a unique estimate; it did not depend on the initial choice of **Q** and **R**. However, in our case (Section 2.8) we find that the algorithm did not give unique estimates.

To address some of these problems and to limit the number of parameters to be estimated, we can parameterize the error covariances:

$$\mathbf{Q} = \sum_{k=1}^{k=K} \alpha_k \mathbf{Q}_k, \quad \mathbf{R} = \sum_{k=1}^{k=L} \alpha_{K+k} \mathbf{R}_k. \tag{2.35}$$

The exact forms of $\mathbf{Q}_k$ and $\mathbf{R}_k$ are specified based on the physical understanding of the model and the measurements. The number of parameters $K$ is much smaller than the number of elements in **Q**, but in principle we can take delta matrices (all zeroes except for one element) as the basis. In that case the algorithm is identical to the original MT algorithm. Ideally, the adaptive method should return error bars on the vector $\boldsymbol{\alpha}$. Unexpectedly large (or small) error bars could signal wrong parametrization, at which point, one would need to change the parametrization (2.35), and rerun the algorithm. It is important to note that the MT algorithm does not provide such error bars, and is one of the reasons why we develop a covariance matching algorithm in Chapter 3.

53

An alternative is then to estimate the coefficients $\alpha$ instead of the full matrix $\tilde{Q}_{mt}$ at every step of the MT algorithm. This can be done by finding the parameter vector which minimizes the difference between the MT estimate and the parametrization (2.35). The updated estimate of the parameter vector $\alpha$ provides an estimate of the error covariances. Note that one can guard against numerical problems, such as non-positive definiteness, by suitably limiting the search space for parameters $\alpha$.

Despite all of the problems discussed above the MT algorithm is appealing. First, the additional cost relative to the cost of the Kalman filter is insignificant. It is very simple and intuitive. In BFC97 the empirical estimate of MT is shown to give results similar to the maximum likelihood method of Dee (1995) (Section 2.9) which is much more computationally expensive, but has been applied in a number of studies, e.g. Dee and da Silva (1997). In addition, it is proven in BFC97 that the MT algorithm is identical to the maximum likelihood estimator of Maybeck (1982), under the same set of assumptions as above.

We investigate convergence and stability properties of the MT method by analyzing a scalar model (Section 2.6). In Section 2.7 we present analysis of the MT algorithm for the model with two degrees of freedom, and show that the algorithm behaves differently depending on the choice of observational network, state model, and error covariances. We then extend these results to a real case through a series of twin experiments with the linear model described in Section 2.3.

## 2.6 Derivation of the Myers and Tapley Algorithm with a Scalar Model

To introduce the MT method, we start by applying the method to a scalar model, i.e. a model with only one DOF[2]. Evidently, when the observations are few or poor we should not be able to obtain good estimates of the model statistics. The results may depend

---

[2]This derivation is original.

on the type of observations, i.e. on the structure of the observation matrix $\mathbf{H}$, and the model $\mathbf{A}$. However, quantifying these intuitive concepts is not simple because adaptive algorithms are non-linear. For example, simply increasing the number of observations or the length of the time series does not guarantee that adaptive algorithms produce reliable estimates. A scalar model oversimplifies the analysis a great deal, but it can be solved analytically and provides useful guidance for the more complicated examples discussed in the following sections.

For the scalar case, all matrices become scalars, but the notation remains the same. We assume that the state-transition matrix $A$ and covariances $Q$ and $R$ are time-invariant, that the direct observation of the state is available at every time step and that the measurement matrix $H = 1$. Note that we can always rescale variables to make $H = 1$. To summarize, the model and the measurement equations are given by

$$p(t+1) = Ap(t) + u(t), \quad u(t) \sim N(0, Q) \tag{2.36}$$

$$y(t) = p(t) + r(t), \quad r(t) \sim N(0, R), \tag{2.37}$$

and $u(t) \sim N(0, R)$ denotes that a variable $u(()t)$ comes from a normal distribution with mean 0 and variance $Q$. The state transition matrix $A$ represents an estimate of the dynamical model – a scalar in this simplest case. It may be, and often is, different from the true transition matrix $\hat{A}$ of the physical system. Because the observations are not perfect, $R \neq 0$, we can define new variables

$$q \equiv Q/R, \quad \Pi^f(t) \equiv \mathbf{\Pi}(t|t-1)/R, \quad \Pi^a(t) \equiv \mathbf{\Pi}(t|t)/R, \tag{2.38}$$

where $\mathbf{\Pi}(t|)$ denotes the uncertainty of the KF forecast, and $\mathbf{\Pi}(t|t)$ denotes the uncertainty of the KF analysis; see section 2.4.

The equations for the uncertainties of the forecast and the assimilated state, given in full form in Section 2.4, reduce to scalar equations:

$$\Pi^f(t+1) \quad = \quad A^2\Pi^a(t) + q \tag{2.39}$$

$$\Pi^a(t+1) \quad = \quad \Pi^f(t+1)/\left(\Pi^f(t+1) + 1\right) \tag{2.40}$$

The Kalman gain $K$ is given by

$$K(t) = \Pi^f(t)/(\Pi^f(t) + 1) \tag{2.41}$$

The matrices $Q$ and $R$ represent our best *a priori* guesses of the variance of the system and measurement noise, respectively. Note that for a stable model ($|A| < 1$) both gains are always less than 1 in absolute value. Because the scalar system as defined in equations (2.36)-(2.37) is both controllable and observable the existence of a steady-state limit is guaranteed; see Fukumori et al. (1993). It is achieved after very few time steps of the Kalman filter and we use the steady-state filter approximation. Solving the system of equations (2.39) -(2.40) we obtain

$$\Pi^a{}_S = \left( (A^2 - 1) - q + \sqrt{(q + (A^2 - 1))^2 + 4q} \right)/(2A^2), \tag{2.42}$$

$$K_S = \frac{\left( (A^2 - 1) + q + \sqrt{(q + (A^2 - 1))^2 + 4q} \right)}{\left( (A^2 + 1) + q + \sqrt{(q + (A^2 - 1))^2 + 4q} \right)}, \tag{2.43}$$

where subscript $s$ denotes steady-state estimates. Note that the steady-state uncertainties and the Kalman gain depend only on the model $A$ and the ratio $q$.

For the case when the mean of the system and the mean of the measurement noise are assumed to be known and equal to zero, an estimate of the system noise is given by a product of the Kalman gain and innovation vector (Section 2.5)

$$\tilde{u}(t) = K_S v(t). \tag{2.44}$$

For the scalar case the innovation sequence $v(t)$ can be easily evaluated given observations $\mathbf{y}(t)$:

$$v(t) = \sum_{k=1}^{k=t} \lambda_{t,k} \mathbf{y}(k), \tag{2.45}$$

$$\text{where} \quad \lambda_{t,k} = 1, k = t;$$

$$\lambda_{t,k} = -AK_S((1 - K_S)A)^{t-k-1}, \ k \neq t.$$

For future use, we now evaluate the sum of squares of the innovation sequence over $S$,

the length of MT averaging window, time steps:

$$\sum_{t=1}^{S} v(t)^2 = \beta_0 \sum_{k=1}^{S} y(k)^2 + 2\beta_1 \sum_{k=2}^{S} y(k)y(k-1) + 2\beta_2 \sum_{k=3}^{S} y(k)y(k-2) + \dots,$$

$$\beta_0 = \sum_{k=1}^{S} \lambda_{S,k}^2 = 1 + A^2 K_S^2 + A^4 K_S^2 (1 - K_S)^2 + \dots, \qquad (2.46)$$

$$\beta_1 = \sum_{k=2}^{S} \lambda_{S,k} \lambda_{S,k-1} = -AK_S + A^2 K_S^2 (1 - K_S) + \dots, \quad \text{and so on.}$$

We can assume that $S$ is sufficiently large, so that the lower index of the summation can be kept the same. Because for a stable model the terms $\beta_0 > \beta_1 > \beta_2....$ are decreasing in magnitude and coefficients $\lambda_{t,k}$ are rapidly decreasing as $(t - k)$ increases we can neglect terms of higher order in equation (2.46).

Next, we obtain an expression for the statistics of observations by using the true model parameters. To do this we assume that the time series are sufficiently long, and the sampling error is negligible. Using equations (2.36) and (2.37) and the assumptions of serial and mutual independence of the system and measurement noise, we derive

$$< p(t)^2 > = \hat{A}^2 < p(t)^2 > + \hat{Q}, \qquad (2.47)$$

$$< y(t)^2 > = < p(t)^2 > + \hat{R}.$$

Using assumptions of stationarity we replace expectations by a sample estimate $\frac{1}{S} \sum_{k=1}^{S} \cdot$. Next, we obtain an estimate of the sum of squares of the observations:

$$\frac{1}{S} \sum_{t=1}^{S-1} y(t)^2 = \hat{R}(1 + \hat{q}/(1 - \hat{A}^2)), \text{where} \quad \hat{q} = \hat{Q}/\hat{R}. \qquad (2.48)$$

In a similar fashion we find that

$$\frac{1}{S} \sum_{t=k+1}^{S-1} y(t)y(t-k) = \hat{R}\hat{q}\hat{A}^k/(1 - \hat{A}^2)) \qquad (2.49)$$

Thus, when $S$ is large and the assumptions of independence are valid, the sum of the squares of the innovation vectors becomes

$$\frac{1}{S} \sum_{t=k+1}^{S-1} v(t)^2 \simeq \beta_0 \hat{R}(1 + \hat{q}/(1 - \hat{A}^2)) + 2\beta_1 \hat{R}\hat{q}\hat{A}/(1 - \hat{A}^2)) + \dots. \qquad (2.50)$$

57

We are now ready to proceed with adaptive estimation of the system noise variance $Q$. In the scalar case only the ratio $q$ can be determined. Therefore, we cannot determine model and measurement error variances separately, only the ratio of the two. This is true about any innovation based approach: only the ratio of the norms, $|Q|/|R|$, can be determined, but not the norms themselves. Substituting expressions above into the equation for MT estimate (2.32), the sample estimate of $q$ is given by

$$
\begin{aligned}
\tilde{q}_{\mathrm{mt,biased}} \quad &= R^{-1}\mathrm{Cov}(\tilde{u}(t)) = K_{\mathrm{S}}^2 R^{-1} \frac{1}{S} \sum_{t=k+1}^{S-1} v(t)^2 \\
&= K_{\mathrm{S}}^2 \left( \beta_0 \gamma(1 + \hat{q}/(1 - \hat{A}^2)) + 2\beta_1 \gamma \hat{q} \hat{A}/(1 - \hat{A}^2) + ... \right) \quad (2.51)
\end{aligned}
$$

where $\gamma \equiv \hat{R}/R$ is the misspecification of the measurement noise variance, and $\beta_0, \beta_1, ...$ are defined in equation (2.46).

When the *a priori* $Q$ is equal to $\hat{Q}$ this estimate can be shown to be statistically biased by $(A^2 - 1)\Pi^a{}_{\mathrm{S}}$, i.e. on average it produces an estimate greater than the true one by the value of the bias term. It is important to realize that when the *a priori* $Q$ is not equal to the true one the bias term expression is no longer correct. That is, if we use a mis-specified $Q$, on average the estimate minus the bias term is not equal to the true one. Nonetheless, we use the bias correction term as it stands above, and demonstrate below that in this setup it leads to a unique and convergent estimate. The nominally unbiased estimate is equal to

$$
\tilde{q} = K_{\mathrm{S}}^2 \left( \beta_0 \gamma(1 + \hat{q}/(1 - \hat{A}^2)) + 2\beta_1 \gamma \hat{q} \hat{A}^l/(1 - \hat{A}^2) + ... \right) - (A^2 - 1)\Pi^a{}_{\mathrm{S}} \quad (2.52)
$$

Substituting from equations (2.42), (2.43), and (2.46) we obtain that the MT posteriori estimate depends only on the prior estimate of $q$, estimate of the model $A$, $\hat{A}, \hat{q}$, and the misspecification of the measurement variance $\gamma \equiv \hat{R}/R$.

To summarize, we have obtained an analytical approximation for the MT adaptive algorithm for estimation of the error statistics which for a scalar model depends only on four parameters $A$, $\hat{A}, \hat{q}$, and $\gamma$. Thus, to understand the adaptive algorithm we turn to the equation (2.52).

## 2.6.1 Analysis of MT Algorithm with a Scalar Model

We are interested in the behavior of the function

$$\tilde{q} = \tilde{q}(q; A, \hat{A}, \hat{q}, \gamma) \tag{2.53}$$

for different choices of parameters. For the analysis we only consider the case of stable model, $|\hat{A}| < 1$, although our estimates of the model, $A$, can be both stable and unstable. The adaptive algorithm is equivalent to the following procedure:

$$\tilde{q}_1 = \tilde{q}(q_0; A, \hat{A}, \hat{q}, \gamma),$$

$$\tilde{q}_2 = \tilde{q}(\tilde{q}_1; A, \hat{A}, \hat{q}, \gamma), \tag{2.54}$$

$$\tilde{q}_3 = \tilde{q}(\tilde{q}_2; A, \hat{A}, \hat{q}, \gamma), \quad \text{and so on.}$$

That is, running the algorithm for $A = \hat{A} = 0.9$ and *a priori* estimate of $q = 5$, we obtain $\tilde{q}_1 = 2$. Using this estimate as the next *a priori* guess for $q$ we obtain $\tilde{q}_2 = 1.4$, and then $\tilde{q}_3 = 1.2$, and so on. Finally, we converge to an estimate of $q = 1$, which is equal to the true value. Figure 2.6 illustrates this recursive process graphically.

The value to which the adaptive algorithm converges is given by a solution of the equation

$$\tilde{q}(q; A, \hat{A}, \hat{q}, \gamma) - q = 0. \tag{2.55}$$

The contour plot of $\tilde{q}$ as a function $q$ and $A$ for a case when we have a correct estimate of $R$, i.e. $\gamma = 1$, the true model is $\hat{A} = 0.9$, and the ratio $\hat{q}$ is 1, is shown on Figure 2.7 (solid lines). It shows posterior estimates of $\tilde{q}$ for estimates of $A$ in the range from 0 to 2 (corresponding to both stable and unstable models), and the prior ratio $q_0$ in the range from 0 to 10 (corresponding to cases with system noise variance less than measurement noise variance and vice versa). For this system, the adaptive algorithm produces a correct estimate of the system noise variance when we have a correct estimate of the measurement noise variance and the model. This result holds in general for any stable model when $A = \hat{A}$ and $\gamma = 1$. This is shown in Appendix B. It has to be stressed again that these
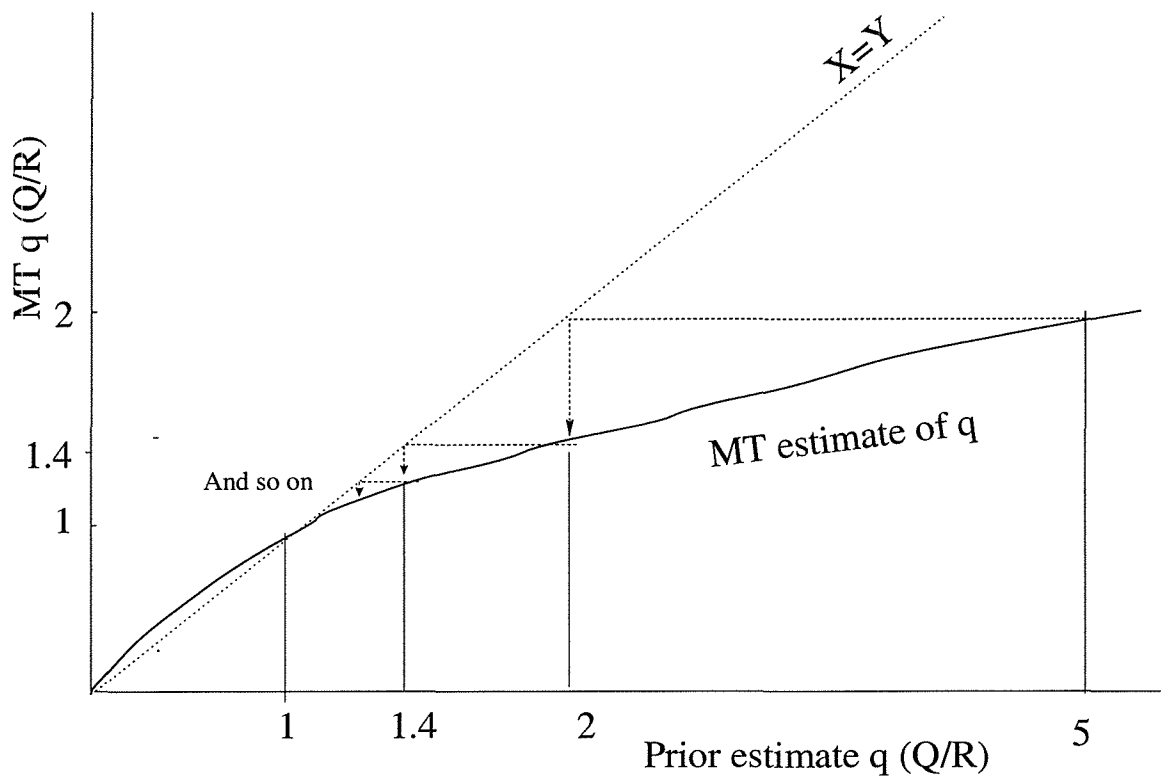
Figure 2.6: A graphical representation of the MT adaptive algorithm. The thick continuous line represents the MT adaptive estimate for different choices of initial $q$. The dotted line is the line $x = y$, used to project the values of the function back onto the abscissa. We start with initial estimate of $q = 5$. After first run of the algorithm we obtain that $q = 2.0$. We then project it back on the $x$-axis by tracing the estimate with a dashed line, and use it as a new prior estimate, and so on. It converges to 1, the true value of $q$.

results are correct only when we have a sufficiently large number of measurements so that the analytical approximations for the covariances, equations (2.48-2.49), remain valid.

The solution to equation (2.55) for all possible choices of the model $A$ is given by a dashed line in Figure 2.7. It demonstrates that the resulting estimate is not very sensitive to the choice of the model used in the analysis: if we use $A = 0.8$, instead of the true value of 0.9, the estimated $\tilde{q}$ is different only by 15 %. We also see that there is a unique estimate for each choice of the model $A$ independent of the initial guess. Thus, when we do not have perfect knowledge of the model the algorithm produces a unique, but wrong, estimate.
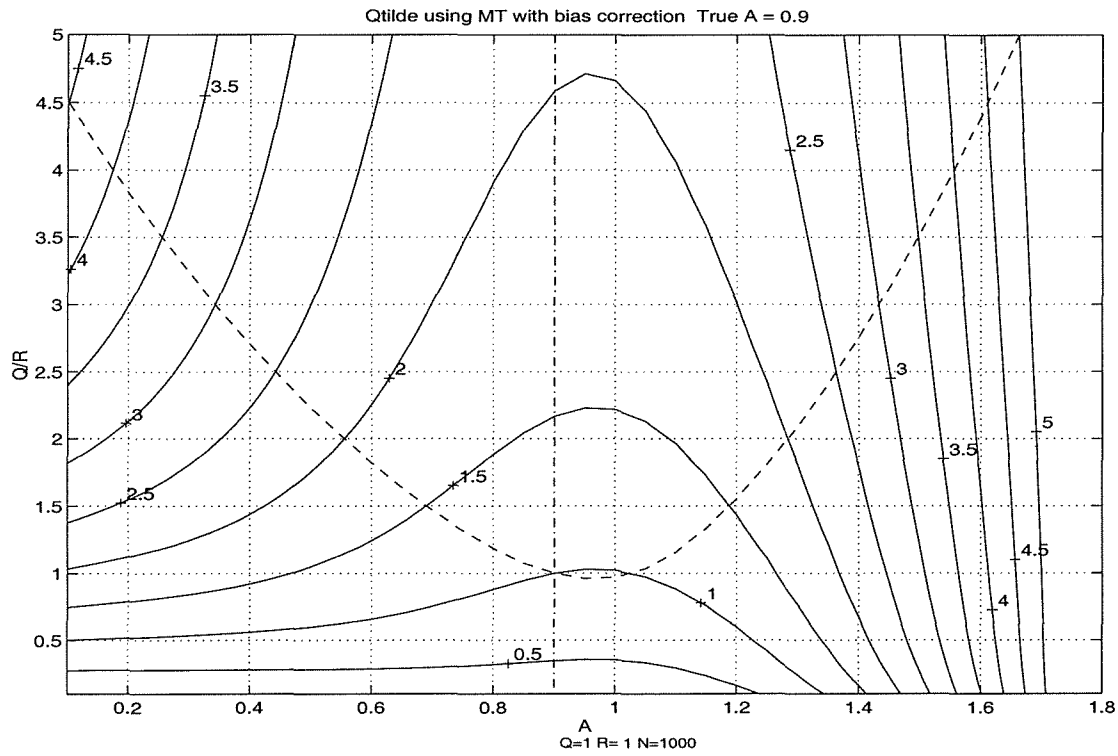
Figure 2.7: Contour plot of $\tilde{q}$ as a function $q$ and $A$ for a case when we have a correct estimate of $R$, the true model is $\hat{A} = 0.9$ (dash-dotted vertical line), and the ratio $\hat{q}$ is 1. Dashed line represents the values to which the algorithm will converge depending on the choice of the model $A$.
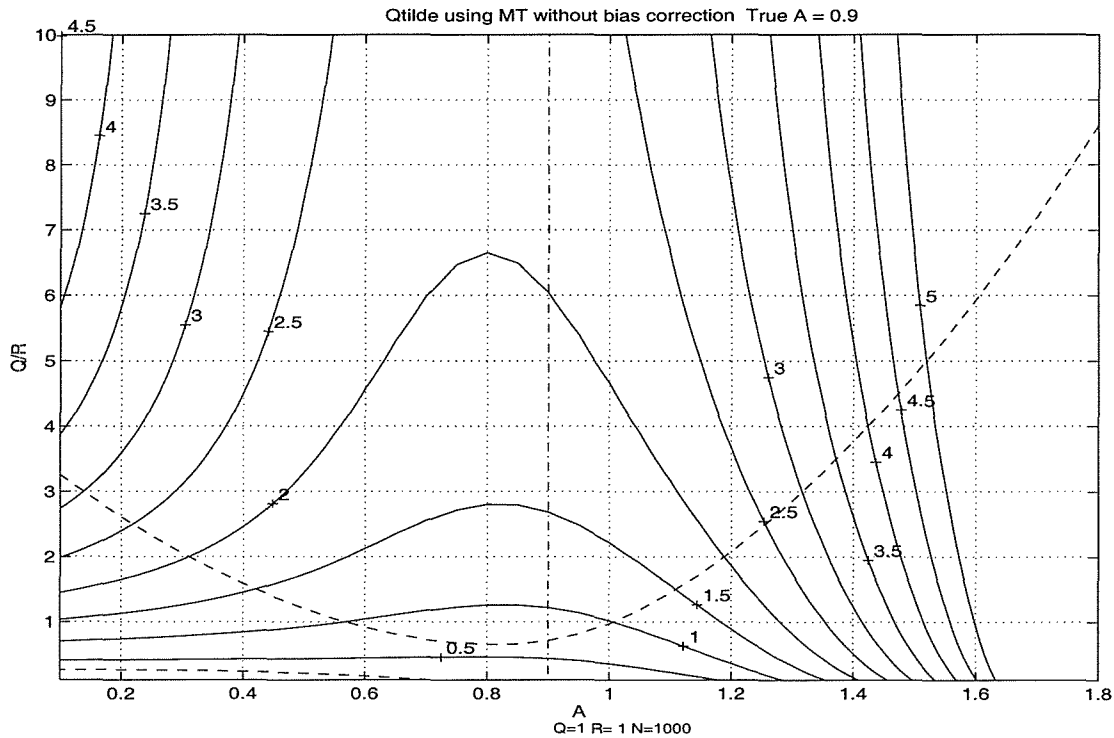
**Figure 2.8:** The contour plot of $\tilde{q}_{\mathrm{mt,biased}}$, without the bias correction, as a function $q$ and $A$ for a case when we have a correct estimate of $R$, the true model is $\hat{A} = 0.9$ (dash-dotted vertical line), and the ratio $\hat{q}$ is 1. Dashed lines represent the values to which the algorithm will converge depending on the choice of the model $A$.

For comparison we show a similar plot for the estimate without the bias correction, i.e. $\tilde{q}_{\mathrm{mt,biased}}(q)$ given in equation (2.51); see Fig. 2.8. The dashed line shows what values the adaptive algorithm converges to when the bias is not accounted for. The results are drastically different. First, the algorithm does not produce a correct estimate of $q$. Second, for slightly different estimates of the model $A$ it either does not converge or it has more than one solution. This again shows that correction for the statistical bias is essential in MT adaptive algorithm.

### 2.6.2  Summary for a MT Algorithm with a Scalar Model

We presented the analysis for a simplest possible case, a scalar model and a scalar observation. This is the only case for which an analytical representation of the MT adaptive algorithm is known to exist. The results suggest that the adaptive MT algorithm is well-behaved if we have a long time series of observations after the correction for the statistical bias. The following results are valid for any stable, i.e. $|\hat{A}| < 1$, model. The adaptive estimate of the system noise variance converges to its true value when we have perfect knowledge of the model, $A$, and the measurement noise variance, $R$. When we do not know the two latter quantities perfectly, we still obtain a unique estimate of $Q$ irrespective of the initial guess. This estimate is close to the true one for sufficiently large misspecification either of the model or of the measurement noise variance. If these conclusions apply to higher-dimensional models for short time series and a general measurement matrix, the MT algorithm is of great practical importance for optimal data assimilation. The scalar case lacks many important characteristics present in multidimensional models. For example, we cannot consider a case of non-identity observation matrix $\mathbf{H}$. Thus, we cannot compare different observation networks, such as altimetric and tomographic ones. In addition, we cannot look at effects of misspecification in cross-correlations in the model and/or observation error, the off-diagonal elements. This is important because very often non-diagonal elements in $Q$ and $R$ are neglected. Thus, though the scalar case provides much hope for the adaptive method it does not shed any light onto many important questions and we have to reserve to numerical experiments.

## 2.7  Derivation of the MT Algorithm for Systems with Several DOF

In this section we generalize analytical results obtained for a scalar model to a two DOF model. This case retains all the essential features of adaptive data assimilation with large

63

models, but is much simpler to analyze. The approach mimics that developed above for the scalar model (Section 2.6). However, we cannot derive closed form solutions. The difficulty is that we need to solve the algebraic Ricatti equation (to obtain the steady Kalman filter, Anderson and Moore, 1979 , p. 155) and the Lyapunov equation (to obtain the model state covariance as a function of the model error covariance, Anderson and Moore, 1979, p. 62 ). Neither the Ricatti equation nor the Lyapunov equation has a closed form solution (except in the scalar case) and we solve them numerically. Similarly to the scalar case, the method described below is identical to running a twin experiment, with sufficiently large number of observations (for 2 DOF model we need at least 500 observations).

Based on the scalar model analysis presented above, we know that the adaptive algorithm can be viewed as a function which has several inputs and outputs. When we have a sufficiently large number of observations, the statistics of the observations can be deduced from the model and measurement equations, and thus the number of parameters is significantly reduced. Namely, instead of providing the time series of observations we specify the true covariances of the model and measurement error to compute statistics of the observations.

The number of parameters is significantly increased from that of the scalar case. When we consider only diagonal $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$ for a two degrees of freedom model, the number of parameters is greater than 10, as compared to 4 for the scalar case. Unlike the scalar case, we need to consider a non-identity observation matrix, and below we show that convergence properties of the algorithm depend on the observation matrix. The adaptive estimates depend on the true model, true model and measurement error covariances, the prior model, prior model and measurement error covariances, and the observation matrix:

$$\tilde{\mathbf{Q}}_{\mathrm{mt}} = \mathrm{function}(\hat{\mathbf{A}}, \hat{\mathbf{Q}}, \hat{\mathbf{R}}, \mathbf{A}, \mathbf{Q}, \mathbf{R}, \mathbf{H}). \tag{2.56}$$

To obtain this function we proceed as in the scalar case. Because we are dealing with non-commuting matrices the results of the scalar case are not directly transferable and we present a derivation below. Combining the model and the measurement equations

64

for the true state, we can obtain estimates of the observation statistics, i.e. covariance and the lag-covariances of the observations. To simplify the notation, we assume that the model error is given on the same grid as the model state, and that there is no deterministic forcing (equation 2.18). Multiplying both sides of (2.18) by $\mathbf{p}(t+1)$ and taking expectations, we obtain the discrete Lyapunov equation,

$$\hat{\mathbf{P}} = \hat{\mathbf{A}}\hat{\mathbf{P}}\hat{\mathbf{A}}^\mathrm{T} + \hat{\mathbf{Q}}, \tag{2.57}$$

which relates the covariance of the state, $\hat{\mathbf{P}}$, to that of the true system error, $\hat{\mathbf{Q}}$. We have assumed that the true state is stationary, i.e. $\hat{\mathbf{P}}(t+1) = \hat{\mathbf{P}}(t) = \hat{\mathbf{P}}$. For stable $\hat{\mathbf{A}}$, the Lyapunov equation is readily solved for $\hat{\mathbf{P}}$ using any of a number of iterative schemes. Similarly, multiplying both sides of (2.37) by $\mathbf{y}(t)^\mathrm{T}$ and taking expectations we have,

$$\mathbf{Y} = < [\mathbf{y}(t) - < \mathbf{y}(t) >][\mathbf{y}(t) - < \mathbf{y}(t) >]^\mathrm{T} > = \mathbf{H}\hat{\mathbf{P}}\mathbf{H}^\mathrm{T} + \hat{\mathbf{R}}, \tag{2.58}$$

and by $\mathbf{y}(t+k)^\mathrm{T}$ and taking expectations we have,

$$\mathbf{Y_k} = < [\mathbf{y}(t+k) - < \mathbf{y}(t+k) >][\mathbf{y}(t) - < \mathbf{y}(t) >]^\mathrm{T} > = \mathbf{H}\hat{\mathbf{A}}^k\hat{\mathbf{P}}\mathbf{H}^\mathrm{T}, k \geq 1, \tag{2.59}$$

which relates the measurement covariance, $\mathbf{Y}$, and the measurement lag covariances, $\mathbf{Y_k}$, to the covariance of the state. These equations are analogous to the scalar case equations (2.48) and (2.49).

To find the MT estimate we need to compute the sum of squares of the innovation vector. We use the steady state Kalman filter. To find the steady state Kalman gain we need to solve the algebraic Ricatti equation, which can be easily derived from (2.27-2.28) and (2.30):

$$\mathbf{\Pi}(t+1|t) = \mathbf{A}\left[\mathbf{\Pi}(t|t-1) - \mathbf{\Pi}(t|t-1)\mathbf{H}^\mathrm{T}\left(\mathbf{H}\mathbf{\Pi}(t|t-1)\mathbf{H}^\mathrm{T} + \mathbf{R}\right)^{-1}\right. \tag{2.60}$$
$$\left.\mathbf{H}\mathbf{\Pi}(t|t-1)\right]\mathbf{A}^\mathrm{T} + \mathbf{Q},$$

$$\mathbf{\Pi_s}(-) = \mathbf{\Pi}(t+1|t) = \mathbf{\Pi}(t|t-1), \mathbf{K_S} = \mathbf{\Pi_s}(-)\mathbf{H}^\mathrm{T}\left(\mathbf{H}\mathbf{\Pi_s}(-)\mathbf{H}^\mathrm{T} + \mathbf{R}\right)^{-1}.$$

Subscript $S$ denotes steady state filter. The hats have been dropped as the filter represents actual Kalman filter assimilation, and therefore the matrices $\mathbf{Q}, \mathbf{R}$, and $\mathbf{A}$ represent

our prior estimates instead of the true values. To solve the Ricatti equation (2.60) we use one of many available numerical techniques – the doubling algorithm (Appendix C). The steady state Kalman gain, $\mathbf{K_S}$, depends on our estimates of the prior model, model and measurement covariance matrices ($\mathbf{A}$, $\mathbf{Q}$, and $\mathbf{R}$). Now, we are ready to evaluate the sum of squares of the innovation vector:

$$\frac{1}{S}\sum_{t=1}^{S}\mathbf{v}(t)\mathbf{v}(t)^{\mathrm{T}} = \sum_{k=0}^{S}\mathbf{\Lambda}_k\mathbf{Y}\mathbf{\Lambda}_k^{\mathrm{T}} + \sum_{k=0}^{S}\sum_{l=1}^{S}(\mathbf{\Lambda}_k\mathbf{Y}_\mathrm{l}\mathbf{\Lambda}_{k+l}^{\mathrm{T}} + \mathbf{\Lambda}_{k+l}\mathbf{Y}_\mathrm{l}^{\mathrm{T}}\mathbf{\Lambda}_k^{\mathrm{T}}) \qquad (2.61)$$

$$\text{where} \qquad \mathbf{\Lambda}_0 = \mathbf{I}, \mathbf{\Lambda}_k = -\mathbf{H}\left[\mathbf{A}(\mathbf{I} - \mathbf{K_S}\mathbf{H})\right]^{k-1}\mathbf{A}\mathbf{K_S}. \qquad (2.62)$$

Note that for a stable model $\mathbf{A}$, the terms $\mathbf{\Lambda}_k$ rapidly decrease with $k$, and, in fact, only a few first terms in the series are important. The length of the averaging window, $S$, is assumed sufficiently large to accommodate all significant terms in the series (2.61).

Then, we can readily obtain the MT estimate; see (2.31-2.32),

$$\tilde{\mathbf{Q}}_{\mathrm{mt}} = \mathbf{K_S}(\frac{1}{S}\sum_{t=1}^{S}\mathbf{v}(t)\mathbf{v}(t)^{\mathrm{T}})\mathbf{K_S^{\mathrm{T}}} - \text{bias correction}, \qquad (2.63)$$

where the bias correction is given by

$$\text{bias correction} = \left(\mathbf{A}\mathbf{\Pi_s}\mathbf{A}^{\mathrm{T}} - \mathbf{\Pi_s}\right), \qquad (2.64)$$

and $\mathbf{\Pi_s}$ is the steady state uncertainty of the model update,

$$\mathbf{\Pi_s} = \mathbf{A}\mathbf{\Pi_s}(-)\mathbf{A}^{\mathrm{T}} + \mathbf{Q}. \qquad (2.65)$$

## 2.7.1 Analysis of the MT method with a Two-DOF Model

We can now investigate the properties of the MT method. We need to address several important questions. First, we need to investigate whether recursive application of the algorithm gives the same result independent of the prior guess– the uniqueness property. Second, we need to consider the sensitivity of the solution to the misspecification of parameters, such as wrong prior measurement covariance $\mathbf{R} \neq \hat{\mathbf{R}}$, or wrong state transition matrix, $\mathbf{A} \neq \hat{\mathbf{A}}$. Third, we need to find out whether the solution is convergent, i.e.

equal to the true one when all the assumptions are valid. Certainly, when a solution is not unique it cannot be convergent. We need to do this for different choices of the state transition matrices, observation networks, and true model and measurement error covariances.

We parameterized the model error covariance in the form suggested by Belanger (1974):

$$\mathbf{Q} = \sum_{i=1}^{K} \alpha_i \mathbf{Q}_i, \tag{2.66}$$

where the matrices $\mathbf{Q}_i$ are known. In principle, one can take the number of parameters $\alpha_i$ to be equal to the number of basis elements of $\mathbf{Q}$ ($N(N+1)/2$ for a symmetric matrix of size $N$), as proposed by Shellenbarger (1967). Mehra (1970), provides an analytical limit on the number of parameters which can be estimated by an innovation based adaptive algorithm. He shows that since the Kalman gain depends only on the $N \times M$ linear functions of $\mathbf{Q}$ only that many parameters can be identified:

$$K_{\max} \leq N \times M. \tag{2.67}$$

Since the parameter space is rather large; see equation (2.56), in each series of experiments we only vary a few parameters. We display the results of a series of twin experiments on one figure (Figure 2.9). The left column of plots shows the varying inputs; and the right column of plots shows the outputs, i.e. the elements of the model error covariance. Thus, in this case we apply the MT algorithm to 30 different initial guesses of prior $\mathbf{Q}$. We vary prior $\mathbf{Q}$ from $[1\ 0; 0\ 1]$ to $[9\ 0; 0\ 9]$, keeping the off-diagonal elements at zero. The output, $\tilde{\mathbf{Q}}_{\mathrm{mt}}$, which lies in 3 dimensional space (symmetric 2 by 2 matrix), is plotted on 3 different plots, one for each element. We decided to plot the difference between the MT estimates and the true ones (thick dots) to ease the comparison between different experiments. The reason is that in this section we are primarily interested whether the MT estimates are close to the true ones, and not in the estimates themselves. In addition on the right column of plots we show the difference between the prior and the true elements (pluses) to facilitate comparison of the prior and adaptive estimates of the model error covariance.

EXPER 115: A=Atrue: R=Rtrue RECURSIVE w. 3 param. estim.

Qtrue=(5.000 2.000) Rtrue=(1.000 0.000) Atrue=(0.800 0.150)  Q =(5.000 0.000)  R =(1.000 0.000)  A =(0.800 0.150)  H =(1.000 1.000)
Qtrue=(2.000 5.000) Rtrue=(0.000 1.000) Atrue=(−0.15 0.900)  Q =(0.000 5.000)  R =(0.000 1.000)  A =(−0.15 0.900)  H =(0.000 0.500)

Figure 2.9: MT adaptive algorithm with a 2 DOF model (Section 2.7). The left column of plots shows the varying inputs against the experiment number (thick dots); and the right column of plots shows the outputs, i.e. the elements of the model error covariance against the experiment number. The plots on the right show estimates for c) $\tilde{Q}_{mt}[1,1] - \hat{Q}[1,1]$ (thick dots), and $Q[1,1] - \hat{Q}[1,1]$ (pluses) d) the same for element $[2,2]$, and e) for element $[1,2]$. Note that since a covariance matrix is symmetric these three elements completely define it in the 2 dimensional case. The values of all the necessary matrices are given below the graphs. In this group of experiments $H$ has rank 2, and the prior $Q$ is changed for each of 30 experiments. The MT algorithm gives a correct estimate of the true covariance independent of the prior guess.

68

First, we establish whether the recursive application of the MT method produces unique solution irrespective of the initial choice of the prior covariance $\mathbf{Q}$. We ran a large number of experiments with different choices of parameters, $\hat{\mathbf{A}}, \hat{\mathbf{Q}}$, etc. If we have two distinct observations, i.e. we observe both elements of the state, the algorithm converges to a unique solution. An example is shown in Figure 2.9 when the measurement matrix is $[1\ 1; 0\ 0.5]^3$ . The state transition matrix has non-zero off diagonal elements and eigenvalues inside the unit circle. The left column of plots shows how we change the prior model error covariance $\mathbf{Q}$ for different runs, and right column shows the resulting estimate of the model error covariance $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ relative to the true $\hat{\mathbf{Q}}$. The right plots show that for each initial guess the MT algorithm produces perfect estimates, as the difference between the MT estimate and the true one is zero. The MT estimate is unique, and equal to the true one.

However, when we repeat the same set of experiments with only one observation, namely the sum of the two state elements, the results are different (Figure 2.10). In this case, the MT estimate depends on the prior $\mathbf{Q}$. The estimate is equal to the true one only when the prior is equal to the true one, experiment 30. The reason is that the rank of the product of the Kalman gain, $K_{\mathrm{S}}$, and the observation matrix, $\mathbf{H}$, is less than the rank of the covariance matrix; see Blanchet et al. (1997). In this case, we only tried to estimate two diagonal elements of $\mathbf{Q}$. The off-diagonal elements are set to zero, the true value. That is, at each step of the iteration, the prior covariance is chosen to be diagonal; that is the covariance is parameterized as

$$\mathbf{Q} = \alpha_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.68}$$

Mehra's criterion, equation (2.67), is satisfied, but the estimate is not unique. That is, Mehra's criterion sets the upper limit for the number of parameters in $\mathbf{Q}$ but even with an infinite amount of data it is not guaranteed that many parameters can be estimated.

---

[3]The rows of the matrix are separated by a semi-colon, i.e. $\begin{bmatrix} 1 & 1 \\ 0 & 0.5 \end{bmatrix} \leftrightarrow [1\ 1; 0\ 0.5]$

EXPER 118: A=Atrue: R=Rtrue RECURSIVE w. 2 param. estim.

Qtrue=(1.000 0.000) Rtrue=(1.000 0.000) Atrue=(0.800 0.150)  Q =(1.000 0.000)  R =(1.000 0.000)  A =(0.800 0.150)  H =(1.000 1.000)
Qtrue=(0.000 1.000) Rtrue=(0.000 0.000) Atrue=(−0.15 0.900)  Q =(0.000 1.000)  R =(0.000 0.000)  A =(−0.15 0.900)  H =(0.000 0.000)

Figure 2.10: MT adaptive algorithm with a 2 DOF model (Section 2.7). The left column of plots shows the varying inputs against the experiment number (thick dots); and the right column of plots shows the outputs, i.e. the elements of the model error covariance against the experiment number. The plots on the right show estimates for c) $\tilde{Q}_{mt}[1,1] - \hat{Q}[1,1]$ (thick dots), and $Q[1,1] - \hat{Q}[1,1]$ (pluses) d) the same for element $[2,2]$, and e) for element $[1,2]$. The values of all the necessary matrices are given below the graphs. In this group of experiments there is only one observation which averages the two state elements, and the prior $Q$ is changing. Only the diagonal elements of $Q$ are estimated. Unlike the case with full rank $H$, figure 2.9, the estimate is wrong except when the prior is accidentally equal to the true.

This sensitivity to the initial guess is a weak point of the method. There is no information on what combination of parameters $\alpha$ can be reliably estimated and what is the uncertainty of the estimates. In these experiments, sampling uncertainty of the statistics of observations is assumed to vanish, and all the assumptions are satisfied. In more realistic applications, when this no longer holds, and we do not have a good prior guess, the algorithm may not converge to the true estimate. We ran a number of experiments with different state-transition matrices with similar results.

However, estimation of only one parameter, the mean variance of the model error, is well-behaved (Figure 2.11). In this experiment at each iteration we set the prior covariance to the identity times the mean of the diagonal of the previous iterative estimate:

$$\mathbf{Q} = \alpha_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.69}$$

The MT method provides the correct estimate independently of the prior guess, the difference between the MT estimate and the true one is identically zero. The true covariance matrix has diagonal elements equal to 5 and 7 respectively, and the adaptive estimate of the mean is equal to 6, i.e. it overestimates the first diagonal element $\hat{\mathbf{Q}}[1,1]$, and underestimates the second diagonal element $\hat{\mathbf{Q}}[2,2]$. These experiments suggest that in the perfect case when we have the right model and measurement error covariance, and infinite time series of observations, the algorithm is stable and converges to the true estimate for the mean model error variance independent of the initial guess for $\mathbf{Q}$. However, the number of iterations (we stopped the algorithm when the MT estimate is less than 1 per cent different from the prior), is on the order of 10. This is a large number since we needed very long observational time series; see equation 2.61, to achieve good estimates of statistics of the innovation sequence.

To understand the dependence of the MT algorithm on the measurement matrix $\mathbf{H}$ we ran the following series of experiments: we changed the second diagonal element of $\mathbf{H}$ from 0, as in the latter case above, to 0.1, for which a MT estimate is unique, and kept everything else unchanged, i.e. we added a second measurement to the original

**EXPER 77: A=Atrue: R=Rtrue RECURSIVE w. 1 param. estim.**

Qtrue=(5.000 0.000) Rtrue=(1.000 0.000) Atrue=(0.800 0.150)   Q =(5.000 0.000)   R =(1.000 0.000)   A =(0.800 0.150)   H =(1.000 1.000)
Qtrue=(0.000 7.000) Rtrue=(0.000 0.000) Atrue=(−0.15 0.900)   Q =(0.000 7.000)   R =(0.000 0.000)   A =(−0.15 0.900)   H =(0.000 0.000)
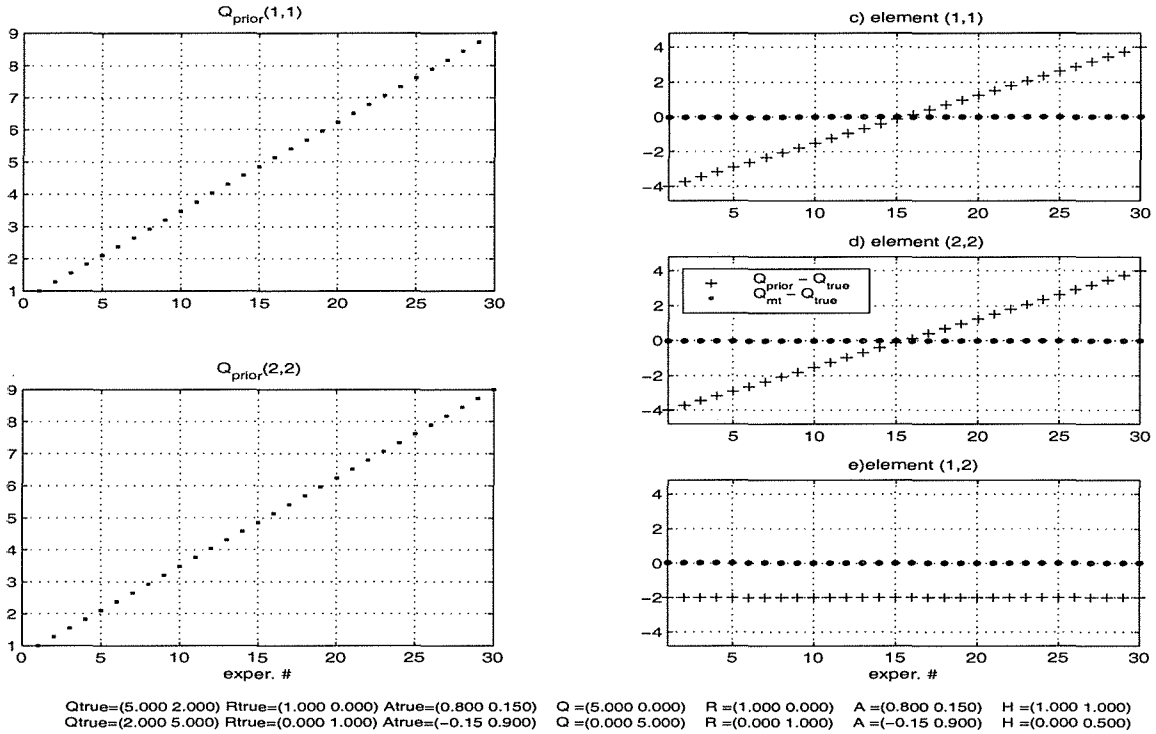
Figure 2.11: MT adaptive algorithm with a 2 DOF model (Section 2.7). The left column of plots shows the varying inputs against the experiment number (thick dots); and the right column of plots shows the outputs, i.e. the elements of the model error covariance against the experiment number. The plots on the right show estimates for c) $\tilde{Q}_{mt}[1,1] - \hat{Q}[1,1]$ (thick dots), and $Q[1,1] - \hat{Q}[1,1]$ (pluses) d) the same for element $[2,2]$, and e) for element $[1,2]$. The values of all the necessary matrices are given below the graphs. In this group of experiments there is only one observation which averages the two state elements, and the prior $Q$ is changing. With the reduced number of parameters to be estimated the algorithm gives a correct estimate of the parameter, the mean of the diagonal of $Q$.

72

averaging measurement of the state. The measurement equation (2.21) can be rescaled, and increasing $\mathbf{H}[2,2]$ is equivalent to reducing uncertainty of the measurement of the second state element. The prior model error covariance $\mathbf{Q}$ is misspecified. We tried to estimate the two parameters as in equation 2.68. The results are displayed in Figure 2.12[4]. The estimates of the model error covariance become closer to the true ones as we increase the value of $\mathbf{H}[2,2]$, or, equivalently, decrease the uncertainty of the second measurement. Therefore, although uniqueness is not guaranteed even with full rank observation matrix, it is achieved when the additional measurement has error variance more than hundred times as much as that of the averaging measurement. The additional information is sufficient to make the algorithm converge to the unique estimate. This unique estimate is equal to the true one.

To study the sensitivity we ran experiments where either the true model transition matrix is different from the model used in the analysis, or the measurement error covariance is misspecified. For example, we show results of experiments where we change both the true and the prior state transition matrices keeping all other parameters the same (Figure 2.13). Namely, we underestimate the diagonal elements of $\hat{\mathbf{A}}$ by 0.1; see the left column of plots. The prior guess for $\mathbf{Q}$ overestimated the variance by a factor of 2. We tried to estimate mean of the diagonal elements of $\mathbf{Q}$, as in equation (2.69). The MT estimate is significantly better than the prior, but is off by as much as 35 percent. In general, the estimates get worse as the eigenvalues of $\hat{\mathbf{A}}$ get closer to one. The results are similar for other choices of the state transition matrix.

In another group of experiments, we misspecified the measurement error covariance (Figure 2.14). The estimates get closer to the truth when the measurement error covariance $\mathbf{R}$ gets larger, and thus misspecification of $\mathbf{R}$ becomes less significant as its weight in the Ricatti equation (2.60) diminishes. Thus, the method is less sensitive to misspecification of $\mathbf{R}$ when the ratio of $|\mathbf{Q}|/|\mathbf{R}|$ is is smaller.

To check the sensitivity of the MT method to off-diagonal elements, we ran exper-

---

[4]The number of iterations of the MT algorithm was limited to 300.

EXPER 66: A=Atrue: R=Rtrue RECURSIVE w. 2 param. estim.

Qtrue=(5.000 0.000) Rtrue=(1.000 0.000) Atrue=(0.800 0.150)  Q =(2.000 0.000)  R =(1.000 0.000)  A =(0.800 0.150)  H =(1.000 1.000)
Qtrue=(0.000 5.000) Rtrue=(0.000 1.000) Atrue=(−0.15 0.900)  Q =(0.000 2.000)  R =(0.000 1.000)  A =(−0.15 0.900)  H =(0.000 0.000)
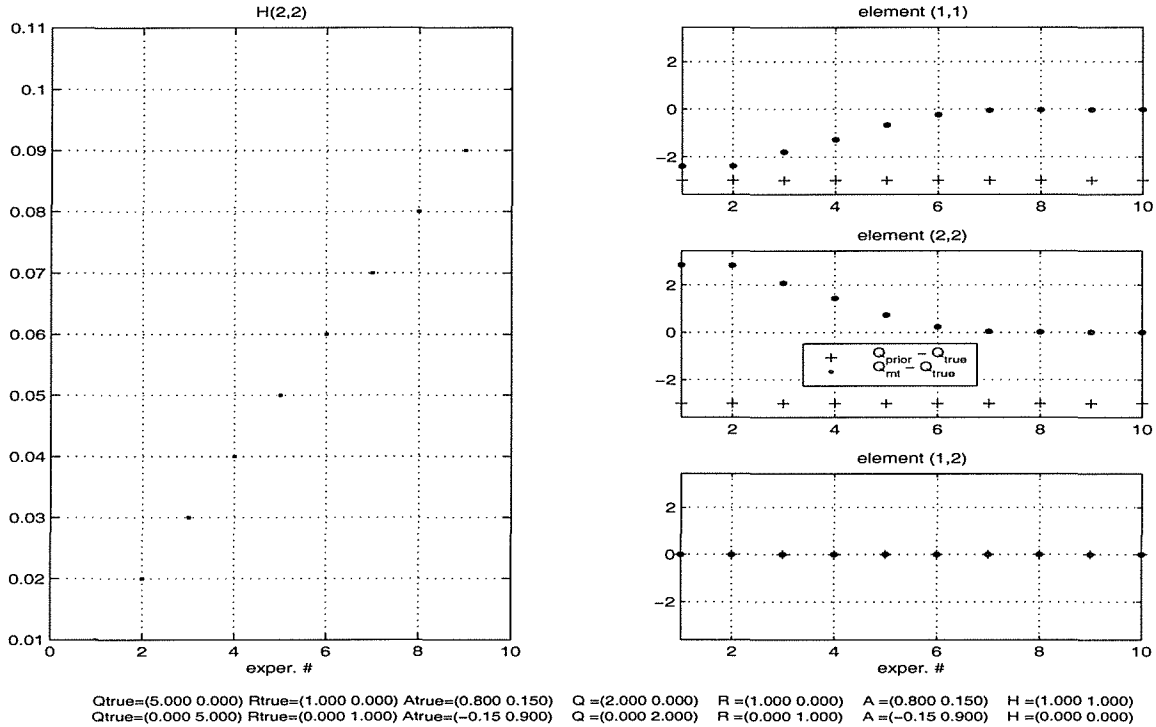
Figure 2.12: MT adaptive algorithm with a 2 DOF model (Section 2.7). The left column of plots shows the varying inputs against the experiment number (thick dots); and the right column of plots shows the outputs, i.e. the elements of the model error covariance against the experiment number. The plots on the right show estimates for c) $\tilde{Q}_{mt}[1,1]-\hat{Q}[1,1]$ (thick dots), and $Q[1,1] - \hat{Q}[1,1]$ (pluses) d) the same for element $[2,2]$, and e) for element $[1,2]$. The values of all the necessary matrices are given below the graphs. In this group of experiments the observation matrix $H$ is changing and the prior $Q$ is kept constant. When we add a second observation which measures the second element of the state to the averaging observation, the algorithm starts to converge to a unique estimate. However, if the uncertainty of such a measurement is great, i.e. $H[2,2] \ll 1$ the algorithm needs a very large number of iterations to converge. Because we limited the number of iterations to 300, it cannot provide the correct estimate for the first 7 experiments.

74

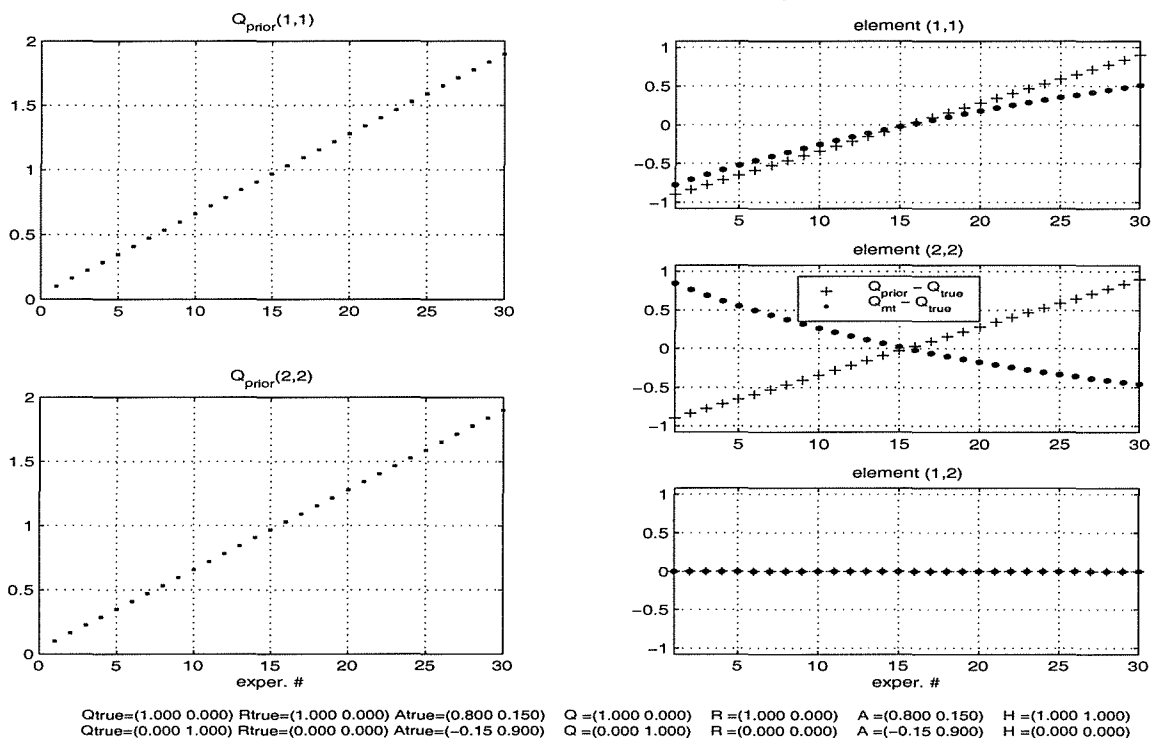EXPER 109: R=Rtrue RECURSIVE w. 1 param. estim.
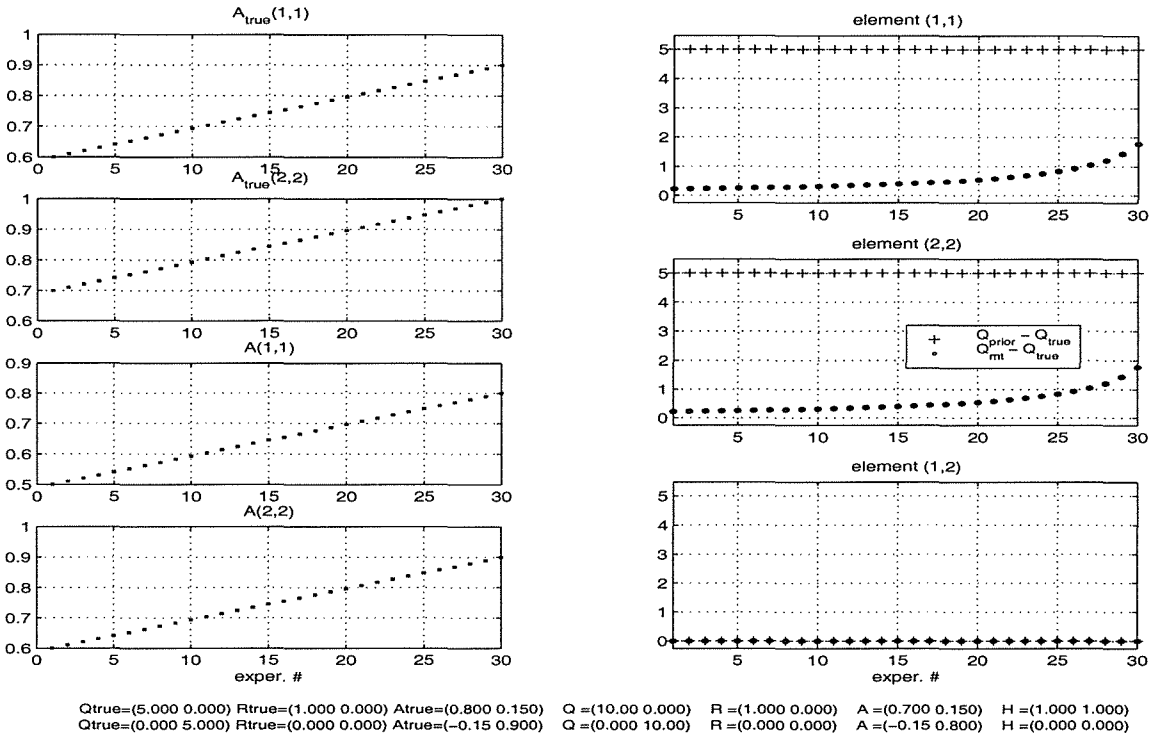
Figure 2.13: MT adaptive algorithm with a 2 DOF model (Section 2.7). The left column of plots shows the varying inputs against the experiment number (thick dots); and the right column of plots shows the outputs, i.e. the elements of the model error covariance against the experiment number. The plots on the right show estimates for c) $\tilde{Q}_{mt}[1,1] - \hat{Q}[1,1]$ (thick dots), and $Q[1,1] - \hat{Q}[1,1]$ (pluses) d) the same for element $[2,2]$, and e) for element $[1,2]$. The values of all the necessary matrices are given below the graphs. In this group of experiments there is one observation which averages the two state elements. The state transition matrix is changing and misspecified. The MT estimates are much closer to the truth than the prior, and get worse with the increasing eigenvalues of $\hat{A}$.

EXPER 110: A=Atrue RECURSIVE w. 1 param. estim.
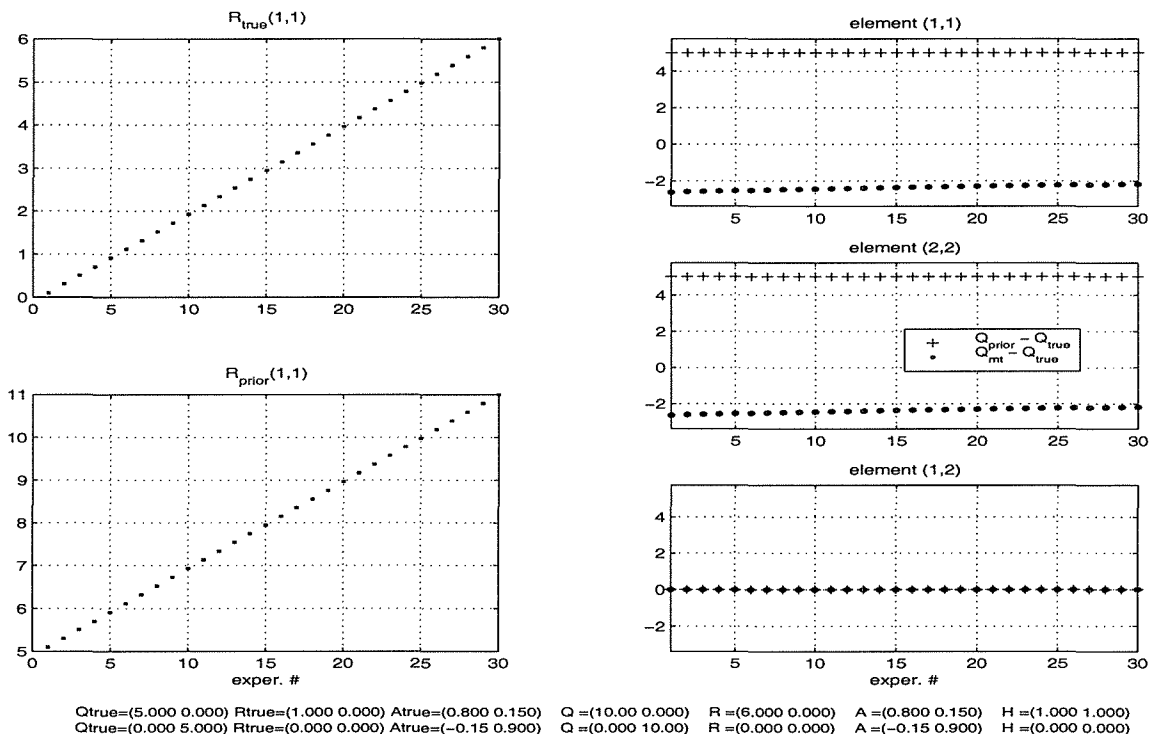
Figure 2.14: MT adaptive algorithm with a 2 DOF model (Section 2.7). The left column of plots shows the varying inputs against the experiment number (thick dots); and the right column of plots shows the outputs, i.e. the elements of the model error covariance against the experiment number. The plots on the right show estimates for c) $\tilde{Q}_{mt}[1,1] - \hat{Q}[1,1]$ (thick dots), and $Q[1,1] - \hat{Q}[1,1]$ (pluses) d) the same for element $[2,2]$, and e) for element $[1,2]$. The values of all the necessary matrices are given below the graphs. In this group of experiments there is one observation which averages the two state elements and changing, and misspecified measurement error covariance matrices $\mathbf{R}$ and $\hat{\mathbf{R}}$. The estimates of $\mathbf{Q}$ are wrong, and get worse with smaller $\mathbf{R}$. This signifies the fact that misspecification of the measurement error is more important when the measurement error is small.

76

iments with a true covariance matrix having non-zero off-diagonal elements, but with a model for the covariance which assumes zero-off-diagonal elements (equations (2.68) and (2.69)). The resulting estimates (Figure 2.15) are wrong, and rather sensitive to the effect of misspecification of the model error covariance model.
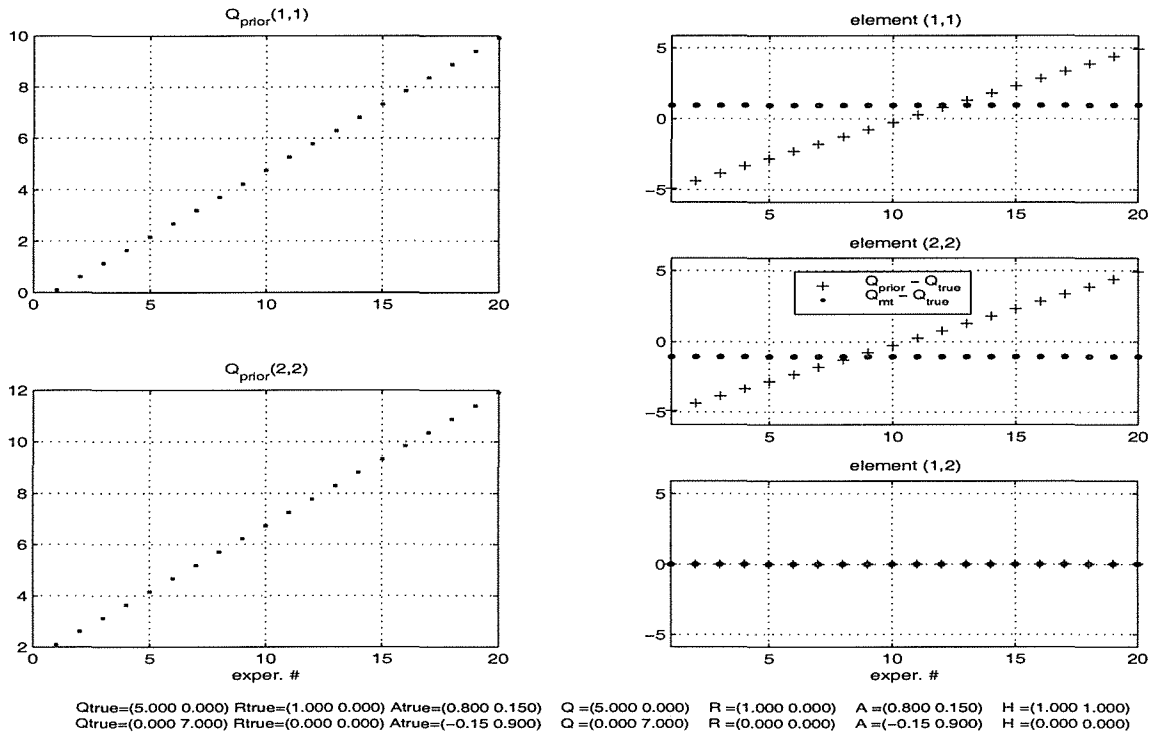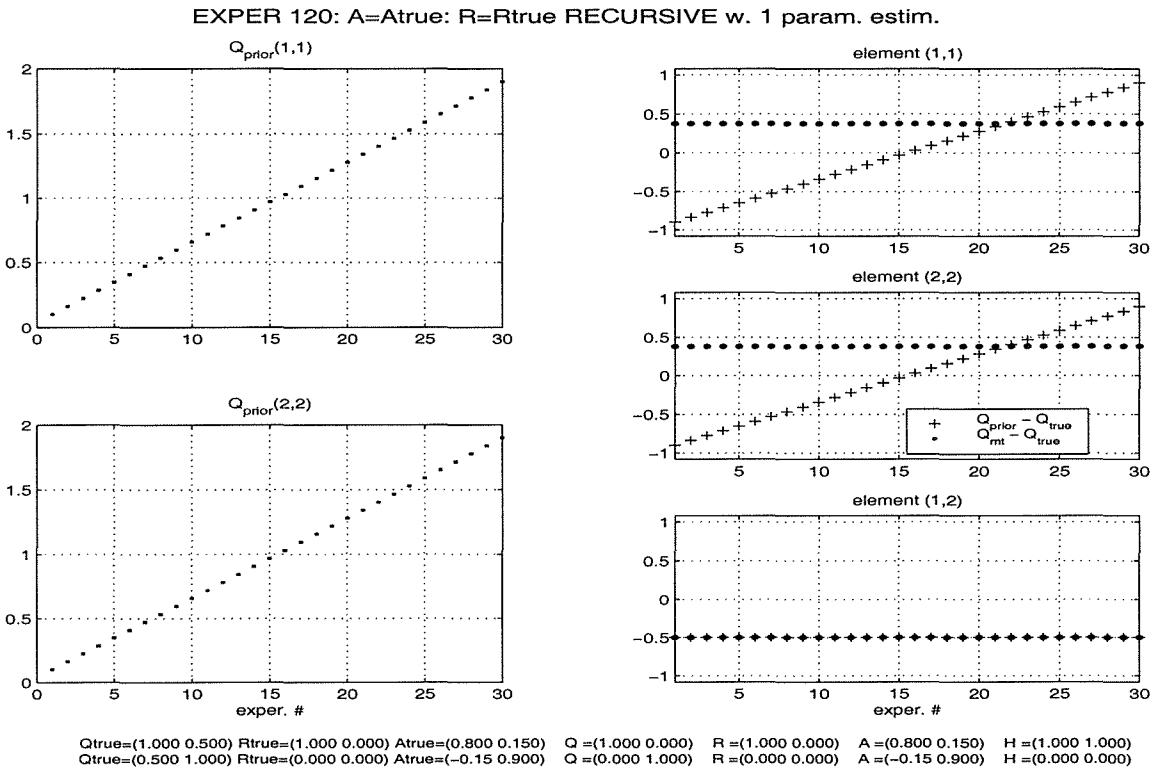


Figure 2.15: MT adaptive algorithm with a 2 DOF model (Section 2.7). The left column of plots shows the varying inputs against the experiment number (thick dots); and the right column of plots shows the outputs, i.e. the elements of the model error covariance against the experiment number. The plots on the right show estimates for c) $\tilde{Q}_{mt}[1,1] - \hat{Q}[1,1]$ (thick dots), and $Q[1,1] - \hat{Q}[1,1]$ (pluses) d) the same for element $[2,2]$, and e) for element $[1,2]$. Note that since a covariance matrix is symmetric these three elements completely define it in the 2 dimensional case. The values of all the necessary matrices are given below the graphs. In this group of experiments $H$ has rank 2, and the prior $Q$ is changed for each of 30 experiments. In addition, the parametrization of the prior $Q$ is wrong (which neglects cross-correlations of the model errors). The estimates are wrong independent of the prior guess for $Q$.

The check for convergence is appropriate only when there is a unique estimate. Thus, based on the discussion above and by running extensive experiments sampling the pa-

77

rameter space we checked that this is the case either when we have two observations or we are trying to estimate a single parameter with one observation, e.g. the mean variance of the model error. An example is shown on Figure 2.16, where we present MT estimates with different choices of state transition matrices. The MT estimates are perfect for all
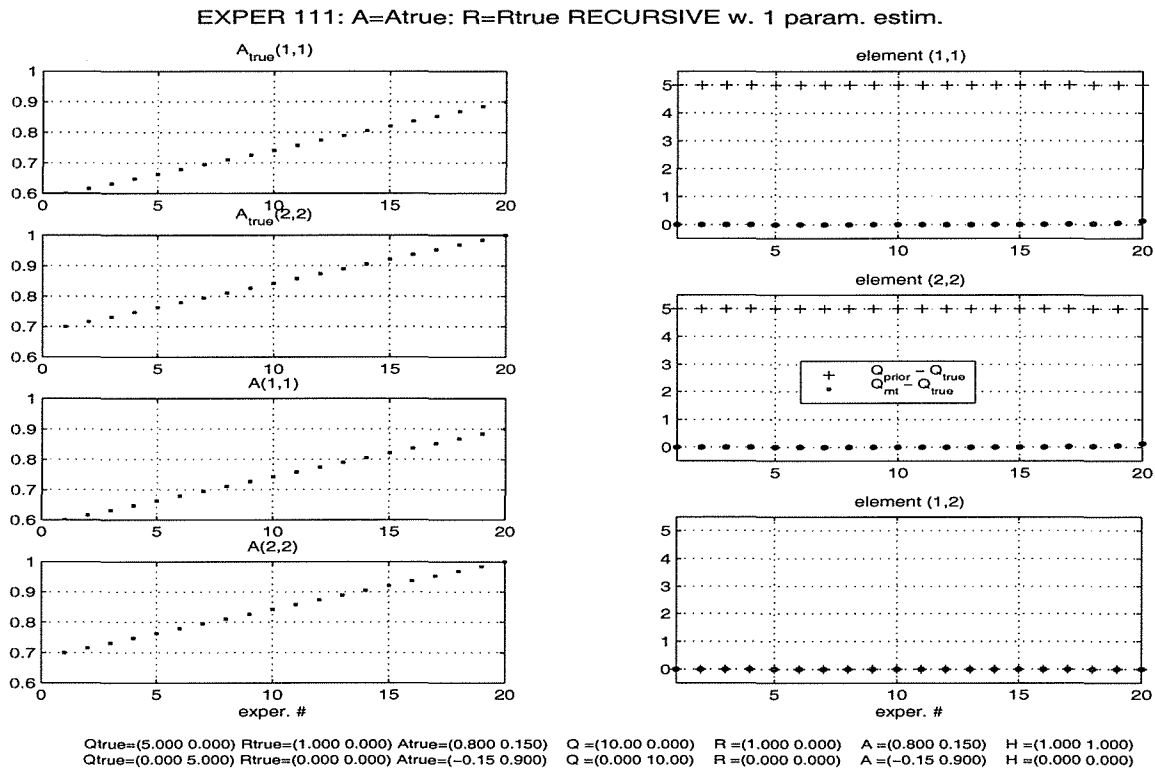


Figure 2.16: MT adaptive algorithm with a 2 DOF model (Section 2.7). The left column of plots shows the varying inputs against the experiment number (thick dots); and the right column of plots shows the outputs, i.e. the elements of the model error covariance against the experiment number. The plots on the right show estimates for c) $\tilde{Q}_{mt}[1,1] - \hat{Q}[1,1]$ (thick dots), and $Q[1,1] - \hat{Q}[1,1]$ (pluses) d) the same for element $[2,2]$, and e) for element $[1,2]$. The values of all the necessary matrices are given below the graphs. With one observation which averages the two state elements and changing, and misspecified, state transition matrix $\mathbf{A}$. Only one parameter is estimated, the mean of the diagonal. The estimate is independent of the prior and equal to the true.

choices of state transition matrices.

### 2.7.2  Summary for The MT method with several DOF Models

We have considered the MT algorithm for a system with two DOF. This allowed us to consider in detail most cases of interest for systems with more than one DOF (i.e. different observational networks, diagonal and off-diagonal covariance matrices, etc). We have considered only the most idealized case when all the assumptions are perfectly satisfied, and we have infinite length time series of observations. Even in this setup, the MT algorithm is sensitive to the initial guess of $\mathbf{Q}$ and the number of parameters estimated, a fact often observed with non-linear algorithms. The limit given by Mehra (1970), is only an upper bound for the number of parameters. Even with infinite time series of observations, the number of parameters that can be reliably estimated is less than that given in equation (2.67). The algorithm produces perfect results when we try to estimate one parameter with one observation, and we have good estimates of the other parameters. The algorithm is sensitive to misspecification in the model and observation error statistics. The algorithm is not efficient, even with perfect knowledge of observational statistics; it takes a lot of iterations for the algorithm to converge.

## 2.8  Twin Experiments with the Linearized MIT GCM and the MT Method

With the details of the linearized MIT GCM (LM) and the data given in Sections 2.1 and 2.2, and the MT method described in Sections 2.5 − 2.7, we turn our attention to twin experiments with the LM. All twin experiments are run in the same fashion. We start by running the model from random initial conditions, and forced stochastically at every time step of the model (1 month for the LM). The stochastic forcing is assumed to be normally distributed in space with zero mean and covariance $\hat{\mathbf{Q}}$ and white in time. This defines the true state of the model. To simulate the observations, we add measurement noise to the projection of the true state onto the measurement space. Measurement noise is
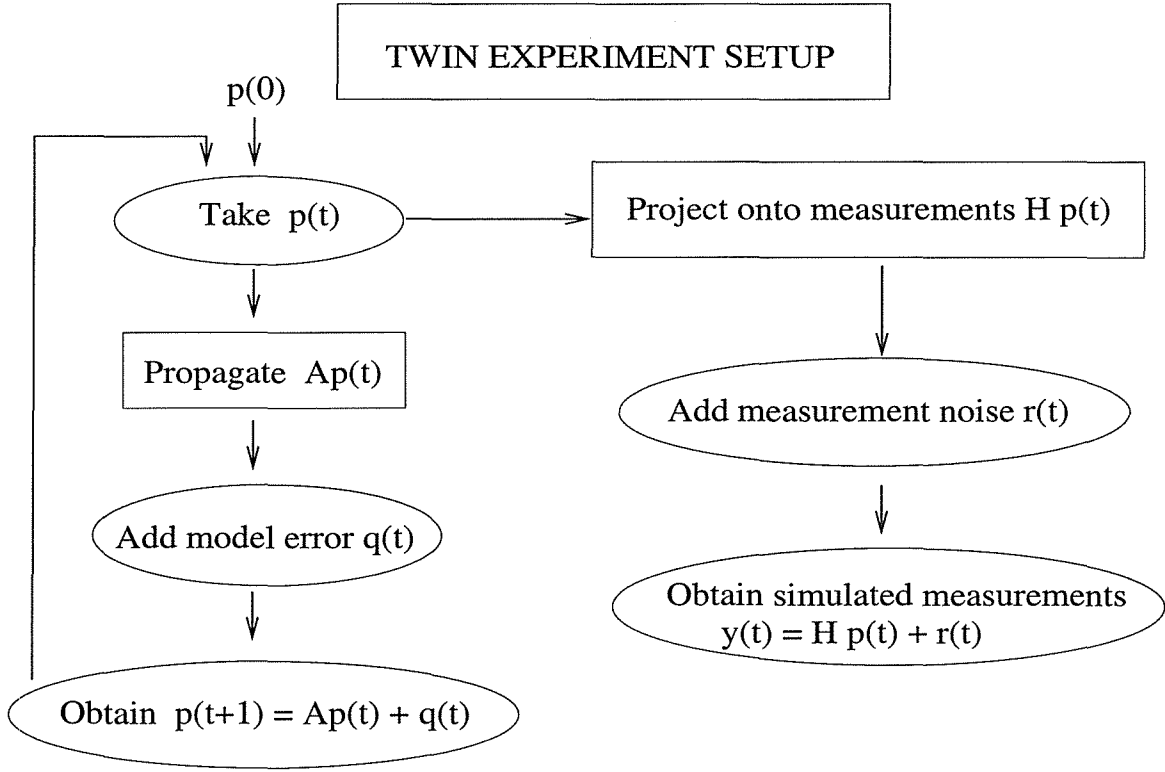
Figure 2.17: Schematic representation of the twin experiment.

assumed to be normally distributed in space with zero mean and covariance $\hat{\mathbf{R}}$ and white in time. These pseudo-observations are then assimilated into the model. The procedure is shown schematically in Figure 2.17.

Rather than trying to estimate the full model error covariance we restrict our attention to the case where the covariance, $\mathbf{Q}$, is block diagonal with four diagonal blocks given by a multiple of the identity:

$$\mathbf{Q} = \begin{bmatrix} \alpha_1 \mathbf{I}_{128} & \mathbf{0}_{128} & \mathbf{0}_{128} & \mathbf{0}_{128} \\ \mathbf{0}_{128} & \alpha_2 \mathbf{I}_{128} & \mathbf{0}_{128} & \mathbf{0}_{128} \\ \mathbf{0}_{128} & \mathbf{0}_{128} & \alpha_3 \mathbf{I}_{128} & \mathbf{0}_{128} \\ \mathbf{0}_{128} & \mathbf{0}_{128} & \mathbf{0}_{128} & \alpha_4 \mathbf{I}_{128} \end{bmatrix}, \quad \text{or} \quad \mathbf{Q} \leftrightarrow [\alpha_1, \alpha_2, \alpha_3, \alpha_4], \qquad (2.70)$$

where $\mathbf{I}_{128}$ is an identity matrix of 128 by 128 and $\mathbf{0}_{128}$ is a zero matrix of 128 by 128. Thus, the problem is reduced to estimating 4 parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, each corresponding to the variance of a particular vertical EOF. The variables are non-dimensional.

The true covariances are given in the same block diagonal form (equation (2.70)). That is, there are no cross-correlations between model errors at different locations. Three choices of true $\hat{\mathbf{Q}}$ are chosen for the experiments,

$$
\begin{aligned}
\hat{\mathbf{Q}}_1 &\leftrightarrow [1, 1, 1, 1], \\
\hat{\mathbf{Q}}_2 &\leftrightarrow [1, 2, 4, 8], \\
\hat{\mathbf{Q}}_3 &\leftrightarrow [8, 4, 2, 1].
\end{aligned} \tag{2.71}
$$

In this very simplified case the correct estimation of the coefficients $\boldsymbol{\alpha}$ is identical to correct estimation of the true covariance matrix. This represents the ideal scenario when our parametrization of the model error covariance matrix is correct. In any practical situation, one would have to use approximations, and there would be additional errors in parameters $\boldsymbol{\alpha}$ associated with the error in parametrization. The true measurement noise covariance is taken to be a multiple of identity. The *a priori* measurement noise covariance $\mathbf{R}$ used in the experiments was chosen to be equal to the true one, unless noted otherwise.

## 2.8.1 Single Posterior Estimate Experiments with The MT method

We first present the results with the MT adaptive algorithm running the Kalman filter for 50 time steps, or 4 years, and then computing the posterior estimate, $\tilde{\mathbf{Q}}$, averaging over the whole time history. Subtracting the bias correction we obtain the unbiased estimate $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ (equation (2.32)). In addition, we compute an estimate of the measurement noise covariance $\tilde{\mathbf{R}}$, and a corresponding unbiased estimate $\tilde{\mathbf{R}}_{\mathrm{mt}}$ using equations (2.33-2.34). Note that we are not trying to estimate both model and measurement error covariances as that procedure is known to be unstable; see Groutage et al. (1987).

First we present results with purely altimetric simulated measurements. $\mathbf{H}$ is a matrix with 128 by 512 elements. The results of the perfect twin experiments for each choice of the true $\hat{\mathbf{Q}}$, defined in (2.71), are presented in table 2.1. The table shows results for each experiment separately, and the experiment number is given in the leftmost column.

| No. | | 1st EOF | 2nd EOF | 3rd EOF | 4th EOF | | | MLF |
|---|---|---|---|---|---|---|---|---|
| 1 | $\hat{\mathbf{Q}}$ | 1 | 1 | 1 | 1 | $\hat{\mathbf{R}}$ | 1 | |
| | $\mathbf{Q}$ | 1 | 1 | 1 | 1 | $\mathbf{R}$ | 1 | |
| | $\tilde{\mathbf{Q}}$ | 0.22 | 0.56 | 0.14 | 0.20 | $\tilde{\mathbf{R}}$ | 2.94 | |
| | $\tilde{\mathbf{Q}}_{mt}$ | 1.01 | 1.05 | 1.00 | 1.02 | $\tilde{\mathbf{R}}_{mt}$ | 1.28 | 1.21 |
| 2 | $\hat{\mathbf{Q}}$ | 1 | 2 | 4 | 8 | $\hat{\mathbf{R}}$ | 1 | |
| | $\mathbf{Q}$ | 1 | 2 | 4 | 8 | $\mathbf{R}$ | 1 | |
| | $\tilde{\mathbf{Q}}$ | 0.20 | 0.96 | 0.52 | 3.42 | $\tilde{\mathbf{R}}$ | 6.61 | |
| | $\tilde{\mathbf{Q}}_{mt}$ | 0.89 | 2.02 | 3.83 | 8.07 | $\tilde{\mathbf{R}}_{mt}$ | 1.38 | 1.63 |
| 3 | $\hat{\mathbf{Q}}$ | 8 | 4 | 2 | 1 | $\hat{\mathbf{R}}$ | 1 | |
| | $\mathbf{Q}$ | 8 | 4 | 2 | 1 | $\mathbf{R}$ | 1 | |
| | $\tilde{\mathbf{Q}}$ | 2.64 | 2.31 | 0.19 | 1.87 | $\tilde{\mathbf{R}}$ | 6.46 | |
| | $\tilde{\mathbf{Q}}_{mt}$ | 7.80 | 3.93 | 1.87 | 0.99 | $\tilde{\mathbf{R}}_{mt}$ | 1.37 | 1.62 |

Table 2.1: Estimates of the parameter vector $\boldsymbol{\alpha}$ for single-posterior estimate MT experiments with perfect initial guess for $\mathbf{Q}$ and simulated altimetric measurements $\mathbf{H}$. The model error covariance is found by substituting the values of $\boldsymbol{\alpha}$ into equation 2.70. For each experiment we present four model error covariances: true $\hat{\mathbf{Q}}$, prior $\mathbf{Q}$, biased MT estimate $\tilde{\mathbf{Q}}$, and unbiased $\tilde{\mathbf{Q}}_{mt}$, and similarly for the measurement error.

The four rows show the true values of the error covariances, $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$, the prior guesses, $\mathbf{Q}$ and $\mathbf{R}$, the sample estimates before the bias correction, $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}$, and the unbiased estimates, $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ and $\tilde{\mathbf{R}}_{\mathrm{mt}}$ (equations (2.32) and (2.34)). In these first experiments the prior guesses are taken to be equal to the true values.

In Experiment 1 the true model error covariance is the identity; in other words the model error variance is equipartitioned, and the prior guess is correct. The measurement error covariance is assumed to be the identity as well. After 50 steps of the KF, we obtain posterior sample estimate, $\tilde{\mathbf{Q}} \leftrightarrow [0.22, 0.56, 0.14, 0.20]$. It grossly underestimates the model error variance for all vertical modes. However, after we apply the bias correction the estimate is nearly indistinguishable from the true one, $\tilde{\mathbf{Q}}_{\mathrm{mt}} \leftrightarrow [1.01, 1.05, 1.00, 1.02]$. Results for the two other choices of the true model error covariance, when the model error variance grows and diminishes with the EOF number, respectively, are very similar. That is, the unbiased estimates are indistinguishable from the truth. In Section 2.6 we have shown that bias correction is essential for satisfactory performance of the adaptive technique. The bias correction works better for the model error variance than for the measurement error variance. The largest error of the model error variance estimate is only 11 per cent, while the largest error for the measurement error variance is 38 per cent. It has to be noted again, that since there are no estimates of the uncertainty for $\tilde{\mathbf{Q}}_{\mathrm{mt}}$, we use posterior estimate of $\tilde{\mathbf{R}}$ to judge whether we have reached good estimates of the model error covariance $\tilde{\mathbf{Q}}$; see additional discussion below. To recapitulate, independent of how the model error variance is partitioned, the MT estimate is very good if we have the correct prior estimates of the model error.

In the next set of experiments, we misspecify the prior model error covariance $\mathbf{Q}$ – a more realistic case. The results are summarized in table 2.2, which is given in the same format as the Table 2.1. In experiment 4, the prior estimate is an overestimate of the true error covariance. We can observe that when the prior $\mathbf{Q}$ is different from the true $\hat{\mathbf{Q}}$, the posterior estimates are are not equal to the true ones, unlike the "perfect" experiments above. However, they are a little closer to the truth than the prior ones.

| No. |  | 1st EOF | 2nd EOF | 3rd EOF | 4th EOF |  |  | MLF |
|---|---|---|---|---|---|---|---|---|
| 4 | $\hat{\mathbf{Q}}$ | 1 | 1 | 1 | 1 | $\hat{\mathbf{R}}$ | 1 | |
|  | $\mathbf{Q}$ | 1 | 2 | 4 | 8 | $\mathbf{R}$ | 1 | |
|  | $\tilde{\mathbf{Q}}$ | 0.08 | 0.48 | 0.27 | 1.61 | $\tilde{\mathbf{R}}$ | 2.90 | |
|  | $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ | 0.79 | 1.56 | 3.58 | 6.22 | $\tilde{\mathbf{R}}_{\mathrm{mt}}$ | -2.29 | 1.34 |
| 5 | $\hat{\mathbf{Q}}$ | 1 | 1 | 1 | 1 | $\hat{\mathbf{R}}$ | 1 | |
|  | $\mathbf{Q}$ | 8 | 4 | 2 | 1 | $\mathbf{R}$ | 1 | |
|  | $\tilde{\mathbf{Q}}$ | 1.29 | 1.19 | 0.09 | 0.12 | $\tilde{\mathbf{R}}$ | 2.81 | |
|  | $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ | 6.46 | 2.80 | 1.77 | 0.92 | $\tilde{\mathbf{R}}_{\mathrm{mt}}$ | -2.28 | 1.31 |
| 6 | $\hat{\mathbf{Q}}$ | 1 | 2 | 4 | 8 | $\hat{\mathbf{R}}$ | 1 | |
|  | $\mathbf{Q}$ | 1 | 1 | 1 | 1 | $\mathbf{R}$ | 1 | |
|  | $\tilde{\mathbf{Q}}$ | 0.49 | 1.26 | 0.30 | 0.49 | $\tilde{\mathbf{R}}$ | 6.77 | |
|  | $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ | 1.26 | 1.74 | 1.16 | 1.30 | $\tilde{\mathbf{R}}_{\mathrm{mt}}$ | 5.104 | 1.97 |
| 7 | $\hat{\mathbf{Q}}$ | 1 | 2 | 4 | 8 | $\hat{\mathbf{R}}$ | 1 | |
|  | $\mathbf{Q}$ | 8 | 4 | 2 | 1 | $\mathbf{R}$ | 1 | |
|  | $\tilde{\mathbf{Q}}$ | 2.84 | 2.58 | 0.22 | 0.21 | $\tilde{\mathbf{R}}$ | 7.26 | |
|  | $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ | 8.00 | 4.19 | 1.90 | 1.01 | $\tilde{\mathbf{R}}_{\mathrm{mt}}$ | 2.18 | 1.67 |
| 8 | $\hat{\mathbf{Q}}$ | 8 | 4 | 2 | 1 | $\hat{\mathbf{R}}$ | 1 | |
|  | $\mathbf{Q}$ | 1 | 2 | 4 | 8 | $\mathbf{R}$ | 1 | |
|  | $\tilde{\mathbf{Q}}$ | 0.20 | 0.97 | 0.56 | 3.30 | $\tilde{\mathbf{R}}$ | 6.71 | |
|  | $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ | 0.91 | 2.04 | 3.87 | 7.95 | $\tilde{\mathbf{R}}_{\mathrm{mt}}$ | 1.48 | 1.64 |

Table 2.2: Estimates of the parameter vector $\boldsymbol{\alpha}$ for single-posterior estimate MT experiments with misspecified $\mathbf{Q}$ and simulated altimetric measurements $\mathbf{H}$. The model error covariance is found by substituting the values of $\boldsymbol{\alpha}$ into equation 2.70. For each experiment we present four model error covariances: true $\hat{\mathbf{Q}}$, prior $\mathbf{Q}$, biased MT estimate $\tilde{\mathbf{Q}}$, and unbiased $\tilde{\mathbf{Q}}_{\mathrm{mt}}$, and similarly for the measurement error.

This is the foundation of the MT algorithm, with the expectation that the recursive update of the covariance will eventually produce a good estimate of the true covariance $\hat{\mathbf{Q}}$. As in the "perfect" experiments above, the sample estimate $\tilde{\mathbf{Q}}$ grossly underestimates the model uncertainty, and the bias correction increases the estimates significantly.

The same is true about Experiment 5, where the prior model error variance is distributed as $[8, 4, 2, 1]$ among the vertical modes of the model. Note that estimates of the mean measurement error variance are 2.9 and $-2.3$ for Experiment 4, and 2.8 and $-2.3$, for Experiment 5, before and after the bias correction correspondingly. The MT estimate of the measurement error is clearly wrong as a variance can never be negative. Ideally, the bias correction would always correct for the error in the prior estimate. Due to the complexity of the Kalman filter algorithm this is not practical. At the same time, the apparent negative variance indicates that our estimate for the prior statistics of the model error was wrong, and we should try something else. The MT algorithm prescribes substituting the posterior estimate $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ for the prior, and keep on going. In fact, the posterior estimate is closer to the true one than the prior, and we would indeed expect to obtain a good estimate after a few iterations. Similar conclusions are reached when we start with an underestimate of the model error variance, e.g. Experiment 6.

Experiment 7, and especially Experiment 8, show that there is a subtlety. In Experiment 7 the true $\hat{\mathbf{Q}}$ has variance partitioned as $[1, 2, 4, 8]$ between the four vertical modes. The prior estimate is partitioned in the opposite way, $[8, 4, 2, 1]$. The posterior MT estimates are within 5 per cent of the prior, that is, essentially the same. Thus, we expect that recursive application of the algorithm is not going change the prior estimate of the model error covariance, and this is indeed the case as is shown in Section 2.8.2. The mean variance of the posterior measurement error variance, $\tilde{\mathbf{R}}_{\mathrm{mt}}$, is close to the prior one, which was taken to be equal to the true. Thus, if we were to hide the information about the true values of the model and measurement error covariances, experiments 7 and 8 should be grouped together with the perfect experiments 1-3 (experiments 2 and 7 share the same true statistics, and so do experiments 3 and 8). But in experiments 7 and

8 we have a totally wrong partitioning of the model error variance. Thus, the adaptive algorithm of MT does not have a unique estimate of the model error with the LM and altimetric measurements. Therefore, we cannot be certain that our estimate of the model error we were to obtain using real altimetric data is the correct one.

The results change dramatically once we assimilate additional observations. First, we present the results with the complete observational network ($\mathbf{H}$ given by the identity matrix). Of course, at present, we have no observational technique which can provide that kind of observations on global scale. But it provides a convenient benchmark for comparison of different observational networks. The results are presented in Table 2.3. These experiments are identical to the ones discussed above, except for the change of $\mathbf{H}$. Only results for experiments with the same choice of true and prior model error covariances as in the experiments 2, 6 and 7 above are shown. Experiment 2a shows that with this measurement matrix the perfect experiments are still the same, that is, the prior and the posterior are nearly the same. But the results for Experiment 7a are now different. The posterior estimate for $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ is no longer the same as the prior $\mathbf{Q}$, and is much closer to the true one. The Kalman filter has efficiently assimilated information available in the observations and eliminated the wrong partitioning of the variance among the modes. Similar conclusions are valid when we reverse the partitioning; see experiment 6. We can see that assimilation with an observation for each element of the state makes a big difference. We should be able to estimate the model error covariance adaptively with this set of parameters.

Next, we present results for experiments where simulated idealized tomographic sections are used in addition to altimetric observations. To test whether tomographic measurements can even in principle be as useful as full observational network, i.e. identity observation matrix, we assume that tomographic sections are given at each latitudinal and longitudinal grid point. This gives us 24 tomographic rays (16 longitudinal grid points and 8 latitudinal grid points). Thus, the number of observations at every time step increases to 224. The results are presented in table 2.4. They indicate that augmenting

| No. | | 1st EOF | 2nd EOF | 3rd EOF | 4th EOF | | | MLF |
|-----|-----|---------|---------|---------|---------|-----|-----|-----|
| 2a | $\hat{Q}$ | 1 | 2 | 4 | 8 | $\hat{R}$ | 1 | |
| | $Q$ | 1 | 2 | 4 | 8 | $R$ | 1 | |
| | $\tilde{Q}$ | 0.83 | 1.98 | 3.85 | 7.53 | $\tilde{R}$ | 5.84 | |
| | $\tilde{Q}_{mt}$ | 0.86 | 1.91 | 3.91 | 7.85 | $\tilde{R}_{mt}$ | 1.42 | 6.80 |
| 6a | $\hat{Q}$ | 1 | 2 | 4 | 8 | $\hat{R}$ | 1 | |
| | $Q$ | 1 | 1 | 1 | 1 | $R$ | 1 | |
| | $\tilde{Q}$ | 0.83 | 1.39 | 1.99 | 3.11 | $\tilde{R}$ | 5.89 | |
| | $\tilde{Q}_{mt}$ | 0.87 | 1.32 | 1.99 | 3.27 | $\tilde{R}_{mt}$ | 4.34 | 8.41 |
| 7a | $\hat{Q}$ | 1 | 2 | 4 | 8 | $\hat{R}$ | 1 | |
| | $Q$ | 8 | 4 | 2 | 1 | $R$ | 1 | |
| | $\tilde{Q}$ | 2.13 | 2.58 | 2.88 | 3.20 | $\tilde{R}$ | 6.00 | |
| | $\tilde{Q}_{mt}$ | 2.28 | 2.55 | 2.91 | 3.34 | $\tilde{R}_{mt}$ | 1.57 | 8.48 |

Table 2.3: Estimates of the parameter vector $\alpha$ for single-posterior estimate MT experiments with with misspecified $Q$ and identity $H$. The model error covariance is found by substituting the values of $\alpha$ into equation 2.70. For each experiment we present four model error covariances: true $\hat{Q}$, prior $Q$, biased MT estimate $\tilde{Q}$, and unbiased $\tilde{Q}_{mt}$, and similarly for the measurement error.

additional tomographic lines makes the results similar to those in which $H$ is given by the identity. However, the MT estimates are not nearly as good as with the identity observation matrix, e.g. $\tilde{Q}_{mt} \leftrightarrow [6.83, 3.81, 2.50, 2.35]$ in 7b versus $\tilde{Q}_{mt} \leftrightarrow [2.28, 2.55, 2.91, 3.34]$ in 7a, with the truth given by $\hat{Q} \leftrightarrow [1, 2, 4, 8]$. Thus, based on these twin experiments we can conclude that a combination of altimetric and latitudinal and longitudinal tomographic measurements has the same information content for estimation of the error properties as measurements of each state element, but it would take a longer time series to reach the true estimates. This is a non-trivial conclusion given that we only have 224 observations at every time step versus 512 state elements, and altimetric observations alone are not capable of differentiating the vertical partitioning of the model error variance.

In addition, we ran experiments with misspecified measurement error covariance, i.e. $R \neq \hat{R}$. The results were not qualitatively different, and are summarized in table 2.5. As

| No. | | 1st EOF | 2nd EOF | 3rd EOF | 4th EOF | | | MLF |
|-----|-----|---------|---------|---------|---------|-----|-----|-----|
| 2b | $\hat{\mathbf{Q}}$ | 1 | 2 | 4 | 8 | $\hat{\mathbf{R}}$ | 1 | |
| | $\mathbf{Q}$ | 1 | 2 | 4 | 8 | $\mathbf{R}$ | 1 | |
| | $\tilde{\mathbf{Q}}$ | 0.46 | 1.18 | 1.27 | 4.79 | $\tilde{\mathbf{R}}$ | 25.32 | |
| | $\tilde{\mathbf{Q}}_{mt}$ | 0.97 | 1.92 | 3.90 | 7.98 | $\tilde{\mathbf{R}}_{mt}$ | 2.53 | 3.70 |
| 6b | $\hat{\mathbf{Q}}$ | 1 | 2 | 4 | 8 | $\hat{\mathbf{R}}$ | 1 | |
| | $\mathbf{Q}$ | 1 | 1 | 1 | 1 | $\mathbf{R}$ | 1 | |
| | $\tilde{\mathbf{Q}}$ | 0.65 | 1.48 | 1.01 | 1.97 | $\tilde{\mathbf{R}}$ | 26.15 | |
| | $\tilde{\mathbf{Q}}_{mt}$ | 1.20 | 1.73 | 1.65 | 2.52 | $\tilde{\mathbf{R}}_{mt}$ | 19.11 | 4.60 |
| 7b | $\hat{\mathbf{Q}}$ | 1 | 2 | 4 | 8 | $\hat{\mathbf{R}}$ | 1 | |
| | $\mathbf{Q}$ | 8 | 4 | 2 | 1 | $\mathbf{R}$ | 1 | |
| | $\tilde{\mathbf{Q}}$ | 2.84 | 2.71 | 1.12 | 1.87 | $\tilde{\mathbf{R}}$ | 27.38 | |
| | $\tilde{\mathbf{Q}}_{mt}$ | 6.83 | 3.81 | 2.50 | 2.35 | $\tilde{\mathbf{R}}_{mt}$ | 4.76 | 4.32 |

Table 2.4: Estimates of the parameter vector $\boldsymbol{\alpha}$ for single-posterior estimate MT experiments with misspecified $\mathbf{Q}$ and simulated altimetric and tomographic (latitudinal and longitudinal sections) measurements $\mathbf{H}$, $N_o$=224. The model error covariance is found by substituting the values of $\boldsymbol{\alpha}$ into equation 2.70. For each experiment we present four model error covariances: true $\hat{\mathbf{Q}}$, prior $\mathbf{Q}$, biased MT estimate $\tilde{\mathbf{Q}}$, and unbiased $\tilde{\mathbf{Q}}_{mt}$, and similarly for the measurement error.

expected, if we underestimate the measurement error variance we tend to overestimate the model error variance (e.g. experiments 1u, 1o and 1, u and o stand for underestimate and overestimate respectively). However, the estimates are not very sensitive and the change of prior measurement error variance by a factor of 4 produces only a change of less than 25 per cent in estimates of $\tilde{\mathbf{Q}}_{mt}$ (experiments 4u and 4o, 6u and 6o).

## 2.8.2 Fully Adaptive Twin Experiments

Having discussed preliminary results where we computed only the posterior estimate, we now present results using fully adaptive estimation. First we have to decide on the size of the averaging window used in the MT adaptive scheme (equation (2.32)). As discussed in BFC97, the results are sensitive to the choice of $N$, the shorter the window the better. Note that since real data is only available over 50 time steps of the LM (i.e. 4 years), and we also have to allow for an initialization period, we have a limited choice for the length of the averaging window. In fact, the experiments with posterior estimate experiments discussed in Section 2.8 can be viewed as adaptive experiments with the window of length 50 and with observations available over 50 time steps. We decided to use a window of size $S = 10$ (equation (2.32)), which is close to that used in BFC97. The difference from the earlier experiments is that we estimate the model error covariance, $\tilde{\mathbf{Q}}_{mt}$, at every time step starting with the 11th time step and then use that estimate in the Kalman filter algorithm, equations (2.26-2.30) to obtain an estimate at the next step, and so on. We estimate only four coefficients, namely the mean variance of the model error for each of four vertical modes. The measurement error covariance $\mathbf{R}$ is not updated at all. In all other respects, these twin experiments are identical to the earlier ones.

The time series of the MT estimates of the mean model error variance for each vertical mode are shown in Figure 2.18. Four different experiments corresponding to different choices of the measurement matrix $\mathbf{H}$ are shown. The results with $\mathbf{H} = \mathbf{I}$, (Figure 2.18, yellow) show that when every state element is directly measured, the MT algorithm is very efficient at changing the model error variance partitioning, and converges to the

| No. | | 1st EOF | 2nd EOF | 3rd EOF | 4th EOF | | |
|-----|------|---------|---------|---------|---------|------|------|
| 1o | $\hat{Q}$ | 1 | 1 | 1 | 1 | $\hat{R}$ | 1 |
| | $Q$ | 1 | 1 | 1 | 1 | $R$ | 2 |
| | $\tilde{Q}$ | 0.12 | 0.38 | 0.09 | 0.12 | $\tilde{R}$ | 2.81 |
| | $\tilde{Q}_{mt}$ | 0.93 | 0.87 | 0.95 | 0.96 | $\tilde{R}_{mt}$ | 0.90 |
| 1u | $\hat{Q}$ | 1 | 1 | 1 | 1 | $\hat{R}$ | 1 |
| | $Q$ | 1 | 1 | 1 | 1 | $R$ | 0.5 |
| | $\tilde{Q}$ | 0.31 | 0.78 | 0.20 | 0.30 | $\tilde{R}$ | 2.90 |
| | $\tilde{Q}_{mt}$ | 1.08 | 1.25 | 1.05 | 1.09 | $\tilde{R}_{mt}$ | 1.39 |
| 4o | $\hat{Q}$ | 1 | 1 | 1 | 1 | $\hat{R}$ | 1 |
| | $Q$ | 1 | 2 | 4 | 8 | $R$ | 2 |
| | $\tilde{Q}$ | 0.07 | 0.38 | 0.20 | 1.10 | $\tilde{R}$ | 2.88 |
| | $\tilde{Q}_{mt}$ | 0.77 | 1.42 | 3.51 | 6.07 | $\tilde{R}_{mt}$ | -2.75 |
| 4u | $\hat{Q}$ | 1 | 1 | 1 | 1 | $\hat{R}$ | 1 |
| | $Q$ | 1 | 2 | 4 | 8 | $R$ | 0.5 |
| | $\tilde{Q}$ | 0.12 | 0.71 | 0.36 | 2.21 | $\tilde{R}$ | 3.04 |
| | $\tilde{Q}_{mt}$ | 0.82 | 1.80 | 3.06 | 6.60 | $\tilde{R}_{mt}$ | -2.75 |
| 6o | $\hat{Q}$ | 1 | 2 | 4 | 8 | $\hat{R}$ | 1 |
| | $Q$ | 1 | 1 | 1 | 1 | $R$ | 2 |
| | $\tilde{Q}$ | 0.33 | 0.90 | 0.21 | 0.31 | $\tilde{R}$ | 7.02 |
| | $\tilde{Q}_{mt}$ | 1.13 | 1.40 | 1.07 | 1.16 | $\tilde{R}_{mt}$ | 5.11 |
| 6u | $\hat{Q}$ | 1 | 2 | 4 | 8 | $\hat{R}$ | 1 |
| | $Q$ | 1 | 1 | 1 | 1 | $R$ | 0.5 |
| | $\tilde{Q}$ | 0.67 | 1.67 | 0.40 | 0.70 | $\tilde{R}$ | 6.68 |
| | $\tilde{Q}_{mt}$ | 1.44 | 2.15 | 1.26 | 1.50 | $\tilde{R}_{mt}$ | 5.18 |

Table 2.5: Estimates of the parameter vector $\alpha$ for single-posterior estimate MT experiments with perfect initial guess for $Q$, misspecified $R$, and simulated altimetric measurements $H$. The model error covariance is found by substituting the values of $\alpha$ into equation 2.70. For each experiment we present four model error covariances: true $\hat{Q}$, prior $Q$, biased MT estimate $\tilde{Q}$, and unbiased $\tilde{Q}_{mt}$, and similarly for the measurement error.
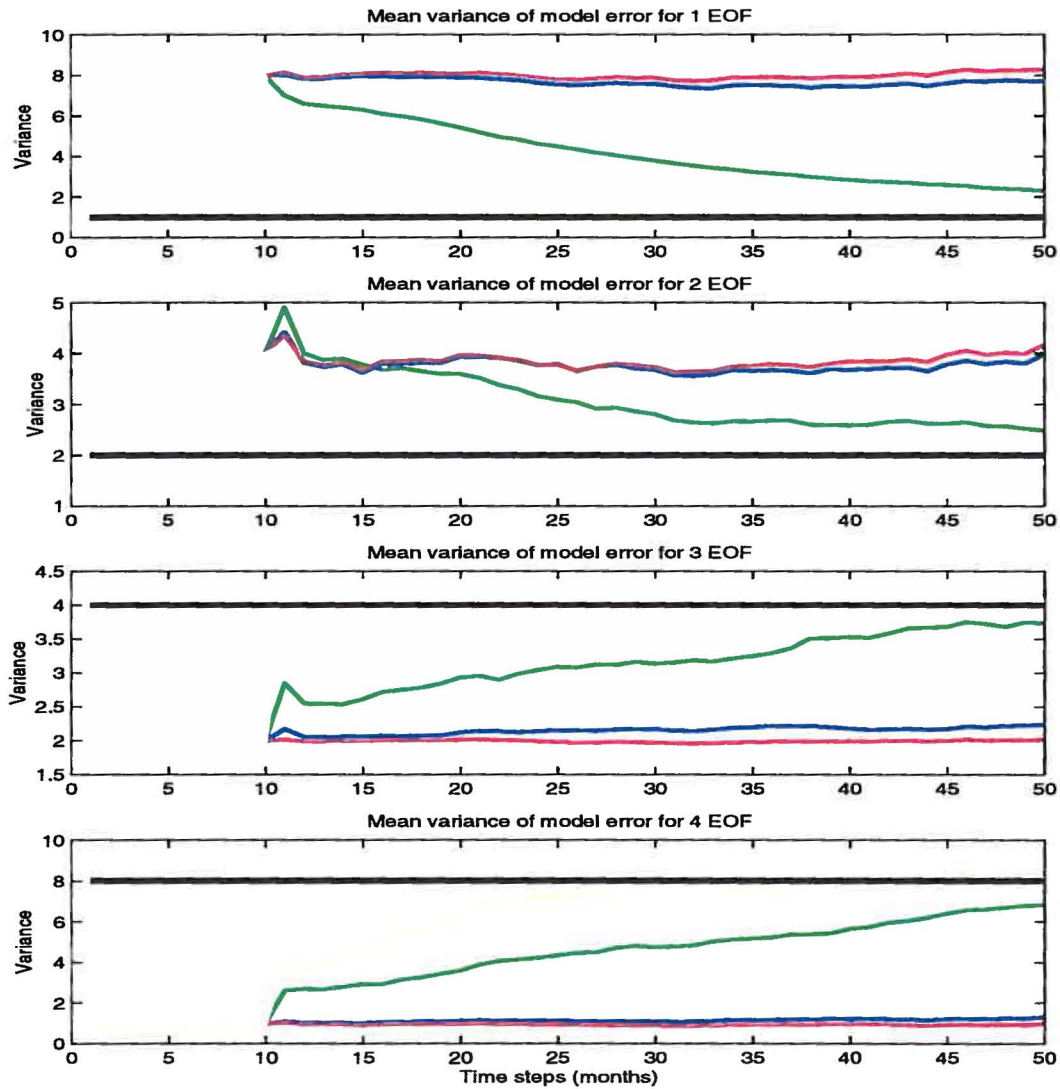
Figure 2.18: Time series for MT estimates $\bar{\mathbf{Q}}_{mt}$ with averaging window of 10 time steps for simulated altimetric measurements (red), altimetric and 4 ATOC tomographic rays (blue), altimetric and 24 idealized tomographic rays (green), and identity observation matrix $\mathbf{H}$ (yellow). The thick black line represents the true values.

specified true value of $[1, 2, 4, 8]$. The estimates fluctuate a little around the true value. The final estimate is within 10 per cent of the true one, even though the prior was under-specified by as much as a factor of 8.

However, when we assimilate simulated altimetric observations, the adaptive results are poor (Figure 2.18, red). We see that the recursive update does not change the prior at all, in spite of the fact that the prior $\mathbf{Q}$ is totally wrong. This was predicted by the results of a single posterior estimate (Experiment 8). Based on these twin experiments we conclude that using the altimetric measurements alone we cannot infer the true model error covariance using the MT method. Since the number of parameters is very small, this implies that the MT method cannot be used for estimation of the model error covariance for this particular linear model.

We next include additional observations from 24 synthetic latitudinal and longitudinal tomographic lines (Section 2.8.1), increasing number of observations to 224 at every time step. Still, we observe less than half the number of degrees of freedom in the model. The performance of the algorithm is now dramatically different (Figure 2.18, blue). This shows that tomographic data can be successfully used to differentiate between model errors variances of different internal modes. In addition, we ran a number of other twin experiments with other choices of true and prior covariances, $\hat{\mathbf{Q}}$ and $\mathbf{Q}$, with similar results.

Last, we ran a twin experiment with four simulated tomographic arrays, those currently available from the ATOC acoustic array (Figure 2.18, green). These four rays are shown by thick solid lines L, K, N, O in Figure 2.4. They are very similar to the ones obtained with altimetric measurements alone. Thus, while in principle the tomographic measurements are capable of saving the situation, the ones currently available do not. Even when all the assumptions required by the method are perfectly satisfied the MT method fails with the kind of observations available in oceanography.

## 2.9 Maximum Likelihood Estimator

The other algorithm which can be applied to oceanographic problems is a maximum likelihood (ML, hereafter) algorithm of Dee (1995). It has been reformulated in BFC97 to allow use of a sequence of observations. The ML algorithm was also shown to be identical to Maybeck's algorithm when the innovations are assumed to be normally distributed, and thus to that of MT. It is directly related to the covariance-matching with innovations algorithms, (e.g. Shellenbarger, 1967; Belanger, 1974).

We start as in the MT algorithm, section 2.5.1, by defining the innovation vector:

$$\mathbf{v}(t) = \mathbf{y}(t) - \mathbf{H}\tilde{\mathbf{p}}(t|t-1), \tag{2.72}$$

which represents additional information provided by the observations. Next, we define a lag covariance of the innovations

$$\mathbf{C}_j(t) = <\mathbf{v}(t)\mathbf{v}^{\mathrm{T}}(t-j)> . \tag{2.73}$$

It can be shown that $\mathbf{C}_j(t) = 0$ for $j \neq 0$, i.e. the innovation sequence is white, if and only if the Kalman filter is optimal, Jazwinski (1969). The lag zero covariance of the innovation sequence is given by

$$\mathbf{C}_0(t) = \mathbf{H}\mathbf{\Pi}(t|t-1)\mathbf{H}^{\mathrm{T}} + \mathbf{R} \tag{2.74}$$

The uncertainty of the Kalman filter forecast, $\mathbf{\Pi}(t|t-1)$, is dependent on the model and measurement error covariances through the KF equations ( 2.27-2.30).

The goal of the adaptive filter is then to compute the error covariances, given by the parameter vector $\boldsymbol{\alpha}$ through equations 2.35, which lead to a "white" innovation sequence with the sample covariance of the innovations matching the expected one. Thus, we define a likelihood function

$$f(\boldsymbol{\alpha}) = \sum_{t=1}^{\text{K steps}} \left( ln[\det \tilde{\mathbf{C}}_0(t, \boldsymbol{\alpha})] + \mathbf{v}(t)^{\mathrm{T}}\tilde{\mathbf{C}}_0^{-1}(t, \boldsymbol{\alpha})\mathbf{v}(t)] \right), \tag{2.75}$$

where $\tilde{\mathbf{C}}_0(i, \boldsymbol{\alpha})$ is related to $\mathbf{Q}$ and $\mathbf{R}$ through equation (2.74). One then proceeds to maximize the likelihood function with respect to the parameter vector $\boldsymbol{\alpha}$. Once a

parameter vector is chosen, the KF can be ran forward again with a new prior guess of $\mathbf{Q}$, similar to the MT algorithm.

## 2.9.1 Twin Experiments With a Maximum Likelihood Estimator

The maximum likelihood estimator of Dee (1995) is described in Section 2.9. The results of BFC97 suggest that this approach yields results very similar to use of the MT algorithm but is much more computationally demanding. In our case the dimension of the state is more than four times that used in BFC97, and therefore the computational cost is even greater. Therefore, instead of doing optimization we compute the values of the likelihood function for the MT method experiments (Section 2.8.1), and show that the ML method gives results similar to the MT method.

While we did not run full tests of the ML algorithm, we computed the value of the likelihood function for each of experiments described in Section 2.8.1. The values (normalized by $10^4$) are shown in the most right column in the tables 2.1–2.5. If we compare values of the likelihood function for experiments with the same true model and measurement error covariances, and therefore the same statistics of observations, we can see which of the prior guesses for $\mathbf{Q}$, is the most likely. For example, experiments 1, 4, and 5 were identical, but used different initial guesses for $\mathbf{Q}$, and the values of the likelihood function are 1.21, 1.31, and 1.34. Thus, for this case with vertical equipartitioning of the model error the likelihood function is considerably smaller when the prior is equal to the true (Experiment 1). However, for the experiments 2, 6, and 7, which all share the same true model error covariance given by $\boldsymbol{\alpha}$ of $[1, 2, 4, 8]$, the likelihood function is 1.63, 1.97, and 1.67, respectively. That is, just as for the MT method the two prior choices, the true one and the one with opposite partitioning of the model error variance, are nearly indistinguishable. We would need a very efficient optimization routine to be able to find the global minimum (assuming it is indeed at the true value of the model error covariance). Considering the fact that we need to run a Kalman filter for many

94

time steps for each evaluation of the likelihood function, this procedure is computationally very expensive. Therefore, the ML method cannot provide a stable estimate of the model error with just altimetric measurements. The ML algorithm would fail to converge to the true estimate in this case, similar to the MT method.

The results are once again quite different if we have a perfect observational network, Table 2.3, and synthetic altimetric and tomographic measurements, Table 2.4. The values of the likelihood function are much lower with the correct choice than with the misspecified model error covariance, and the ML algorithm quickly converges to the true estimates of the model error statistics.

## 2.10  Summary

Our immediate goal was to obtain a practical method for the quantitative estimation of large scale baroclinic GCM errors in the North Pacific using TOPEX/POSEIDON altimeter and ATOC tomographic data. This goal remains elusive. Following the suggestion of Blanchet et al. (1997), we singled out the Myers and Tapley (1976) adaptive algorithm for the investigation. Simultaneous estimation of the model and measurement error statistics with such adaptive methods is unstable, and we concentrated on the problem adaptive estimation of the model error statistics.

In a series of twin experiments, we applied the MT algorithm to a reduced state linear model which approximates the dynamics of large scale GCM errors (large scale is here defined as 4 vertical degrees of freedom and 8° horizontal sampling). The twin experiments were carried out under the simplifying assumption that GCM errors are consistent with a diagonal, horizontally homogeneous, covariance matrix for the linear model system error. Instead of estimating the full covariance matrix as proposed in the original MT algorithm we estimated only several parameters (the mean diagonal values), as described in Dee (1991). Our principal conclusion is that even under this simplifying assumption, the simulated altimeter data fail to provide sufficient information for quantifying GCM

errors. In particular, we find that the MT algorithm converges to different solutions depending on the initial guess. The addition of a simulated tomographic array consisting of 8 zonal and 16 meridional basin-wide acoustic paths forces the MT algorithm to converge to a unique solution. However our numerical experiments also indicate that, due to its limited spatial and temporal coverage, the available ATOC data is insufficient to uniquely constrain the present adaptive estimation problem.

These negative conclusions about the MT algorithm are supported by the analysis of low-dimensional systems presented in Sections 2.6 and 2.7. The results with low-dimensional models, tested over a very large parameter space, showed that the MT algorithm is sensitive to the initial guess of the model error covariance and misspecification of the dynamic model matrix and the measurement error statistics. Even with infinitely long time series of observations, it does not always converge to the true estimates. The limit on the number of parameters given by Mehra (1970), can only serve as an upper limit. There is no information on the uncertainty of the estimates, and no information on what choices of parameters can be estimated. The algorithm is very inefficient; it takes many iterations to produce correct estimates in the idealized setup with infinite time series of observations. Thus, while it may work in some cases (for example, the problem considered in Blanchet et al. (1997), where it is assumed that the prior is wrong by a factor of 50), it is not guaranteed to succeed in other cases. Similar conclusions were reached with a ML algorithm of Dee (1995). This leads us to development of a different, but closely related approach described in the next chapter.

It has to be noted that the issue of systematic error, i.e non-zero mean of the system and measurement noise, has been completely neglected here. The reason is that it has been shown that estimating first-order statistics is more difficult than second-order statistics (see Anderson and Moore, 1979). We return to this issue in the next chapter where we consider a new covariance matching approach.

# Chapter 3

# Covariance Matching Approach (CMA)

In this chapter we present a new method, a covariance matching approach (CMA), for estimation of error statistics. It is closely related to the adaptive methods presented in Chapter 2 and the method of Fu et al. (1993) (reviewed in Appendix D), but is different in several important ways. The name comes from the main idea of the method: matching expectations for the data covariances with the sample estimates. Unlike the covariance matching methods developed in the control literature (Shellenbarger 1967, Belanger 1974) we use observations directly and not through the innovation sequence. The covariance matching with innovations provided basis for the adaptive methods discussed in Chapter 2. Other differences are that we use a Green's function approach, utilize several lag-difference covariances at once, and provide a reliable way to estimate uncertainties of the estimates.

We start by presenting the method and discussing various theoretical and practical aspects. We then present a small numerical example in Section 3.4, and contrast it with similar tests with the MT method (Section 2.7). Differences and similarities with other adaptive methods are discussed in Section 3.6. We also extend the method to estimate not only the covariances of the model and measurement errors but also the trends and

annual cycles (Section 3.5).

To test the CMA in a more realistic setup we ran a series of twin experiments using the same setup as that used to test the MT method (Section 4.3). Experimental results with real data are presented in Chapter 4.

Major part of this and the next chapter is presented in the article Menemenlis and Chechelnitsky (1999). The discussion below follows the one in the article, but a number of important theoretical and practical questions are addressed in much greater detail.

## 3.1   Statistical Modeling

Here we summarize the notation and setup which are described in detail in Section 2.3. Let $\mathbf{p}(t)$ represent GCM simulation errors, which we model dynamically as

$$\mathbf{p}(t+1) \;\; = \;\; \mathbf{A}(t)\mathbf{p}(t) + \mathbf{\Gamma}(t)\mathbf{u}(t). \tag{3.1}$$

where $\mathbf{A}(t)$ is the state transition matrix and $\mathbf{\Gamma}(t)\,\mathbf{u}(t)$, the model, or system, error (errors in the boundary conditions, indeterminate GCM parameters, etc.). The residual of oceanographic observations, $\boldsymbol{\eta}_{\text{ocean}}(t)$ and GCM predictions, $\boldsymbol{\zeta}_{\text{GCM}}(t)$, can be expressed as a noisy linear (or linearized) combination of $\mathbf{p}(t)$,

$$\mathbf{y}(t) = \mathbf{H}(t)\mathbf{p}(t) + \mathbf{r}(t), \tag{3.2}$$

where $\mathbf{H}(t)$ is the measurement matrix and $\mathbf{r}(t)$ is measurement error. In addition to instrument errors, $\mathbf{r}(t)$ includes oceanic signal that is not resolved by the GCM, the so-called representation error. Vectors $\mathbf{u}(t)$ and $\mathbf{r}(t)$ are taken to be random variables and are described by their means, $\langle \mathbf{u}(t) \rangle$ and $\langle \mathbf{r}(t) \rangle$, and by their covariance matrices, $\mathbf{Q}(t) \equiv \text{cov}\,\mathbf{u}(t)$ and $\mathbf{R}(t) \equiv \text{cov}\,\mathbf{r}(t)$, where the covariance operator is defined in the usual way, $\text{cov}\,\mathbf{u} \equiv \langle [\mathbf{u} - \langle \mathbf{u} \rangle][\mathbf{u} - \langle \mathbf{u} \rangle]^{\text{T}} \rangle$, $\langle \cdot \rangle$ is the expectation operator, and $^{\text{T}}$ is the transpose. This is a complete statistical description of the errors if the random vectors $\mathbf{u}(t)$ and $\mathbf{r}(t)$ have multivariate normal distribution (Mardia et al. 1979), that is, if the errors can

be modeled as resulting from a set of stationary Gaussian processes. If the errors are non-Gaussian, the mean and covariance remain useful, though incomplete, descriptors. Our objective is to use measurements $\mathbf{y}(t)$ to estimate $\langle \mathbf{u}(t) \rangle$, $\langle \mathbf{r}(t) \rangle$, $\mathbf{Q}(t)$, and $\mathbf{R}(t)$ (See Table A.1 for a summary of the notation, Section 2.3 for a discussion of the errors).

## 3.2   The Basic Algorithm

We start by considering the case where $\mathbf{A}$, $\boldsymbol{\Gamma}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$ are steady (time-independent); $\mathbf{A}$, $\boldsymbol{\Gamma}$, and $\mathbf{H}$ are known; $\mathbf{A}$ is stable, that is, all its eigenvalues are less than one by absolute value; $\boldsymbol{\Gamma}$ is the identity; and vectors $\mathbf{u}(t)$ and $\mathbf{r}(t)$ have zero mean and are independent of $\mathbf{p}(t)$,

$$\langle \mathbf{u}(t) \rangle = \mathbf{0}, \quad \langle \mathbf{r}(t) \rangle = \mathbf{0}, \quad \langle \mathbf{p}(t)\, \mathbf{u}(t)^{\mathrm{T}} \rangle = \mathbf{0}, \quad \langle \mathbf{p}(t)\, \mathbf{r}(t)^{\mathrm{T}} \rangle = \mathbf{0}. \tag{3.3}$$

(The assumption of independence between $\mathbf{u}(t)$ and $\mathbf{p}(t)$ is less restrictive than that used by Fu et al. (1993) who assumed the model simulation error to be independent of the true state, $\langle \mathbf{p}(t)\, \hat{\boldsymbol{\zeta}}_{\mathrm{GCM}}(t)^{\mathrm{T}} \rangle = \mathbf{0}$, Appendix D). For stable $\mathbf{A}$, equations (3.1), (3.2), and (3.3), imply that $\langle \mathbf{y}(t) \rangle = \langle \mathbf{p}(t) \rangle = \mathbf{0}$; equations (3.1) and (3.3) imply that $\langle \mathbf{u}(t_1)\, \mathbf{u}(t_2)^{\mathrm{T}} \rangle = \mathbf{0}$ for $t_1 \neq t_2$. Finally we parameterize $\mathbf{Q}$ and $\mathbf{R}$ as

$$\mathbf{Q} = \sum_{k=1}^{K} \alpha_k \mathbf{Q}_k, \quad \mathbf{R} = \sum_{k=1}^{L} \alpha_{K+k} \mathbf{R}_k. \tag{3.4}$$

Multiplying equation (3.1) by its transpose and taking expectations produces the steady state Lyapunov equation (Anderson and Moore, 1979, p. 62),

$$\mathbf{P} \equiv \mathrm{cov}\, \mathbf{p} = \mathbf{A}\mathbf{P}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}, \tag{3.5}$$

which relates the covariance of the GCM error to that of the system error. The Lyapunov equation can readily be solved for $\mathbf{P}$ using a numerical scheme in Section 3.3.2. Similarly, multiplying equation (3.2) by its transpose and taking expectations yields

$$\mathbf{Y} \equiv \mathrm{cov}\, \mathbf{y} = \mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R}. \tag{3.6}$$

From equations (3.5) and (3.6) it follows that each element of $\mathbf{Y}$ is linearly related to the elements of $\mathbf{Q}$ and $\mathbf{R}$, and hence to parameters $\alpha_k$ in equation (3.4). An elegant way to solve this system of equations is through the use of Green's functions, $\mathbf{G}_{\mathbf{Y},k}$, here defined as the response of the measurement covariance matrix, $\mathbf{Y}$, to unit perturbations of $\mathbf{Q}_k$ or $\mathbf{R}_k$, that is,

$$\mathbf{G}_{\mathbf{Y},k} = \mathbf{H}\mathbf{P}_k\mathbf{H}^{\mathrm{T}}, \qquad \mathbf{G}_{\mathbf{Y},K+k} = \mathbf{R}_k, \tag{3.7}$$

where $\mathbf{P}_k$ can be calculated from $\mathbf{Q}_k$ by the Lyapunov equation (3.5). Rewriting $\mathbf{Y}$ and $\mathbf{G}_{\mathbf{Y},k}$ as column vectors, $(\vdots)$, yields a set of linear equations,

$$\mathbf{Y}(\vdots) = [\mathbf{G}_{\mathbf{Y},1}(\vdots) \dots \mathbf{G}_{\mathbf{Y},K+L}(\vdots)] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{K+L} \end{bmatrix}, \tag{3.8}$$

which can be solved for parameters $\alpha_k$ using any of several discrete linear inverse methods, e.g. Menke (1989), Wunsch (1996). To reduce computational cost, the column operator $(\vdots)$ (Section 3.3.3) in equation (3.8) can be replaced by some representative subset of matrix $\mathbf{Y}$, for example its diagonal elements, arranged in column vector format. For any given definition of the operator $(\vdots)$ and set of matrices $\mathbf{A}$ and $\mathbf{H}$, linear inverse theory provides powerful tools for understanding which, and how well, combinations of parameters $\alpha_k$ in equation (3.8) can be determined.

This completes a basic description of the estimation algorithm. We next consider a series of algorithmic refinements and the effects of relaxing some of the simplifying assumptions. One issue is whether $\mathbf{R}$ and $\mathbf{Q}$ can be estimated simultaneously, that is, whether an arbitrary set of parameters $\alpha_k$ in equation (3.8) can be resolved independently (e.g. Groutage et al., 1987 ; Maybeck, 1982 ). In Sections 3.2.1 and 3.2.2 we demonstrate that, under a very general set of conditions, $\mathbf{R}$ and $\mathbf{Q}$ can be resolved by making use of time-lag correlations in the data. A more serious limitation is that $\mathbf{Y}$ is estimated as the sample covariance of $\mathbf{y}(t)$: the consequences of sampling uncertainty are discussed in Section 3.3. The algorithm is illustrated with a small numerical example

in Section 3.4. Systematic and time-correlated errors are considered in Sections 3.5.1 and 3.5.2. Section 3.5.3 deals with time-dependent models. Section 3.5.4 discusses statistical consistency tests. The comparison with innovations based approaches is given in Section 3.6.

### 3.2.1 Using Lag-Difference Covariance Matrices

The covariance matrix $\mathbf{Y}$ does not describe temporal correlations in the data. It is therefore reasonable to expect that estimates of $\mathbf{Q}$ and $\mathbf{R}$ might be improved by making use of lag-difference covariance matrices. From a recursive application of equation (3.1) and from the definitions of $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{P}$, the covariance matrix of the lag-$s$ difference is

$$
\begin{aligned}
\mathbf{D}_s &\equiv \operatorname{cov}[\mathbf{y}(t+s) - \mathbf{y}(t)] \\
&= 2\mathbf{H}\mathbf{P}\mathbf{H}^\mathrm{T} - \mathbf{H}\mathbf{A}^s\mathbf{P}\mathbf{H}^\mathrm{T} - \mathbf{H}\mathbf{P}\mathbf{A}^{s\prime}\mathbf{H}^\mathrm{T} + 2\mathbf{R} \\
&= \mathbf{H}(\mathbf{A}^s - \mathbf{I})\mathbf{P}(\mathbf{A}^s - \mathbf{I})^\mathrm{T}\mathbf{H}^\mathrm{T} + \sum_{k=1}^{s} \mathbf{H}\mathbf{A}^{s-k}\mathbf{Q}\mathbf{A}^{(s-k)\prime}\mathbf{H}^\mathrm{T} + 2\mathbf{R},
\end{aligned} \tag{3.9}
$$

where it is assumed that measurement errors are uncorrelated in time,

$$
\langle \mathbf{r}(t_1)\,\mathbf{r}(t_2)^\mathrm{T} \rangle = \mathbf{0} \quad \text{for} \quad t_1 \neq t_2.
$$

$\mathbf{Y}$ and several lag-$s$-difference covariance matrices can be combined in an equation of type $\mathbf{d} = \mathcal{G}\alpha$, that is,

$$
\begin{bmatrix} \mathbf{Y}(:) \\ \mathbf{D}_1(:) \\ \vdots \\ \mathbf{D}_S(:) \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{\mathbf{Y},1}(:) & \cdots & \mathbf{G}_{\mathbf{Y},K+L}(:) \\ \mathbf{G}_{\mathbf{D}_1,1}(:) & \cdots & \mathbf{G}_{\mathbf{D}_1,K+L}(:) \\ & \vdots & \\ \mathbf{G}_{\mathbf{D}_S,1}(:) & \cdots & \mathbf{G}_{\mathbf{D}_S,K+L}(:) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{K+L} \end{bmatrix}. \tag{3.10}
$$

As before $\mathbf{G}_{\mathbf{D}_s,k}$ represents the Green's function associated with the data covariance matrix $\mathbf{D}_s$ and parameter $\alpha_k$. Because $\mathbf{D}_r - \mathbf{D}_s$ is independent of $\mathbf{R}$ for any $r \neq s$, it is possible to resolve a particular $\mathbf{Q}_k$ independently of $\mathbf{R}$ (see Section 3.2.2), provided there are sufficient data, and provided $\mathbf{Q}_k$ is observable in the sense that $\mathbf{H}\mathbf{A}^s\mathbf{Q}_k\mathbf{A}^{s\prime}\mathbf{H}^\mathrm{T} \neq \mathbf{0}$

for several $s \geq 1$. An equation of type (3.10) can also be written in terms of the lag covariance,

$$\mathbf{Y}_s \equiv \langle \mathbf{y}(t+s)\mathbf{y}(t)^{\mathrm{T}} \rangle = \mathbf{H}\mathbf{A}^s\mathbf{P}\mathbf{H}^{\mathrm{T}}. \tag{3.11}$$

Whether it is preferable to use lag rather than lag-difference covariance matrices, that is, $\mathbf{Y}_s$ rather than $\mathbf{D}_s$, is addressed in section 3.3.

## 3.2.2 Maximum Number of Resolvable Parameters

If the number of observations is less than the dimension of the state ($M \leq N$) the maximum number of parameters ($K$ and $L$ defined in equation 3.4) which can be estimated by using the CMA is:

$$K_{\mathrm{max}} = M\left(N + 1 - (M + 1)/2\right), \tag{3.12}$$

when only the model error covariance $\mathbf{Q}$ is estimated; and

$$K_{\mathrm{max}} + L_{\mathrm{max}} = M(N + 1), \tag{3.13}$$

when both the model and measurement error covariances are estimated.

First, we consider the case when $\mathbf{R}$ is not estimated. The maximum number of parameters is given by the rank of the matrix $\mathcal{G}$, defined in equation (3.10). This matrix is made up by the Green functions $\mathbf{G}_{\mathbf{Y},k}(:), \mathbf{G}_{\mathbf{D}_1,k}(:)$, etc. After subtracting twice the first row $\mathbf{G}_{\mathbf{Y},1}(:), \ldots, \mathbf{G}_{\mathbf{Y},K}(:)$ from the negative of all the other rows we find that the rank of $\mathcal{G}$ is equal to the rank of the following matrix

$$\mathcal{F} = \begin{bmatrix} \mathbf{F}_{0,1}(:) & \ldots & \mathbf{F}_{0,K}(:) \\ \mathbf{F}_{1,1}(:) & \ldots & \mathbf{F}_{1,K}(:) \\ & \vdots & \\ \mathbf{F}_{S,1}(:) & \ldots & \mathbf{F}_{S,K}(:) \end{bmatrix}, \tag{3.14}$$

where $\mathbf{F}_{l,k} = \mathbf{H}\left(\mathbf{A}^l\mathbf{S}_k + \mathbf{S}_k(\mathbf{A}^l)^{\mathrm{T}}\right)\mathbf{H}^{\mathrm{T}}. \tag{3.15}$

To prove that the maximum rank of $\mathcal{F}$ is given by equation (3.12) we use the fact, proven in Appendix E, that the dimension of the null space of the operator

$$\mathbf{H}\left(\mathbf{A}^l \cdot + \cdot (\mathbf{A}^l)^{\mathrm{T}}\right)\mathbf{H}^{\mathrm{T}} \tag{3.16}$$

is equal to the number of linearly independent symmetric matrices of the dimension $N - M$ $((N - M)(N - M + 1)/2)$. The rank of $\mathcal{F}$ is equal to the difference between the dimension of the domain, equal to the number of linearly independent symmetric matrices of size $N$, and the dimension of the null space,

$$K_{\mathrm{max}} = N(N + 1)/2 - (N - M)(N - M + 1)/2 = M\left(N + 1 - (M + 1)/2\right)$$

Of course, the matrix of the Green functions, $\mathcal{G}$ defined in equation (3.10), must have at least $M\left(N + 1 - (M + 1)/2\right)$ rows to achieve the maximum rank. When the operator $(\vdots)$ represents upper triangular of the covariance matrix (symmetric matrix), i.e. there are $M(M + 1)$ elements in $\mathbf{Y}(\vdots)$,

$$S \geq \frac{M\left(N + 1 - (M + 1)/2\right)}{(M + 1)M/2} = \frac{2N + 1 - M}{M + 1}$$

The proof for the case when both the state and the measurement noise covariance are estimated is analogous to the one above. The only difference is that there are $M(M+1)/2$ additional columns in $\mathcal{F}$ and therefore the maximum rank of $\mathcal{G}$ is increased to

$$K_{\mathrm{max}} + L_{\mathrm{max}} = N(N + 1)/2 - (N - M)(N - M + 1) + M(M + 1)/2 = M(N + 1)$$

If the number of independent observations is greater than the dimension of the state $(M > N)$, then assuming that there are no temporal correlations in the model and measurement errors at some lag $S$, all elements of $\mathbf{Q}$ and $\mathbf{R}$ can be estimated by using the lag-difference covariance, $\mathbf{D}_s$.

## 3.3 Finite Number of Measurements

The discussion so far has assumed that covariance matrices $\mathbf{Y}$ and $\mathbf{D}_s$ are exact. In practice, a finite number of measurements is available and we work with sample estimates

$\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{D}}_s$, the sample covariance of $\mathbf{y}(t)$ is

$$\tilde{\mathbf{Y}} \equiv \frac{1}{T} \sum_{t=1}^{T} [\mathbf{y}(t) - \overline{\mathbf{y}}][\mathbf{y}(t) - \overline{\mathbf{y}}]^{\mathrm{T}}, \qquad (3.17)$$

where $T$ is the total number of time steps and

$$\overline{\mathbf{y}} \equiv \frac{1}{T} \sum_{t=1}^{T} \mathbf{y}(t) \qquad (3.18)$$

is the sample mean.

The first algorithmic modification required concerns the computation of Green's functions. If $T$ spans less than about 20 $e$-folding periods for each observable normal mode of linear system $\mathbf{p}(t+1) = \mathbf{A}\mathbf{p}(t)$, the steady state limit given by the solution to the Lyapunov equation (3.5) will be inaccurate. A Monte Carlo approach can instead be used to estimate $\mathbf{P}_k$ by driving linear model (3.1) with random system noise generated using covariance $\mathbf{Q}_k$; $\mathbf{P}_k$ is estimated by averaging over a large number of independent simulations, each with finite time span $T$.

A second modification is required to accommodate uncertainty in the sample covariance matrices. This is achieved by appending an error term, vector $\boldsymbol{\varepsilon}$, to equation (3.10):

$$\mathbf{d} = \mathcal{G}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}. \qquad (3.19)$$

The probability distribution of sample covariance matrices is approximately normal (Section 3.3.1); therefore it is appropriate to use variance minimizing methods. For example, parameter vector $\boldsymbol{\alpha}$ in equation (3.19) can be determined by minimizing the weighted least-squares cost function,

$$J(\boldsymbol{\alpha}) = \boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{R}_{\varepsilon}^{-1} \boldsymbol{\varepsilon} + (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{\mathrm{o}})^{\mathrm{T}} \mathbf{R}_{\alpha}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{\mathrm{o}}), \qquad (3.20)$$

where $\boldsymbol{\alpha}_{\mathrm{o}}$, $\mathbf{R}_{\alpha}$, and $\mathbf{R}_{\varepsilon}$ represent prior knowledge for $\langle \boldsymbol{\alpha} \rangle$, cov $\boldsymbol{\alpha}$, and cov $\boldsymbol{\varepsilon}$, respectively.

The uncertainty variance of a sample covariance is $\mathrm{O}[\sigma_1^2 \sigma_2^2 (1 + \rho^2)/p]$, where $\sigma_1^2$ and $\sigma_2^2$ denote variances for the two random variables being compared, $\rho$ is the correlation coefficient, and $p$ is the number of degrees of freedom, approximately the number of

independent measurements (Section 3.3.1). It follows that for a given sample size, the smaller the variances, the more accurately sample covariances can be determined.

For example, in the twin experiments of Section 4.3.2, statistically significant error estimates are possible using lag-diference covariance matrices, $\mathbf{D}_s$, but not with lag covariance matrices, $\mathbf{Y}_s$. In those experiments the errors propagate slowly relative to the duration of a time step, that is, the state transition matrix is approximately identity, so that, for small values of lag $s$, (3.9) simplifies to $\mathbf{D}_s \approx s\mathbf{HQH}^\mathrm{T} + 2\mathbf{R}$. The sample uncertainty of $\tilde{\mathbf{D}}_s$ therefore scales with the diagonal elements of $(s\mathbf{HQH}^\mathrm{T} + 2\mathbf{R})$. By comparison, the uncertainty of $\tilde{\mathbf{Y}}_s$ scales with the diagonal elements of $(\mathbf{HPH}^\mathrm{T} + \mathbf{R})$ which are much larger. As a rule of thumb, it is preferable to work with $\tilde{\mathbf{D}}_s$ when $\mathbf{A} \approx \mathbf{I}$ and $|\mathbf{R}| \ll |\mathbf{HPH}^\mathrm{T}|$.

Next we turn to a discussion of how to implement these changes in practice. A numerical example follows in Section 3.4.

### 3.3.1 Uncertainty of a Sample Covariance

The uncertainty of the sample estimates of the data covariance, $\mathbf{R}_\varepsilon$ in equation (3.20), can be estimated using the theory of distributions for sample estimators, see Anderson (1971). We assume that the process, $\mathbf{y}(t)$, is stationary and Gaussian. Accordingly, it is completely described by the first and second order moments. We derive expressions which give uncertainty of the sample covariance (of any lag $h$) in terms of the covariances themselves.

For a scalar Gaussian time series $\mathbf{y}(t)$, the covariance of the sample covariance [1] is

---

[1]The sample covariance of $\mathbf{y}(t + h)$ and $\mathbf{y}(t)$ is defined as

$$\tilde{\mathbf{Y}}(h) \equiv \frac{1}{T - h} \sum_{t=1}^{T-h} [\mathbf{y}(t + h) - \bar{\mathbf{y}}_{h+}] [\mathbf{y}(t) - \bar{\mathbf{y}}_h]^\mathrm{T}, h \in \{0, \ldots, T - 2\}, \tag{3.21}$$

where $T$ is the total number of time steps and

$$\bar{\mathbf{y}}_h = \frac{1}{T - h} \sum_{t=1}^{T-h} \mathbf{y}(t), h = 0, 1, \ldots, T - 2, \tag{3.22}$$

defined as

$$\text{cov}\ (\sigma(q), \sigma(r)) \equiv\ <[\sigma(q) - <\sigma(q)>][\sigma(r) - <\sigma(r)>]>, \qquad (3.23)$$

where we modified the notation for the variance to that which is commonly used with scalar time series. It can be obtained by substituting the sample estimate (3.21) into the definition above and computing the expectation of the entire sum (see Anderson (1971), p. 452),

$$(T - q)\ (T - r)\ \text{cov}\ (\sigma(q), \sigma(r)) \qquad (3.24)$$

$$\sum_{t=1}^{T-q} \sum_{t'=1}^{T-r} [\sigma(t - t')\sigma(t + q - t' - r) + \sigma(t - t' - r)\sigma(t + q - t')] -$$

$$\frac{1}{T - r} \sum_{t=1}^{T-q} \sum_{t',s'=1}^{T-r} [\sigma(t - t')\sigma(t + q - s' - r) + \sigma(t - s' - r)\sigma(t + q - t')] -$$

$$\frac{1}{T - q} \sum_{t,s=1}^{T-q} \sum_{t'=1}^{T-r} [\sigma(t - t')\sigma(s + q - t' - r) + \sigma(t - t' - r)\sigma(s + q - t')] +$$

$$\frac{1}{(T - q)(T - r)} \sum_{t,s=1}^{T-q} \sum_{t',s'=1}^{T-r} [\sigma(t - t')\sigma(s + q - s' - r) + \sigma(t - s' - r)\sigma(s + q - t')].$$

In addition, Anderson (1971) also shows that the distribution of $\text{cov}\ (\sigma(q), \sigma(r))$ is approximately normal.

For a vector time series the covariance of the sample covariance is a tensor of rank 4. Therefore, we present the element-wise analog to equation (3.25). The difficulty is mainly in keeping track of the indices.

Denote the $(i, j)$ element of $\mathbf{Y}(q)$ by $\mathbf{Y}_{(i,j)}(q)$. Then, the resulting equation for the covariance of the estimators $\mathbf{Y}_{(i,j)}(q)$ and $\mathbf{Y}_{(l,k)}(r)$ is given by

$$(T - q)(T - r)\ \text{cov}\left[\mathbf{Y}_{(i,j)}(q)\ \mathbf{Y}_{(k,l)}(r)\right] = \qquad (3.25)$$

$$\sum_{s=1}^{T-q} \sum_{u=1}^{T-r} \left[\hat{\mathbf{Y}}_{(i,k)}(s - u)\ \hat{\mathbf{Y}}_{(j,l)}(s + q - u - r) + \hat{\mathbf{Y}}_{(i,l)}(s - u - r)\ \hat{\mathbf{Y}}_{(j,k)}(s + q - u)\right] -$$

$$\bar{y}_{h+} = \frac{1}{T - h} \sum_{t=1}^{T-h} y(t + h), h = 0, 1, \ldots, T - 2$$

are the sample means.

106

$$\frac{1}{T-r} \sum_{s=1}^{T-q} \sum_{u,v=1}^{T-r} \left[ \hat{\mathbf{Y}}_{(i,k)}(s-u) \, \hat{\mathbf{Y}}_{(j,l)}(s+q-v-r) + \hat{\mathbf{Y}}_{(i,l)}(s-v-r) \, \hat{\mathbf{Y}}_{(j,k)}(s+q-u) \right] -$$

$$\frac{1}{T-q} \sum_{s,t=1}^{T-q} \sum_{u=1}^{T-r} \left[ \hat{\mathbf{Y}}_{(i,k)}(s-u) \, \hat{\mathbf{Y}}_{(j,l)}(t+q-u-r) + \hat{\mathbf{Y}}_{(i,l)}(s-u-r) \, \hat{\mathbf{Y}}_{(j,k)}(t+q-u) \right] +$$

$$\frac{1}{(T-q)(T-r)} \sum_{s,t=1}^{T-q} \sum_{u,v=1}^{T-r} \left[ \hat{\mathbf{Y}}_{(i,k)}(s-u) \, \hat{\mathbf{Y}}_{(j,l)}(t+q-v-r) + \hat{\mathbf{Y}}_{(i,l)}(s-v-r) \, \hat{\mathbf{Y}}_{(j,k)}(t+q-u) \right],$$

where true covariance is denoted by the hat:

$$\hat{\mathbf{Y}}(h) = < [\mathbf{y}(t+h) - < \mathbf{y}(t) >][\mathbf{y}(t) - < \mathbf{y}(t) >]^{\mathrm{T}} > . \tag{3.26}$$

In practice, to calculate the uncertainty of the sample lag covariances we substitute sample estimates, equation (3.21), for the true ones $\hat{\mathbf{Y}}_{(i,k)}(h)$ into equation (3.25).

Uncertainty for covariance of lag $s$ difference, $\boldsymbol{D}_s$, defined by

$$\mathbf{D}_s \equiv \mathrm{cov}[\mathbf{y}(t+s) - \mathbf{y}(t)]$$

can be computed by observing that

$$\boldsymbol{D}(s) = 2\mathbf{Y}(0) - (\mathbf{Y}(s) + \mathbf{Y}(s)^{\mathrm{T}}).$$

Therefore, using the relation for the covariance of the sum we obtain

$$
\begin{aligned}
\mathrm{cov}\left(\boldsymbol{D}(q), \boldsymbol{D}(r)\right) \quad &= 4\,\mathrm{cov}\left(\mathbf{Y}(0), \mathbf{Y}(0)\right) - 2\,\mathrm{cov}\left(\mathbf{Y}(0), \mathbf{Y}(r)\right) - 2\,\mathrm{cov}\left(\mathbf{Y}(0), \mathbf{Y}(r)^{\mathrm{T}}\right) \\
&\quad + \mathrm{cov}\left(\mathbf{Y}(q), \mathbf{Y}(r)\right) + \mathrm{cov}\left(\mathbf{Y}(q)^{\mathrm{T}}, \mathbf{Y}(r)^{\mathrm{T}}\right) - 2\,\mathrm{cov}\left(\mathbf{Y}(q), \mathbf{Y}(0)\right) \\
&\quad -2\,\mathrm{cov}\left(\mathbf{Y}(q)^{\mathrm{T}}, \mathbf{Y}(0)\right) - \mathrm{cov}\left(\mathbf{Y}(q), \mathbf{Y}(r)^{\mathrm{T}}\right) - \mathrm{cov}\left(\mathbf{Y}(q)^{\mathrm{T}}, \mathbf{Y}(r)\right).
\end{aligned}
$$

If the process is non-normal, we would have to include fourth-order cumulants, which were assumed to be zero in derivation of the equation equation (3.25). In practice, for short time series, sample estimates of fourth order cumulants are very noisy, and one benefits by setting them to zero.

A useful approximation is

$$\mathrm{cov}\left(\tilde{\mathbf{Y}}_{(i,j)}(0), \ \tilde{\mathbf{Y}}_{(k,l)}(0)\right) \approx \frac{1}{p}\left[\tilde{\mathbf{Y}}_{(i,k)}(0)\, \tilde{\mathbf{Y}}_{(j,l)}(0) + \tilde{\mathbf{Y}}_{(i,l)}(0)\, \tilde{\mathbf{Y}}_{(j,k)}(0)\right], \tag{3.27}$$

where $p \leq T$ is the decorrelation number (roughly, the number of time steps $T$ divided by the $e$-folding correlation time). This equation provides an extension of a well-known result for a white (in time) time series where the covariance of the diagonal elements of data covariance is equal to:

$$\text{cov}\left(\mathbf{Y}_{(i,i)}(0), \mathbf{Y}_{(k,k)}(0)\right) = \frac{2(T-1)}{T^2}(\hat{\mathbf{Y}}_{(i,k)}(0))^2,$$

i.e. the decorrelation scale is set to T.

Validity of the general formula (3.25) was tested by Monte Carlo experiments with different time series and different time and space correlation structure. (MATLAB-callable software for estimating uncertainty of sample covariance is available via anonymous FTP to gulf.mit.edu, IP Address 18.83.0.149, from directory pub/misha. It is described in Appendix F.)

### 3.3.2 Lyapunov Equation

Solving the Lyapunov equation (3.5),

$$\mathbf{P} \equiv \text{cov} \, \mathbf{p} = \mathbf{A}\mathbf{P}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}, \tag{3.28}$$

for systems with more than 500 variables is non-trivial. We found that the following approach from Gajic and Qureshi (1995), is best suited for the case when we need to solve the equation for several different matrices $\mathbf{Q}_k$ (equations (3.5) and (3.7)).

First, compute the Jordan form of the matrix $\mathbf{A}$ by finding its eigenvectors and eigenvalues,

$$\mathbf{A} = \mathbf{V} \, \mathbf{\Lambda} \, \mathbf{V}^{-1}. \tag{3.29}$$

Next, we multiply the Lyapunov equation (3.5) by $\mathbf{V}^{-1}$ on the left and $\mathbf{V}$ on the right to obtain:

$$\mathbf{\Lambda}\mathbf{J}\mathbf{\Lambda}^{\mathrm{T}} - \mathbf{J} = -\mathbf{Z}, \quad \text{where } \mathbf{J} = \mathbf{V}^{-1}\mathbf{P}(\mathbf{V}^{-1})^{\mathrm{T}}, \; \mathbf{Z} = \mathbf{V}^{-1}\mathbf{Q}(\mathbf{V}^{-1})^{\mathrm{T}}. \tag{3.30}$$

108

This equation has block structure, and each block can be solved individually. The resulting covariance $\mathbf{P}$ is easily found by applying the transformation inverse to equation (3.30).

When all eigenvalues of $\mathbf{A}$ are distinct, equation (3.30) breaks down into element-wise group of equations:

$$\mathbf{J}[i,j] = \mathbf{Z}[i,j]/(1 - \lambda_i \lambda_j),, \quad \text{where} \ \ \mathbf{\Lambda} = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_N). \tag{3.31}$$

We also note that computing inverse of the matrix of eigenvectors can be unstable. When all eigenvalues of $\mathbf{A}$ are distinct we can instead compute left eigenvectors, i.e. eigenvectors of the transpose $\mathbf{A}^{\mathrm{T}}$, columns of $\mathbf{W}$. Because eigenvectors are defined up to an arbitrary constant, the inverse of $\mathbf{V}$ is

$$\mathbf{V}^{-1} = \mathbf{W}^{\mathrm{T}}\,\mathbf{S}, \text{where} \ \ \mathbf{S}_{i,i} = \mathrm{diag}\,(v_i\ w_i^{\mathrm{T}}), \mathbf{V} = (v_1, \ldots, v_N), \mathbf{W} = (w_1, \ldots, w_N). \tag{3.32}$$

This method is much faster than the traditional ones. Significant computational saving comes from the fact that once the Jordan form of $\mathbf{A}$ is computed and stored, six multiplications with complex matrices of size $N$ are sufficient to solve the Lyapunov equation.

When $T$ is short, i.e. it spans few $(> 20)$ $e$-folding periods for each observable eigenmode of linear system $\mathbf{p}(t+1) = \mathbf{A}\mathbf{p}(t)$, a Monte Carlo approach can instead be used to estimate $\mathbf{P}$. Namely, we propagate random noise from normal distribution with zero mean and covariance $\mathbf{Q}$ through the system (3.1), as described in Section 2.8. The sample covariance of the resulting state gives a sample $\mathbf{P}$. Repeating this procedure with many different realizations of the random noise we obtain an approximation to the solution of the Lyapunov equation.

### 3.3.3 The Column Operator ($\vdots$)

The column operator (equation (3.10)) denotes reshaping a full symmetric matrix of size $M$ into a vector of size $M(M+1)/2$. If the number of observations $M$ is large, the full matrices need to be replaced by a representative subset of the covariance matrices (for

109

example, diagonal elements, or only the first two diagonals). A different subset of matrix elements can be taken for different matrices $\mathbf{Y}, \mathbf{D}_1$, etc, as long as it is the same on both sides of equation (3.10). Physical considerations of the problem should guide what elements are used. For example, if one believes that spatial correlation is significant, off-diagonal elements should be used, and vice versa. The need to provide corresponding uncertainties provides an additional constraint on what linear combinations can be used in the CMA (Section 3.3.1).

## 3.4 Numerical Example with a Covariance Matching Approach

The covariance matching recipe is next illustrated using a numerical example. Consider the system of equations (3.1), (3.2) with

$$\mathbf{A} = \begin{bmatrix} 0.8 & 0.2 \\ -0.1 & 0.9 \end{bmatrix}, \quad \mathbf{B} = \mathbf{I}, \quad \mathbf{H} = \begin{bmatrix} 1 & 1 \end{bmatrix}. \tag{3.33}$$

The system and measurement error covariance matrices are parameterized

$$\mathbf{Q} = \alpha_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{R} = \alpha_4. \tag{3.34}$$

From the steady-state Lyapunov equation (3.5) we obtain the covariance matrices

$$\mathbf{P}_1 = \begin{bmatrix} 2.5 & -0.4 \\ -0.4 & 0.5 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 1.9 & 1.7 \\ 1.7 & 3.7 \end{bmatrix}, \quad \mathbf{P}_3 = \begin{bmatrix} 5.9 & 3.2 \\ 3.2 & 2.5 \end{bmatrix}, \tag{3.35}$$

corresponding to unit perturbations of parameters $\alpha_1$, $\alpha_2$, and $\alpha_3$, respectively, in (3.34). Computing the Green's functions associated with $\mathbf{Y}$ and $\mathbf{D}_s$ results in the following

110

system of equations:

$$
\begin{bmatrix} \mathbf{Y} \\ \mathbf{D}_1 \\ \mathbf{D}_2 \\ \mathbf{D}_3 \end{bmatrix} = \begin{bmatrix} 2.2 & 9.1 & 14.9 & 1 \\ 1.3 & 1.1 & 4.4 & 2 \\ 2.3 & 2.6 & 8.6 & 2 \\ 3.2 & 4.4 & 12.6 & 2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}. \tag{3.36}
$$

The kernel matrix in (3.36) has rank 3 (singular values $[24.6 \ \ 3.7 \ \ 0.9 \ \ 0.0]^{\mathrm{T}}$) which indicates that only three independent combinations of parameters $\alpha_k$ can be resolved. It turns out that the addition of $\mathbf{D}_3$, or of higher lag covariance matrices, does not contribute new information. Rules regarding the total number of resolvable parameters are set forward in Appendix 3.2.2.

Simulated data were generated for $\boldsymbol{\alpha} = [1 \ 1 \ 0 \ 1]^{\mathrm{T}}$, $T = 500$. We seek to estimate $\boldsymbol{\alpha}$ using the simulated data and the recipe of Section 3.2.1. From inverse theory, only projections onto singular vectors of $\mathcal{G}$ corresponding to non-zero singular values can be determined (Wunsch, 1996; p.147 ). The full solution is

$$
\tilde{\boldsymbol{\alpha}} = [0.08 \ \ 0.25 \ \ 0.47 \ \ 1.00]^{\mathrm{T}} + \lambda \, [-0.88 \ \ -0.34 \ \ 0.34 \ \ 0.00]^{\mathrm{T}}, \tag{3.37}
$$

where $\lambda$ is an arbitrary constant multiplying null space contributions; $\lambda$ cannot be determined without additional information. To set $\lambda$ we assume that there is a priori knowledge that the system error covariance matrix is diagonal, that is, $\alpha_3 = 0$. This assumption requires that $\lambda = -1.4$ and hence that $\tilde{\boldsymbol{\alpha}} = [1.3 \ \ 0.7 \ \ 0 \ \ 1.0]^{\mathrm{T}}$.

Next we seek to estimate the solution uncertainty, $\mathbf{P}_\alpha \equiv \mathrm{cov} \ \tilde{\boldsymbol{\alpha}}$. Formally $\mathbf{P}_\alpha$ is a function of a priori covariance matrices $\mathbf{R}_\varepsilon$ and $\mathbf{R}_\alpha$ in (3.20). $\mathbf{R}_\alpha$ is the a priori covariance of parameter vector $\boldsymbol{\alpha}$, the only a priori knowledge assumed here being that $\alpha_3 = 0$. Matrix $\mathbf{R}_\varepsilon$ describes the sample uncertainty of $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{D}}_s$. An estimate of $\mathbf{R}_\varepsilon$, consistent with the available data, can be obtained using the expressions derived in

Section 3.3.1:

$$\mathbf{R}_{\varepsilon} \approx \begin{bmatrix} 1.8 & 0.1 & 0.2 & 0.5 \\ 0.1 & 0.1 & 0.1 & 0.1 \\ 0.2 & 0.1 & 0.2 & 0.3 \\ 0.5 & 0.1 & 0.3 & 0.6 \end{bmatrix}. \tag{3.38}$$

The solution uncertainty matrix is $\mathbf{P}_{\alpha} = (\mathcal{G}^{\mathrm{T}} \mathbf{R}_{\varepsilon}^{-1} \mathcal{G})^{-1}$. Therefore $\tilde{\alpha} = [1.3 \pm 0.4 \quad 0.7 \pm 0.2 \quad 0 \quad 1.0 \pm 0.2]^{\mathrm{T}}$, consistent with the parameter vector $\alpha = [1 \ 1 \ 0 \ 1]^{\mathrm{T}}$ used to generate the simulated data. (Unless otherwise specified, uncertainty is reported using one standard deviation.)

From a set of numerical experiments, like that above, we conclude that the covariance matching method gives consistent and statistically significant estimates, provided the total number of available measurements is much greater than the number of parameters $\alpha_k$, that is, $MT \gg K + L$, where $M$ is the length of the measurement vector, $T$ is the number of time steps, and $K + L$ is the total number of parameters in (3.4), (3.4). The requirement for a large number of observations per parameter is a direct consequence of the large uncertainty of sample covariance matrices.

What happens if instead of assuming $\alpha_3 = 0$, which is the condition used to generate the simulated data, it is instead assumed that $\alpha_1 = 0$? This assumption implies that $\lambda = 0.09$ in (3.37) and leads to a second solution $\tilde{\alpha} = [0 \quad 0.2 \pm 0.3 \quad 0.5 \pm 0.2 \quad 1.0 \pm 0.2]^{\mathrm{T}}$. From the data alone there is no way to decide whether this solution is better or worse than the previous one. In fact, there exist a large number of consistent solutions depending on particular choices of $\lambda$ and of other a priori assumptions. For this particular example, a second independent measurement at every time step would permit $\mathbf{Q}$ to be determined uniquely. But for real oceanographic problems there is rarely, if ever, sufficient data to fully determine $\mathbf{Q}$, and one must therefore rely on physical intuition to choose suitable models for the errors.

(MATLAB script files and functions which implement this example, and which can be customized for different applications, are available by contacting the author or via anony-

mous FTP to gulf.mit.edu, IP Address 18.83.0.149, from directory pub/misha/Paper1/.)

### 3.4.1 Comparison of the Numerical Example with the CMA and the MT Method

The analysis of the MT method with 2 DOF models presented in Section 2.7 allows us to contrast the two adaptive methods. First, the MT method is unstable when we try to estimate both model and measurement error covariances, and we had to provide a guess for $\mathbf{R}$ and estimate only $\mathbf{Q}$. The CMA algorithm can estimate both error covariances. Second, the CMA method does not require an initial estimate of the model error covariance, unlike the MT algorithm. In fact, when we tried to estimate two diagonal elements of $\mathbf{Q}$ in the same setup as above, the MT method estimates were sensitive to the initial guess for $\mathbf{Q}$ (Figure 2.10). Furthermore, the MT method does not provide estimates of the uncertainty of the estimates for $\boldsymbol{\alpha}$, while the CMA method does.

In addition, we ran a similar numerical experiment with non-zero elements on the diagonal, i.e. $\alpha_3 = 0.5$, but CMA estimates obtained assuming $\alpha_3 = 0$.. The resulting estimates $\tilde{\boldsymbol{\alpha}} = [2 \pm 0.5 \quad 0.7 \pm 0.3 \quad 0 \quad 1.0 \pm 0.3]^{\mathrm{T}}$. This can be contrasted with Figure 2.15, where a similar experiment is shown with the MT method (using infinitely long time series of observations). The CMA shows that in this case the estimates $\boldsymbol{\alpha}$ are worse since the uncertainty is increased. The MT method does not provide an estimate of the uncertainty, or any other information on the reliability of the estimates.

## 3.5 Extensions of Covariance Matching

The same approach can be used for estimation of other statistics, such as systematic errors and time correlated errors.

## 3.5.1 Systematic Errors

Systematic errors, or biases, refer to the quantities $\langle \mathbf{r}(t) \rangle$ and $\langle \mathbf{u}(t) \rangle$. These errors are important because, even if very small, they can accumulate over long numerical integrations and degrade the predictive skill of a model. A first scenario is that of a stable, time-independent system, as before, but with $\langle \mathbf{r} \rangle \neq \mathbf{0}$, $\langle \mathbf{u} \rangle \neq \mathbf{0}$. Notice that the estimators which have been developed for $\mathbf{R}$, $\mathbf{Q}$, and $\mathbf{P}$ are not, to first order, affected by the presence of measurement and model biases because the sample mean is subtracted from the data in (3.21) and because the biases cancel out when computing lagged data differences.

Model bias correction in the context of atmospheric data assimilation was recently discussed by Evensen et al. (1998) and by Dee and da Silva (1997). They described on-line algorithms suitable for sequential estimation approaches. Off-line algorithms, whereby biases are removed prior to data assimilation, are also available, e.g. Fukumori et al.(1999), and are discussed below for completeness. From equations (3.1) and (3.2) it follows that

$$\langle \mathbf{y} \rangle = \mathbf{H}(\mathbf{I} - \mathbf{A})^{-1} \langle \mathbf{u} \rangle + \langle \mathbf{r} \rangle, \tag{3.39}$$

that is $\langle \mathbf{y} \rangle$ is linearly related to the biases, $\langle \mathbf{q} \rangle$ and $\langle \mathbf{r} \rangle$. The sample mean, $\overline{\mathbf{y}}$ in equation (3.22), is an unbiased estimator of $\langle \mathbf{y} \rangle$ with uncertainty (Anderson, 1971 ; Section 8.2),

$$\text{cov}\,(\overline{\mathbf{y}}) = \sum_{r=1-T}^{T-1} \left( \frac{T - |r|}{T^2} \right) \langle \mathbf{y}(t+r)\mathbf{y}(t)^{\mathrm{T}} \rangle, \tag{3.40}$$

which, using equations (3.1), (3.2)), and (3.5), reduces to

$$\text{cov}\,(\overline{\mathbf{y}}) = \frac{\mathbf{Y}}{T} + \sum_{r=1}^{T-1} \left( \frac{T - r}{T^2} \right) \left( \mathbf{H}\mathbf{A}^r\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{H}\mathbf{P}(\mathbf{A}^r)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} \right), \tag{3.41}$$

where $\mathbf{Y}$ is the data covariance matrix(equation 3.6) and $\mathbf{P}$ is the GCM error covariance matrix equation (3.5), estimates for both matrices having been obtained earlier. Without additional information, it is not possible to discriminate between system bias $\langle \mathbf{u} \rangle$ and data bias $\langle \mathbf{r} \rangle$.

114

A second scenario, that of a gradual change, or trend, in the system error, is discussed in Section 3.5.3 which deals with time dependent models.

## 3.5.2 Time Correlated Errors

So far we have assumed that measurement and system errors are uncorrelated in time, that is $\langle \mathbf{r}(t_1) \mathbf{r}(t_2)^{\mathrm{T}} \rangle = \mathbf{0}$, $\langle \mathbf{u}(t_1) \mathbf{u}(t_2)^{\mathrm{T}} \rangle = \mathbf{0}$ for $t_1 \neq t_2$. These conditions are required to evaluate lag-difference covariance matrices equation (3.9) — but it is not required to evaluate the data covariance matrix $\mathbf{Y}$ which can therefore be used as before. The latter condition is implicit in equation (3.3) and presents a more difficult modeling challenge. Under the Kalman filter formalism this situation is usually addressed by appending additional parameters to the state vector and jointly estimating time-correlated and uncorrelated errors.

Under the Kalman filter formalism this situation is usually addressed by appending additional parameters to the state vector and jointly estimating time-correlated and uncorrelated errors. These parameters can also be estimated off-line. Consider for example the specific case of an annual cycle in the system and/or in the measurement errors, a situation which is of direct practical relevance to oceanographic applications. Taking the Fourier transform of equations (3.1) and (3.2), it follows that each frequency component of $\mathbf{y}(t)$ is linearly related to the same frequency component in $\mathbf{u}(t)$ and $\mathbf{r}(t)$,

$$\mathbf{y}_a = \mathbf{H}\mathbf{p}_a + \mathbf{r}_a, \tag{3.42}$$

$$\mathbf{p}_a \exp(i\omega/12) = \mathbf{A}\mathbf{p}_a + \mathbf{u}_a, \tag{3.43}$$

where the subscript $a$ indicates the complex annual cycle amplitude, that is, $\mathbf{y}_a = \mathbf{a}\exp(i\phi)$, $\mathbf{a}$ is the amplitude, $\phi$ is the phase, $\omega = 2\pi/\text{year}$. We have assumed a time step of 1 month in equation (3.43). Again, additional information is required to partition the annual cycle error between system and data errors. The important point, however, is that it can be removed from the model-data residual to avoid biasing estimates of second order statistics.

### 3.5.3 Time Dependent Models

We consider two types of time dependence. The first type is "known" time dependencies in the linear models, $\mathbf{A}$, $\mathbf{\Gamma}$ and $\mathbf{H}$, and also possibly in the measurement error covariance matrix, $\mathbf{R}$. These are readily accommodated by using a Monte Carlo approach to compute the Green's functions, $\mathbf{G}_{\mathbf{D},k}$. An example of this approach, with a time-varying $\mathbf{H}(t)$, is the treatment of acoustic time series of differing lengths in Section 4.4.

The second type of time dependence is due to fluctuations of the "unknown" model parameters, $\alpha_k(t)$ in equation (3.4). In principle, this situation can be addressed through piecewise estimates of $\alpha_k(t)$ for periods that are short relative to the time scales over which $\alpha_k$ varies. A better approach is to parameterize the time dependency and to estimate these parameters using all the available data. An example is the detection of a trend, $\langle \partial \mathbf{u}/\partial t \rangle \neq \mathbf{0}$, in the system error. From equation (3.39), and assuming $\langle \partial \mathbf{r}/\partial t \rangle = \mathbf{0}$, the first difference of $\mathbf{u}(t)$ is related to the first difference of $\mathbf{y}(t)$ by

$$\langle \mathbf{y}(t+1) - \mathbf{y}(t) \rangle = \mathbf{H}(\mathbf{I} - \mathbf{A})^{-1} \langle \mathbf{u}(t+1) - \mathbf{u}(t) \rangle. \tag{3.44}$$

The expression $\langle \mathbf{y}(t+1) - \mathbf{y}(t) \rangle$ can be approximated using least-squares (or other suitable estimators) and in turn used to estimate the quantity $\langle \partial \mathbf{u}/\partial t \rangle$ (Section 4.4).

### 3.5.4 Tests of Consistency

The final step of any estimation study is to test the resulting estimates for statistical consistency with all prior assumptions. One possible test is the comparison of estimation residuals ($\varepsilon$ in equations 3.19, 3.20) to the expected posterior covariance,

$$\text{cov}\,\tilde{\varepsilon} = \mathbf{R}_\varepsilon \left( \mathbf{R}_\varepsilon \left( \mathcal{G} \mathbf{R}_\alpha \mathcal{G}^{\mathrm{T}} + \mathbf{R}_\varepsilon \right)^{-1} \right)^{\mathrm{T}}. \tag{3.45}$$

In addition, when $\mathbf{Q}$ and $\mathbf{R}$ are used in conjunction with a Kalman filter, whiteness tests can be applied to the innovation vectors, Daley (1992b).

The description of the algorithm is now complete. In the remainder of this chapter we compare the CMA with the innovation based methods discussed in Chapter 2. In

Chapter 4 we illustrate this algorithm by estimating the large scale ($>$1000 km) baroclinic errors in a particular implementation and linearization of a GCM, first with twin experiments (Section 4.3), and with real data (Section 4.4).

## 3.6 Comparison of CMA with Innovations Based Methods

In this section we present a short summary of covariance matching with innovations approach (CMIA, hereafter) and compare it to the CMA described in detail in Sections 3.2– 3.3. We start by introducing the innovations and CMIA, and then compare the two methods. In comparison to CMA, CMIA requires additional approximations and is more computationally expensive. Therefore, for the problem of interest CMA is preferred. Note that CMIA is very similar to other methods which use innovations, see Blanchet et al. (1997).

### 3.6.1 Covariance Matching with Innovations Approach (CMIA)

Matching sample lag covariances to their respective theoretical expectations has been first proposed in the context of Kalman filtering, Shellenbarger (1967) and Belanger (1974). Innovations are defined as a difference between observations and one-step Kalman filter forecast,

$$\mathbf{v}(t) = \mathbf{y}(t) - \mathbf{H}\tilde{\mathbf{p}}(t|t-1). \tag{3.46}$$

In the Kalman filter, the forecast $\tilde{\mathbf{p}}(t|t-1)$ is a one-step model forecast from the best estimate at time $t-1$, which is computed using all available information up to time $t-1$ (Section 2.4). Belanger showed that by linearizing the filtering problem, theoretical expectations for lag covariances of the innovations can be linearly related to the prior model and measurement error covariances, see Dee et al. (1985) for an insightful discussion. When the model and measurement error covariances are parameterized as in

117

equation (3.4), we obtain a linear relationship, similar to equation (3.10), between the sample estimate of the lag covariances of innovations and the parameters $\alpha_k$. One then proceeds as in Section 3.2 and finds the parameters which give the best fit to the linear equation.

### 3.6.2   Comparison of CMIA and CMA

In this section we explain the similarities and the differences between the two methods. The only difference between CMA, described in the main text, and CMIA, described above, is that the former uses the observations directly instead of using the innovations. In principle, observations and innovations contain the same statistical information, Kailath (1968). We illustrate this at the end of this section by deriving the equation for the covariance of observations from that for the covariance of innovations.

In practice, the two methods may produce different estimates. The difference is due to two factors. First, quality of sample estimates for the covariances and lag-$s$ covariances of observations and innovations may be different. In CMIA not only do we have to linearize the filter around the prior values of $\alpha$ in CMIA, we also need to use the prior values of $\alpha$ to compute the innovations. That is, CMIA is not guaranteed to converge if we start with a bad prior guess for $\alpha$, Moghaddamjoo and Kirlin (1993).

However, CMIA can have an advantage when the initial estimates are close to the true ones, as the innovation statistics are nearly white and we do not need to use higher lag-$s$ covariances. In such a case, use of innovations offers an advantage as all the information is compressed into the covariance of innovations. On the other hand, CMA requires use of higher order lag-$s$ covariances, and, therefore, requires more computation and can have greater error.

The second factor is the computational cost, and the approximations we have to make when the cost is prohibitively high. To compute the innovations one has to update the Kalman filter, which is computationally expensive, $O(N^3)$ operations (even if we resort to approximations, such as steady state Kalman filter, the cost is still very high). The

118

CMA does not require running the Kalman filter, and is therefore computationally much cheaper.

### 3.6.3 Comparison of the CMA and the MT Method

Relation between CMA and the MT algorithm, see Chapter 2, is similar to that between CMA and CMIA. The similarity is best seen by comparing equations derived for the MT method with low-dimensional models, Section 2.7, with those of the CMA. In the derivation of the MT approach we derived two equations (2.58-2.59) which relate the covariance and lag-covariances of the state to those of the observations. These equations are identical to those used in the CMA, equations (3.6-3.9), except that in the CMA we use covariance of the lag differences, $\mathbf{D}_s$, instead of the lag-covariance, $\mathbf{Y}_s$. In MT method we then relate statistics of innovations to the observations through Ricatti equation, and by computing the bias correction. These introduces additional uncertainties into the MT method as discussed in Section 2.7. Another important difference is that in CMA we use all observations directly instead of going to innovations. In situations when observational time series are short, and only a few iterations of the MT method are possible, uncertainty of the CMA estimates can be considerably smaller.

### 3.6.4 Illustration of the Similarity of CMIA and CMA

The two covariance matching approaches are in theory identical. We illustrate this by deriving the equation for the covariance of observations from that for the covariance of the innovations. This shows that when the sample covariances of observations and innovations are perfect, the estimates from the two methods are identical. Similar results can be derived for lag-$s$ covariances.

The goal is to obtain the expression for the covariance of the observations ( equation 3.6) from the equation for the covariance of the innovations derived in Belanger:

$$< \mathbf{v}(t)\mathbf{v}(t)^{\mathrm{T}} >= \mathbf{H}\Xi(t)\mathbf{H}^{\mathrm{T}} + \mathbf{R}; \quad \Xi(t) =< \boldsymbol{\xi}(t)\boldsymbol{\xi}(t)^{\mathrm{T}} >, \tag{3.47}$$

119

where the forecast error $\boldsymbol{\xi}(t)$ is defined as the difference between the forecast and the true state,

$$\boldsymbol{\xi}(t) = \tilde{\mathbf{p}}(t|t-1) - \hat{\mathbf{p}}(t). \tag{3.48}$$

Combining equations (3.46) and (3.48) we obtain the following sum:

$$< \mathbf{v}(t)\mathbf{v}(t)^{\mathrm{T}} > = < [\mathbf{y}(t) - \mathbf{H}\hat{\mathbf{p}}(t) - \mathbf{H}\boldsymbol{\xi}(t)] [\mathbf{y}(t) - \mathbf{H}\hat{\mathbf{p}}(t) - \mathbf{H}\boldsymbol{\xi}(t)]^{\mathrm{T}} >$$

$$= \; < \mathbf{y}(t)\mathbf{y}(t)^{\mathrm{T}} > + \mathbf{H} < \hat{\mathbf{p}}(t)\hat{\mathbf{p}}(t)^{\mathrm{T}} > \mathbf{H}^{\mathrm{T}} + \mathbf{H} < \boldsymbol{\xi}(t)\boldsymbol{\xi}(t)^{\mathrm{T}} > \mathbf{H}^{\mathrm{T}} \tag{3.49}$$

$$- \left( < (\mathbf{y}(t) - \mathbf{H}\hat{\mathbf{p}}(t)) \, \boldsymbol{\xi}(t)^{\mathrm{T}} > \mathbf{H}^{\mathrm{T}} + \text{ transpose} \right) \tag{3.50}$$

$$- \left( < \mathbf{y}(t)\hat{\mathbf{p}}(t)^{\mathrm{T}} > \mathbf{H} + \text{ transpose} \right). \tag{3.51}$$

The terms in equation (3.49) are given by their theoretical expectations:

$$< \mathbf{y}(t)\mathbf{y}(t)^{\mathrm{T}} > = \mathbf{Y}, < \hat{\mathbf{p}}(t)\hat{\mathbf{p}}(t)^{\mathrm{T}} > = \mathbf{P}, < \boldsymbol{\xi}(t)\boldsymbol{\xi}(t)^{\mathrm{T}} > = \boldsymbol{\Xi}(t). \tag{3.52}$$

The terms in equation (3.50) vanish because observation error is independent of the forecast:

$$< (\mathbf{y}(t) - \mathbf{H}\hat{\mathbf{p}}(t)) \, \boldsymbol{\xi}(t)^{\mathrm{T}} > = < \mathbf{r}(t)\boldsymbol{\xi}(t)^{\mathrm{T}} > = \mathbf{0}. \tag{3.53}$$

The terms in equation (3.51) can be computed as follows:

$$< \mathbf{y}(t)\hat{\mathbf{p}}(t)^{\mathrm{T}} > \mathbf{H}^{\mathrm{T}} = < (\mathbf{H}\hat{\mathbf{p}}(t)\hat{\mathbf{p}}(t)^{\mathrm{T}} > \mathbf{H}^{\mathrm{T}} + < (\mathbf{r}(t)\hat{\mathbf{p}}(t)^{\mathrm{T}} > \mathbf{H}^{\mathrm{T}} = \mathbf{HPH}^{\mathrm{T}}, \tag{3.54}$$

since observation error is independent of the true state. Combining equations (3.47), and (3.49 – 3.54), we obtain that

$$\mathbf{H}\boldsymbol{\Xi}(t)\mathbf{H}^{\mathrm{T}} + \mathbf{R} = \mathbf{Y} + \mathbf{HPH}^{\mathrm{T}} + \mathbf{H}\boldsymbol{\Xi}(t)\mathbf{H}^{\mathrm{T}} - \mathbf{0} - 2\mathbf{HPH}^{\mathrm{T}}, \tag{3.55}$$

which after the cancellations gives equation (3.6). Similar equations can be derived for the lag-difference covariances (Section 3.2.1), but lag-differences of the innovations are difficult to work with and have not been used in CMIA.

In summary, while the MT and other innovations based methods and CMA have common roots, the latter approach is more efficient and robust in situations when we have a large number of degrees of freedom and few observations, as confirmed by the twin experiments in Sections 2.8 and 4.3.

120

## 3.7 Summary

The CMA we propose is similar to the methods described by Shellenbarger (1966) and Belanger (1974), but we use GCM-data residuals directly instead of the innovation sequence. Innovation sequence approaches have been preferred by the engineering community because they are more readily amenable to online applications and to the tracking of slowly varying statistics. When first guess error statistics are accurate, the innovation sequence will be less correlated (in time) than the GCM-data residual and, therefore, the available information will collapse into a small number of lag covariance matrices.

However, for the systems of large dimension which are of interest to oceanographic studies, it is preferable to work with the GCM-data residual directly for the following reasons. First, sample covariances of the residuals can be computed offline, thus avoiding the computational burden associated with repeated integrations of the Kalman filter. Second, system and measurement error covariance matrices, $\mathbf{Q}$ and $\mathbf{R}$, are linearly related to those of the GCM-data residual, $\mathbf{Y}$ and $\mathbf{D}_s$. In contrast, the innovation sequence variants of the algorithm require linearization about some first guess error statistics and, therefore, convergence is not guaranteed, Moghaddamjoo and Kirlin (1993). Third, the GCM-data residuals contain information about absolute matrix norms $|\mathbf{Q}|$ and $|\mathbf{R}|$, while the innovation sequence can only be used to determine the relative ratio $|\mathbf{Q}|/|\mathbf{R}|$. The key to making direct use of the GCM-data residual is the use of lag-difference covariance matrices (equation 3.9). For a diagonally dominant model, matrix $\mathbf{A}$ in equation (3.1), the lag difference collapses useful statistical information to a small number of lags.

The proposed CMA is both a powerful diagnostic tool for addressing theoretical questions and an efficient approach for practical applications. Assuming system and data errors to be uncorrelated from each other and from the oceanic state, theoretical questions are addressed within the context of least squares (equation 3.19). For a particular GCM and set of measurements, the Green's function matrix, $\mathcal{G}$ in equation (3.19), establishes which GCM and data error components are resolvable. Component $\mathbf{Q}_k$ of system error covariance matrix $\mathbf{Q}$ is resolvable provided $\mathbf{H}\mathbf{A}^s\mathbf{Q}_k\mathbf{A}^{s\prime}\mathbf{H}^{\mathrm{T}} \neq \mathbf{0}$ for several $s \geq 1$ (Section

3.2.1). When all modeled $\mathbf{Q}_k$ in equation (3.4) are resolved, the data error covariance matrix $\mathbf{R}$ is fully resolvable. At least $N$ independent measurements and two covariance matrices (from the set $\mathbf{Y}$, $\mathbf{D}_1$, $\mathbf{D}_2$...) are required to fully resolve an $N \times N$ matrix $\mathbf{Q}$ (Appendix 3.2.2).

A major obstacle to obtaining statistically significant results is the large uncertainty of sample covariance matrices, $O(2\sigma^4/p)$ where $\sigma^2$ is the variance and $p$ is the number of degrees of freedom (Section 3.3.1). The sample uncertainty is represented by $\mathbf{R}_\varepsilon$ in equation (3.20) and standard least squares tools can be used to evaluate the statistical significance of the error estimates (Sections 3.3 and 3.4). The covariance of the covariance matrix can be computed using equations for the theoretical uncertainty of the second-order moments (Section 3.3.1 and Appendix F). In general, the number of error covariance parameters, $\alpha_k$ in equation (3.4), which can be determined with some degree of confidence is two to three orders of magnitude smaller than the total number of independent data. The goal is to find a small dimensional error model which can be made consistent with the data.

t with the data.

122

# Chapter 4

# Experimental Results with Covariance Matching Approach

In this chapter we use the CMA developed in Chapter 3 to estimate the error statistics of a linearized GCM. We use a linearized GCM of the North Pacific, where more than a year of high quality acoustic data are available in addition to the altimetric data. We start by addressing the following question: "For a linear model with four vertical modes can we estimate the mean variance of model error for each mode based on the two kinds of available measurements: altimetric measurements of the sea surface height and acoustic tomography measurements of sound speed converted into temperature anomalies?" Firstly, we perform a series of twin experiments, and show that the CMA can in principle provide reliable estimates of the mean variance of the model errors with the acoustic data but not the altimetric data. This is contrasted with the results of Chapter 2 where we showed that other adaptive methods failed with both data sets. We then use the T/P altimetric data to estimate spatial structure of the model errors. Using the ATOC acoustic data we estimate vertical structure of the model and measurement errors. In addition to this, we show that the CMA can be used to estimate other statistics of the model and measurement errors, namely trends and annual cycles.

The major part of this chapter is presented in the article by Menemenlis and Chechel-

nitsky (1999).

## 4.1 Circulation and Measurement Models

The circulation and measurement models, described below, were used in Chapter 2 (Section 2.1) and are common to both the twin and the real experiments. The GCM is that of Marshall et al. (1997a, 1997b) integrated in a global configuration with realistic topography and driven by surface wind and buoyancy fields obtained from twice-daily National Centers for Environmental Prediction (NCEP) meteorological analyses. Horizontal grid spacing is 1° and there are 20 vertical levels.

A linear, time-independent model for GCM errors in the North Pacific is constructed by systematically perturbing the GCM with large scale temperature anomalies (Menemenlis and Wunsch, 1997 ). The linear model is defined in a region bounded by 5°–60° N and 132°–252° E (Figure 2.4). It operates on a reduced state vector that has 8°-sampling in the horizontal, 4 vertical temperature Empirical Orthogonal Functions (EOFs, see Figure 2.2), and a time step of 1 month. In this representation, sea surface pressure errors in the GCM caused by barotropic or salinity effects, or by scales not resolved by the reduced state vector, become part of the measurement error $\mathbf{r}(t)$, and are described by covariance matrix $\mathbf{R}$. The state vector dimension is reduced from $5 \times 10^6$ in the GCM to 512 in the linear model. Away from coastal regions, this reduced-state linear model describes the large-scale temperature perturbation response of the GCM with considerable skill for periods up to two years. Similar types of state reduction and linearization are commonly used for propagating the error covariance matrix in data assimilation studies (Fukumori and Malanotte-Rizzoli, 1995 ; Cane et al., 1996 ).

The acoustic tomography data from ATOC are first inverted to produce equivalent range-averaged oceanographic temperature perturbations along each section, the ATOC Consortium, 1998 . Data/GCM discrepancy is then projected onto the four vertical EOFs and the monthly sampling of the reduced state vector described above. Therefore, the

124

measurement matrix for acoustic tomography data consists of a range-average for each vertical EOF and for each section. Acoustic data from five sections (Figure 2.4) are used for a total of 20 data points (projections onto the four vertical EOFs for each section), once per month.

The measurement matrix $\mathbf{H}$ appropriate for altimetry consists of a weighted sum of the four vertical EOFs at each horizontal location of the reduced state grid. The weights are chosen to represent the relative contribution of each EOF to sea surface dynamic height at each location. In this representation, sea surface pressure errors in the GCM caused by barotropic or salinity effects, or by scales not resolved by the reduced state vector, become part of the measurement error $\mathbf{r}(t)$, and are described by covariance matrix $\mathbf{R}$.

Before applying the estimation algorithm described in Section 3.1 to real data, we test the algorithm in a series of twin experiments using simulated data with known statistical properties. But first we present an overview of the CMA.

## 4.2   Overview of the Covariance Matching Approach

In this section we present a brief overview of the covariance matching approach (CMA). The full description is given in Chapter 3. We start by parameterizing the model and measurement error covariances:

$$\mathbf{Q} = \sum_{k=1}^{K} \alpha_k \mathbf{Q}_k, \mathbf{R} = \sum_{k=1}^{L} \alpha_{K+k} \mathbf{R}_k. \tag{4.1}$$

Next, for each element matrix $\mathbf{Q}_k$ we obtain its corresponding reduced state covariance matrix $\mathbf{P}_k$ by solving the Lyapunov equation

$$\mathbf{P}_k = \mathbf{A}\mathbf{P}_k\mathbf{A}^\mathrm{T} + \mathbf{Q}_k. \tag{4.2}$$

125

We then solve a simultaneous set of linear equations for the coefficients $\alpha_k$:

$$
\begin{bmatrix} \mathbf{Y}(\vdots) \\ \mathbf{D}_1(\vdots) \\ \vdots \\ \mathbf{D}_S(\vdots) \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{\mathbf{Y},1}(\vdots) & \ldots & \mathbf{G}_{\mathbf{Y},K+L}(\vdots) \\ \mathbf{G}_{\mathbf{D}_1,1}(\vdots) & \ldots & \mathbf{G}_{\mathbf{D}_1,K+L}(\vdots) \\ & \vdots & \\ \mathbf{G}_{\mathbf{D}_S,1}(\vdots) & \ldots & \mathbf{G}_{\mathbf{D}_S,K+L}(\vdots) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{K+L} \end{bmatrix}. \tag{4.3}
$$

where we use the covariance of the model-data residual $\mathbf{y}(t)$ and the covariances of the temporal lag $s$ residuals:

$$
\mathbf{Y} \equiv \operatorname{cov} \mathbf{y}, \quad \mathbf{D}_s \equiv \operatorname{cov}[\mathbf{y}(t+s) - \mathbf{y}(t)], \tag{4.4}
$$

and the Green's functions defined as:

$$
\begin{aligned}
\mathbf{G}_{\mathbf{Y},k} &= \mathbf{H}\mathbf{P}_k\mathbf{H}^{\mathrm{T}}, \qquad \mathbf{G}_{\mathbf{Y},K+k} = \mathbf{R}_k, \\
\mathbf{G}_{\mathbf{D}_s,k} &= \mathbf{H}(\mathbf{A}^s - \mathbf{I})\mathbf{P}(\mathbf{A}^s - \mathbf{I})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} + \sum_{i=1}^{s}\mathbf{H}\mathbf{A}^{s-i}\mathbf{Q}\mathbf{A}^{(s-i)\prime}\mathbf{H}^{\mathrm{T}}, \quad \mathbf{G}_{\mathbf{D}_s,K+k} = 2\mathbf{R}_k.
\end{aligned} \tag{4.5}
$$

Acknowledging the fact there are errors in the sample estimates of the covariance matrices on the left hand side of equation (4.3) and that the parameterizations (4.1) may be incomplete, we append an error to the equations (4.3),

$$
\mathbf{d} = \mathcal{G}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}. \tag{4.6}
$$

The parameter vector $\boldsymbol{\alpha}$ in equation (4.1) is determined by minimizing the weighted least-squares cost function,

$$
J(\boldsymbol{\alpha}) = \boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{R}_{\varepsilon}^{-1} \boldsymbol{\varepsilon} + (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{\mathrm{o}})^{\mathrm{T}} \mathbf{R}_{\alpha}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{\mathrm{o}}), \tag{4.7}
$$

where $\boldsymbol{\alpha}_{\mathrm{o}}$, $\mathbf{R}_{\alpha}$, and $\mathbf{R}_{\varepsilon}$ represent prior knowledge for $\langle\boldsymbol{\alpha}\rangle$, $\operatorname{cov}\boldsymbol{\alpha}$, and $\operatorname{cov}\boldsymbol{\varepsilon}$, respectively. For a discussion on how to estimate the prior $\mathbf{R}_{\varepsilon}$ see section (3.3.1). This completes the description of the basic CMA algorithm.

126

## 4.3  Twin Experiments

### 4.3.1  Generation of simulated data.

We parameterize $\mathbf{Q}$ in equation (3.4) as a diagonal matrix with four parameters, $\alpha_1 \ldots \alpha_4$, each representing the system error variance associated with each of the four vertical EOFs, that is, we assume that the system error is horizontally homogeneous and white. The measurement error covariance, $\mathbf{R}$ in equation (3.4), is also modeled as a diagonal matrix with two parameters, $\alpha_5$ and $\alpha_6$, corresponding to the measurement error variance associated with acoustic tomography and altimeter data, respectively. We do not claim that this simple model is correct or unique; our objective is limited to testing whether this particular model can be made consistent with the available data.

The test data are generated using the reduced state linear model and the acoustic and altimetric measurement models described above, and by driving equations (3.1) and (3.2) with white system and measurement noise characterized by parameters $\alpha_1 \ldots \alpha_6$ (Figure 2.17).

### 4.3.2  Tests with Pseudo-Acoustic Data

The first set of twin experiments is carried out with noise-free, $\mathbf{R} = \mathbf{0}$, simulated acoustic tomography data. It is both impractical, because of computational cost, and unnecessary, because of information overlap, to match all available lag-difference data covariance matrices as in (3.10). An appropriate subset of data covariance matrices must be selected by trial and error and by reference to the guidelines of Section 3.3, that is, a preference for sample covariance matrices with small matrix norms and hence smaller relative uncertainties. The sample uncertainties of $\tilde{\mathbf{Y}}$, $\tilde{\mathbf{D}}_1$, and $\tilde{\mathbf{D}}_2$ are displayed on Fig. 4.1 as a function of number of years of simulated data. Note that $\tilde{\mathbf{D}}_1$ and $\tilde{\mathbf{D}}_2$ have smaller relative uncertainties than $\tilde{\mathbf{Y}}$, suggesting that matching $\tilde{\mathbf{D}}_1$ or $\tilde{\mathbf{D}}_2$ will produce better estimates of $\mathbf{Q}$ and $\mathbf{R}$ than matching $\tilde{\mathbf{Y}}$.

Figure 4.2 displays estimates of parameters $\alpha_1 \ldots \alpha_4$, based on matching $\tilde{\mathbf{D}}_1$, as a
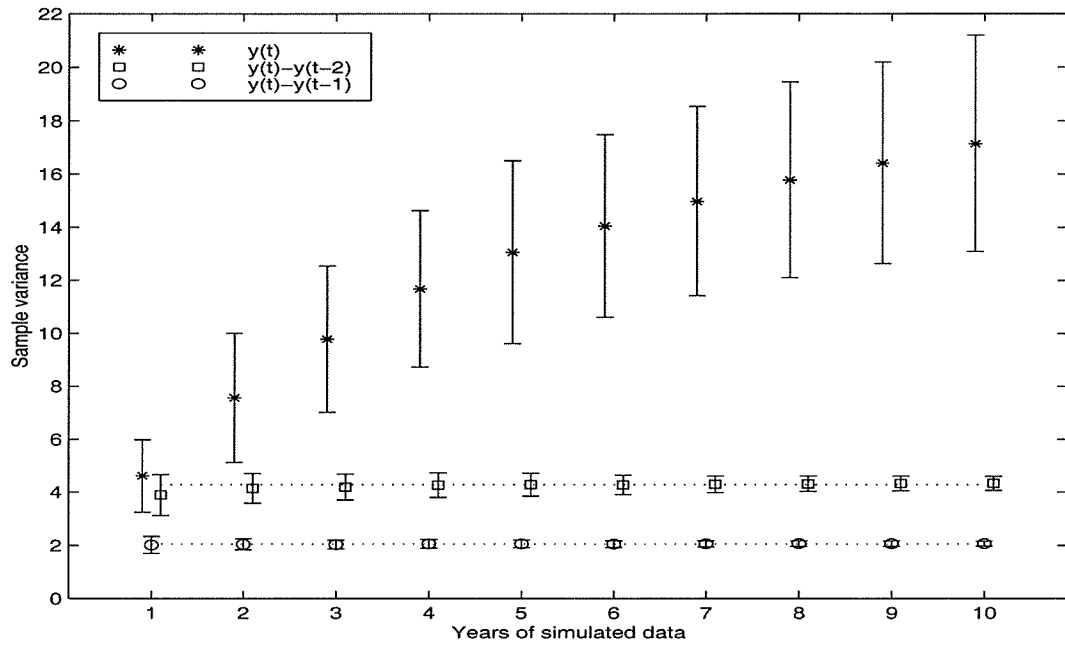
Figure 4.1: Mean diagonal values of sample covariance matrices $\tilde{\mathbf{Y}}$, $\tilde{\mathbf{D}}_1$, and $\tilde{\mathbf{D}}_2$, as a function of years of simulated data. Error bars represent the associated standard uncertainty. Dotted lines are the steady state values. (The steady state value associated with $\tilde{\mathbf{Y}}$, 30, is not shown. Because the leading eigenvalue of the reduced-state dynamical model corresponds to an $e$-folding time scale of 19 years, a few hundred years of data are needed for sample covariance $\tilde{\mathbf{Y}}$ to reach a steady value). Lag-1 estimates have the smallest relative uncertainty, suggesting that matching the lag-1 data covariance matrix will provide the most accurate estimates of system and measurement error.
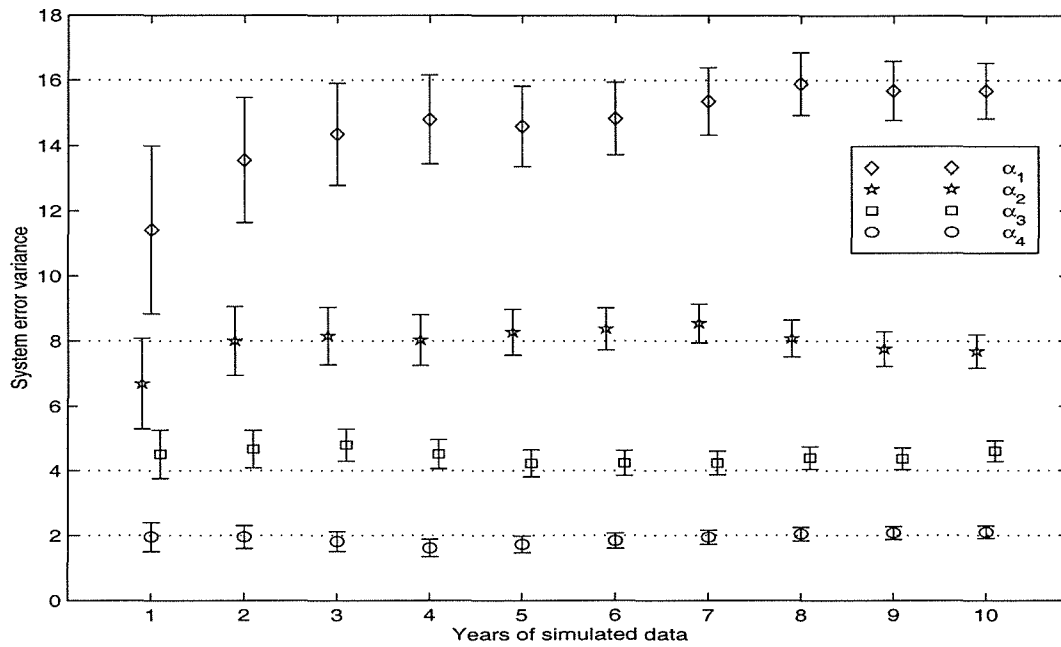
Figure 4.2: Estimates of system error variance based on the lag-1 difference sample covariance, $\tilde{\mathbf{D}}_1$, for simulated acoustic tomography data. Dotted lines indicate the values of $\alpha_1 \ldots \alpha_4$ used to generate the test data. The error bars represent the standard error of the estimates. The figure demonstrates the increasing accuracy of the algorithm with increasing number of measurements.

| estimate | parameters of covariance matrix $\mathbf{Q}$ | | | |
| --- | --- | --- | --- | --- |
| description | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ |
| truth | 16 | 8 | 4 | 2 |
| from $\tilde{\mathbf{Y}}$ | $10.8 \pm 5.7$ | $9.2 \pm 3.2$ | $5.5 \pm 1.6$ | $2.1 \pm 0.7$ |
| from $\tilde{\mathbf{D}}_1$ | $12.6 \pm 2.7$ | $8.0 \pm 1.4$ | $5.3 \pm 0.8$ | $1.9 \pm 0.4$ |
| from $\mathrm{diag}(\tilde{\mathbf{D}}_1)$ | $11.8 \pm 3.4$ | $8.8 \pm 1.8$ | $5.2 \pm 0.9$ | $1.7 \pm 0.5$ |
| from $\tilde{\mathbf{D}}_2$ | $12.5 \pm 3.5$ | $5.2 \pm 1.8$ | $4.2 \pm 0.9$ | $1.6 \pm 0.5$ |

Table 4.1: Estimates of system error covariance matrix $\mathbf{Q}$ based on 14 months of simulated acoustic tomography data. The measurements are assumed perfect, that is, $\mathbf{R} = \mathbf{0}$. The best estimates are obtained by matching the lag-1 difference covariance matrix, $\tilde{\mathbf{D}}_1$, with an average standard error of 18% as compared to 38% for $\tilde{\mathbf{Y}}$.

function of number of years of simulated data. Error bars are obtained as in Section 3.4. Contrary to the empirical algorithm of Myers and Tapley (1976), which failed to converge in this particular example, see Section 2.8, the present algorithm provides useful estimates of system error even with two years of data.

The results of a series of tests based on 14 months of simulated data are summarized in Table 4.1 (at the time of this study 14 months of ATOC data were available). Each particular estimate is not expected to match the true variance of $\mathbf{Q}$ exactly, but over a large number of realizations the estimates are unbiased and their standard deviation matches the standard uncertainty reported on Table 4.1. For the dynamical and measurement models used here, the lag-1 difference sample covariance matrix, $\tilde{\mathbf{D}}_1$, provides the most accurate estimates, with mean standard uncertainty of 18% as compared to 38% for $\tilde{\mathbf{Y}}$. Matching only the diagonal elements of $\tilde{\mathbf{D}}_1$ leads to a standard uncertainty of 23% similar to that obtained by using the full lag-2 difference covariance matrix, $\tilde{\mathbf{D}}_2$.

Next we report on results from a series of experiments with noisy measurements, $\mathbf{R} = \mathbf{I}$ (Table 4.2). Measurement error degrades the estimates of $\mathbf{Q}$ considerably: the standard error for estimates obtained using $\tilde{\mathbf{D}}_1$ is 52%. The uncertainty can be reduced by using several lag-$s$ difference covariance matrices simultaneously: using $\tilde{\mathbf{D}}_1$ and $\tilde{\mathbf{D}}_2$ simultaneously reduces the estimation uncertainty to 38%.

| estimate description | parameters of covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ | | | | |
|---|---|---|---|---|---|
| | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ |
| truth | 16 | 8 | 4 | 2 | 1 |
| from $\tilde{\mathbf{Y}}$ | $10.4 \pm 6.2$ | $10.5 \pm 3.6$ | $5.8 \pm 2.1$ | $4.0 \pm 1.3$ | $0.6 \pm 0.3$ |
| from $\tilde{\mathbf{D}}_1$ | $10.9 \pm 4.5$ | $12.7 \pm 3.1$ | $0.4 \pm 2.4$ | $-.2 \pm 1.6$ | $1.3 \pm 0.2$ |
| from $\tilde{\mathbf{D}}_2$ | $8.8 \pm 4.7$ | $11.3 \pm 2.9$ | $4.3 \pm 2.1$ | $2.9 \pm 1.4$ | $1.0 \pm 0.2$ |
| from $\tilde{\mathbf{D}}_1$ and $\tilde{\mathbf{D}}_2$ | $8.9 \pm 3.8$ | $9.5 \pm 2.4$ | $6.0 \pm 1.7$ | $2.0 \pm 1.1$ | $1.0 \pm 0.1$ |

Table 4.2: Estimates of system and measurement error variance based on 14 months of simulated acoustic tomography data with $\mathbf{R} = \mathbf{I}$. The addition of measurement error increases the uncertainty of the estimates as compared to those of Table 4.1. Nevertheless, usable estimates of $\mathbf{Q}$, with a standard error of 38%, are possible by simultaneously matching the lag-1 and lag-2 difference covariance matrices.

In summary, the estimation uncertainty decreases with increasing years of available data and with increasing ratio $|\mathbf{Q}|/|\mathbf{R}|$. The simulation results indicate that 14 months of acoustic data are sufficient to produce usable estimates of $\mathbf{Q}$ and $\mathbf{R}$, provided the circulation and measurement models of 4.1 are valid and provided $|\mathbf{Q}| \gg |\mathbf{R}|$.

### 4.3.3  Tests with pseudo-altimeter data

A third set of twin experiments is conducted using simulated altimeter data. In theory, it is possible to separate baroclinic modes in the altimeter data by making use of their different temporal evolutions at the sea surface (Holland and Malanotte-Rizzoli, 1989 ). The results presented below, however, suggest that even with perfect measurements, $\mathbf{R} = \mathbf{0}$, and with perfect knowledge of the dynamical and measurement models, $\mathbf{A}$ and $\mathbf{H}$, altimeter data on their own are ill-suited to the estimation of the vertical GCM error statistics. Figure 4.3 is an attempt to estimate the system error using up to ten years of perfect altimeter data. At the conclusion of the tenth year, the standard uncertainty of the estimates remains too large for the estimates to be of any practical interest.

At the writing of this manuscript, 48 months of high quality TOPEX/POSEIDON altimeter data were available. We therefore performed a further series of tests using 48
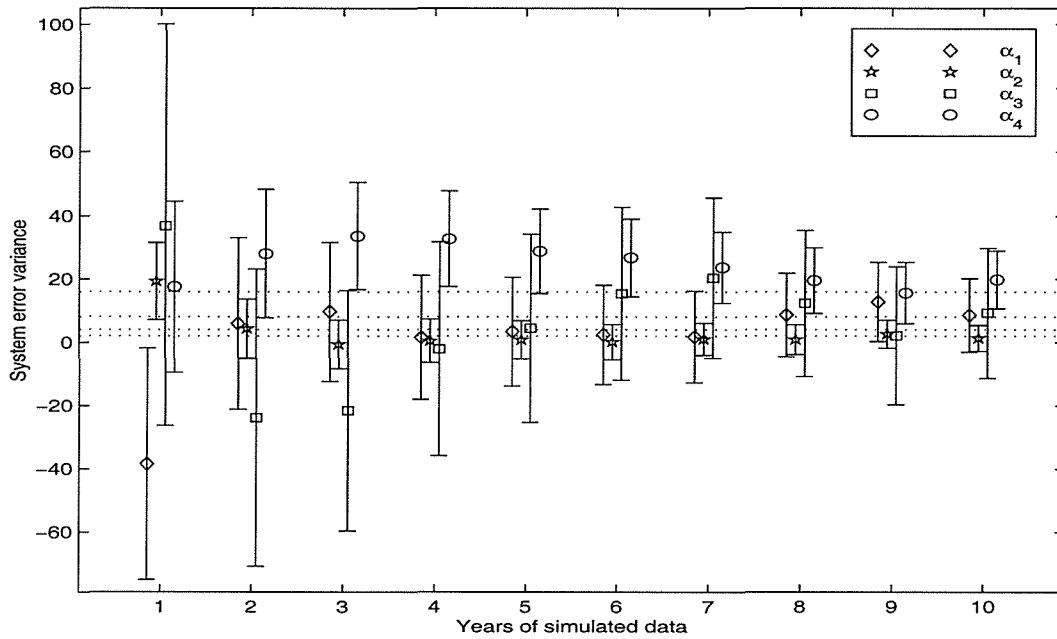
Figure 4.3: Estimates of system error variance based on the diagonal elements of $\tilde{\mathbf{D}}_1$ for simulated altimeter data. Dotted lines indicate variances used to generate the data. Error bars represent the standard uncertainty of the estimates and they can be compared to those of Fig. 4.2 which was created using simulated acoustic data. The large error bars associated with the altimetric estimates suggest that altimeter data are ill-suited to the estimation of the vertical GCM error structure.

| estimate | parameters of covariance matrix $\mathbf{Q}$ | | | |
| --- | --- | --- | --- | --- |
| description | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ |
| truth | 16 | 8 | 4 | 2 |
| from diag($\tilde{\mathbf{Y}}$) | $30 \pm 14$ | $15 \pm 8$ | $-1 \pm 20$ | $-15 \pm 9$ |
| from diag($\tilde{\mathbf{D}}_1$) | $-1 \pm 16$ | $2 \pm 7$ | $-2 \pm 30$ | $32 \pm 14$ |
| from diag($\tilde{\mathbf{D}}_2$) | $30 \pm 16$ | $7 \pm 7$ | $-41 \pm 28$ | $10 \pm 12$ |
| from diag($\tilde{\mathbf{D}}_3$) | $35 \pm 14$ | $18 \pm 8$ | $-45 \pm 28$ | $-10 \pm 11$ |
| from diag($\tilde{\mathbf{D}}_4$) | $36 \pm 13$ | $22 \pm 8$ | $-36 \pm 29$ | $-19 \pm 11$ |
| from diag($\tilde{\mathbf{D}}_5$) | $38 \pm 13$ | $27 \pm 9$ | $-35 \pm 30$ | $-26 \pm 10$ |
| from all the above | $8 \pm 6$ | $10 \pm 3$ | $15 \pm 9$ | $3 \pm 4$ |

Table 4.3: Estimates of system error covariance matrix $\mathbf{Q}$ based on 48 months of perfect, $\mathbf{R} = \mathbf{0}$, simulated altimeter data. The last row of numbers are estimates obtained using the diagonal elements from all six data covariance matrices, $\tilde{\mathbf{Y}}, \tilde{\mathbf{D}}_1 \ldots \tilde{\mathbf{D}}_5$, simultaneously.

months of simulated altimeter data (see Table 4.3). Because of the large dimensions of the sample covariance matrices, only their diagonal elements have been matched. The first six rows of Table 4.3 correspond to estimates from matching $\mathbf{Y}$, and $\mathbf{D}_1$ through $\mathbf{D}_5$. The last row summarizes results from matching all six data covariance matrices simultaneously. The standard errors for this last case range from 35% to 235%. The situation is worse when measurement errors are included. We conclude that covariance matrices for the vertical GCM error structure cannot, in the present setup, be quantified from TOPEX/POSEIDON data alone.

## 4.4 Experimental Results with Real Data

### 4.4.1 TOPEX/POSEIDON data

The covariance matching approach is next applied to TOPEX/POSEIDON altimeter data and to a particular integration of the Marshall et al. (1997a, 1997b) GCM. Figure 4.4 compares measured sea level anomaly variance to that predicted by the GCM. Both the altimetric data and the GCM have been processed in a way consistent with the
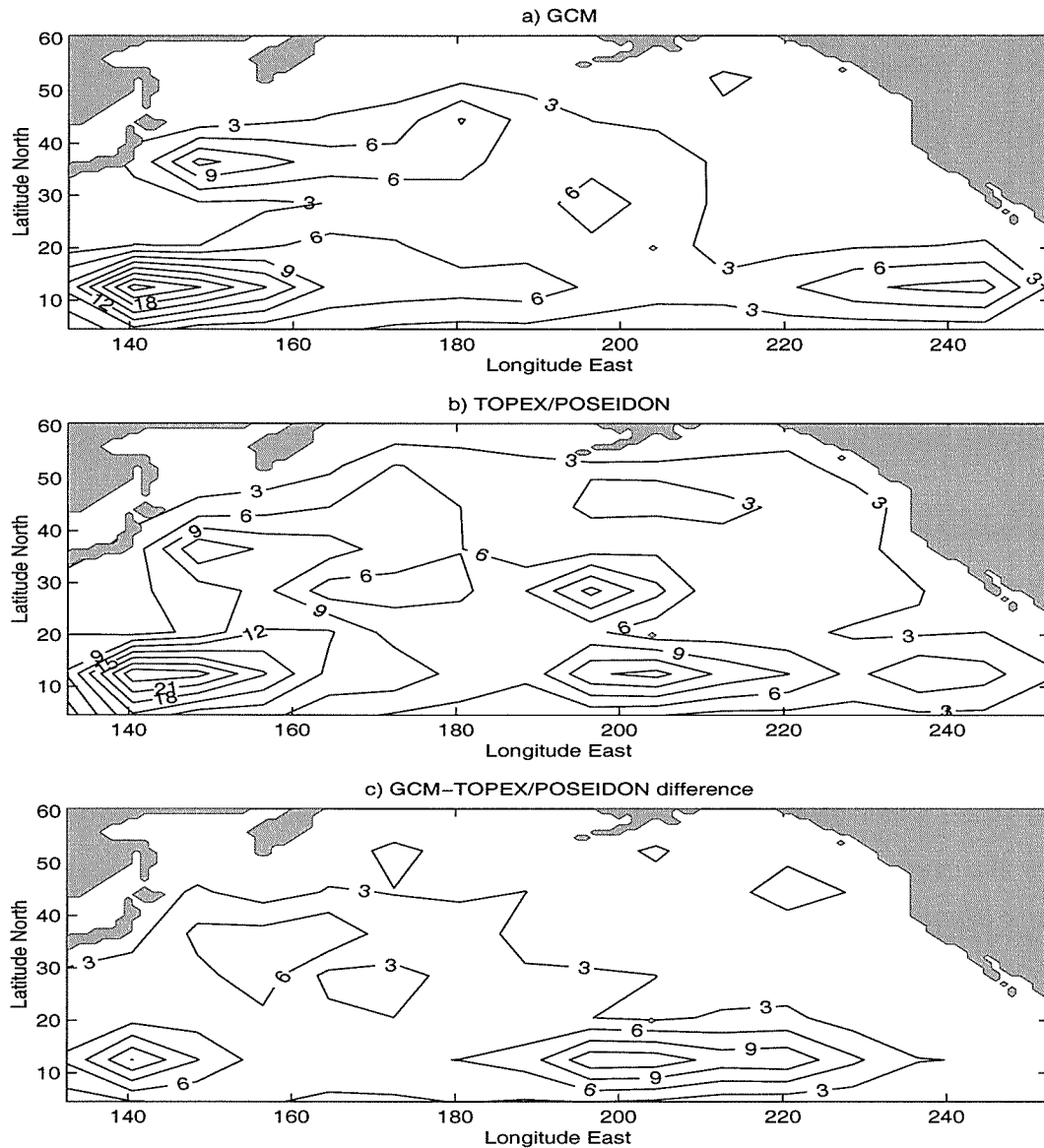
Figure 4.4: North Pacific sea level anomaly variance for a) GCM output, b) TOPEX-POSEIDON data, and c) GCM-TOPEX/POSEIDON residual, during the period 1 October 1992 – 31 May 1997. Annual cycles, trends, periods shorter than two months, and length scales smaller than 16° have been removed. Contour intervals are 3 cm$^2$.

reduced state described in Section 4.1, that is, periods shorter than 2 months and length scales smaller than 16° have been low-pass filtered. In addition, annual cycles and trends have been removed at every location; these will be studied separately. Altimetric data and GCM output exhibit the same general patterns of enhanced variability near the Kuroshio, the Hawaiian Ridge, and in a band north of the Equator. The GCM variability, however, is on average 30% less than that measured by the altimeter, and in some regions, notably in the Eastern Tropical Pacific, the altimetric and GCM time series are uncorrelated. The variance of the GCM-TOPEX/POSEIDON residual (Fig. 4.4c) is 60% that of TOPEX/POSEIDON, indicating that the GCM explains 40% of the observed low frequency/wavenumber variability. Our objective is to determine which fraction of the GCM-TOPEX/POSEIDON residual can be attributed to system error, $\mathbf{HPH}^\mathrm{T}$ in (3.6), and which fraction results from measurement and representation errors, $\mathbf{R}$ in (3.6).

The twin experiments conducted earlier indicate that it is not possible to determine covariance matrices for the vertical GCM error structure from four years of altimetric data. We therefore consider a number of statistical models for covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ which assume equipartition of the variance between the four vertical EOFs. The first model is an attempt to estimate the full spatial structure of the error variance under the assumption that $\mathbf{Q}$ and $\mathbf{R}$ have zero off-diagonal elements. This model results in estimates that have no statistical significance; on average the standard uncertainty of the estimates is fifteen times larger than the estimates themselves for the diagonal elements of $\mathbf{Q}$ and two times larger for the diagonal elements of $\mathbf{R}$.

To obtain statistically significant error estimates, it is necessary to reduce the number of parameters to be estimated. Therefore the second model considered is one of homogeneous and spatially uncorrelated system and measurement error, $\mathbf{Q} = \alpha_1\mathbf{I}$ and $\mathbf{R} = \alpha_2\mathbf{I}$, respectively. Matching this model to covariance matrices $\tilde{\mathbf{Y}}$, $\tilde{\mathbf{D}}_1$, $\tilde{\mathbf{D}}_2$, and $\tilde{\mathbf{D}}_3$ yields $\tilde{\alpha}_1 = 0.25 \pm 0.02$, $\tilde{\alpha}_2 = 1.00 \pm 0.03$. Standard uncertainties are computed using a set of 100 Monte Carlo experiments whereby covariance matching is applied to 100 sets of simulated data generated using normally distributed $\mathbf{q}(t)$ and $\mathbf{r}(t)$ with variance

0.25 and 1.00, respectively. Assuming the statistical model chosen to be the correct one, the standard deviation of the Monte Carlo estimates represents a lower bound for the standard uncertainty of the real estimates. These estimates imply that on average 70% of the GCM-TOPEX/POSEIDON residual variance can be explained by system error, that is, the ratio of the diagonal elements of $\mathbf{HPH}^T$ in (3.6) to those of $\mathbf{Y}$ is approximately 70% (see Fig. 4.5a).

The homogeneous model, however, does not account for some of the regions of enhanced variability in Fig. 4.4c. A third plausible model is $\mathbf{Q} = \alpha_1 \mathbf{Q}_1$ and $\mathbf{R} = \alpha_2 \mathbf{R}_1$, where $\mathbf{Q}_1$ and $\mathbf{R}_1$ are diagonal matrices with a spatially varying structure proportional to that of the GCM-TOPEX/POSEIDON residual variance. Matching this model to the data yields $\tilde{\alpha}_1 = 0.047 \pm 0.006$, $\tilde{\alpha}_2 = 0.28 \pm 0.02$ which implies that 60% of the GCM-TOPEX/POSEIDON residual variance is explained by system error (Fig. 4.5b). To within the sample and estimation uncertainties, the prior variance predicted by this second model is consistent with the data.

A fourth model, that proposed by (Fu et al. 1993), results from assuming that the ocean state is independent from the GCM simulation error, $\langle \mathbf{x}_{\text{ocean}} \mathbf{p}^T \rangle = 0$ (see Appendix D). When this assumption holds,

$$\mathbf{HPH}^T = \frac{1}{2} \left( \text{cov} \left( \boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM,r}} \right) - \text{cov} \, \boldsymbol{\eta}_{\text{ocean}} + \mathbf{H} \, \text{cov} \, \boldsymbol{\zeta}_{\text{GCM,r}} \, \mathbf{H}^T \right), \quad (4.8)$$

$$\mathbf{R} = \text{cov} \left( \boldsymbol{\eta}_{\text{ocean}} - \mathbf{E}\boldsymbol{\zeta}_{\text{GCM}} \right) - \qquad\qquad\qquad\qquad (4.9)$$
$$\frac{1}{2} \left( \text{cov} \left( \boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM,r}} \right) - \text{cov} \, \boldsymbol{\eta}_{\text{ocean}} + \mathbf{H} \, \text{cov} \, \boldsymbol{\zeta}_{\text{GCM,r}} \, \mathbf{H}^T \right),$$

where subscript r denotes the coarse (reduced state) model run. On average, this third model predicts that 15% of the GCM-TOPEX/POSEIDON residual variance is caused by system error (Figure 4.5c). This relatively low value, compared to the earlier estimate of 60%, is consistent with a correlation coefficient of $\rho = -0.55$ between $\mathbf{Hp}(t)$ and $\mathbf{H}\boldsymbol{\zeta}_{\text{GCM}}(t)$ (see Appendix D).
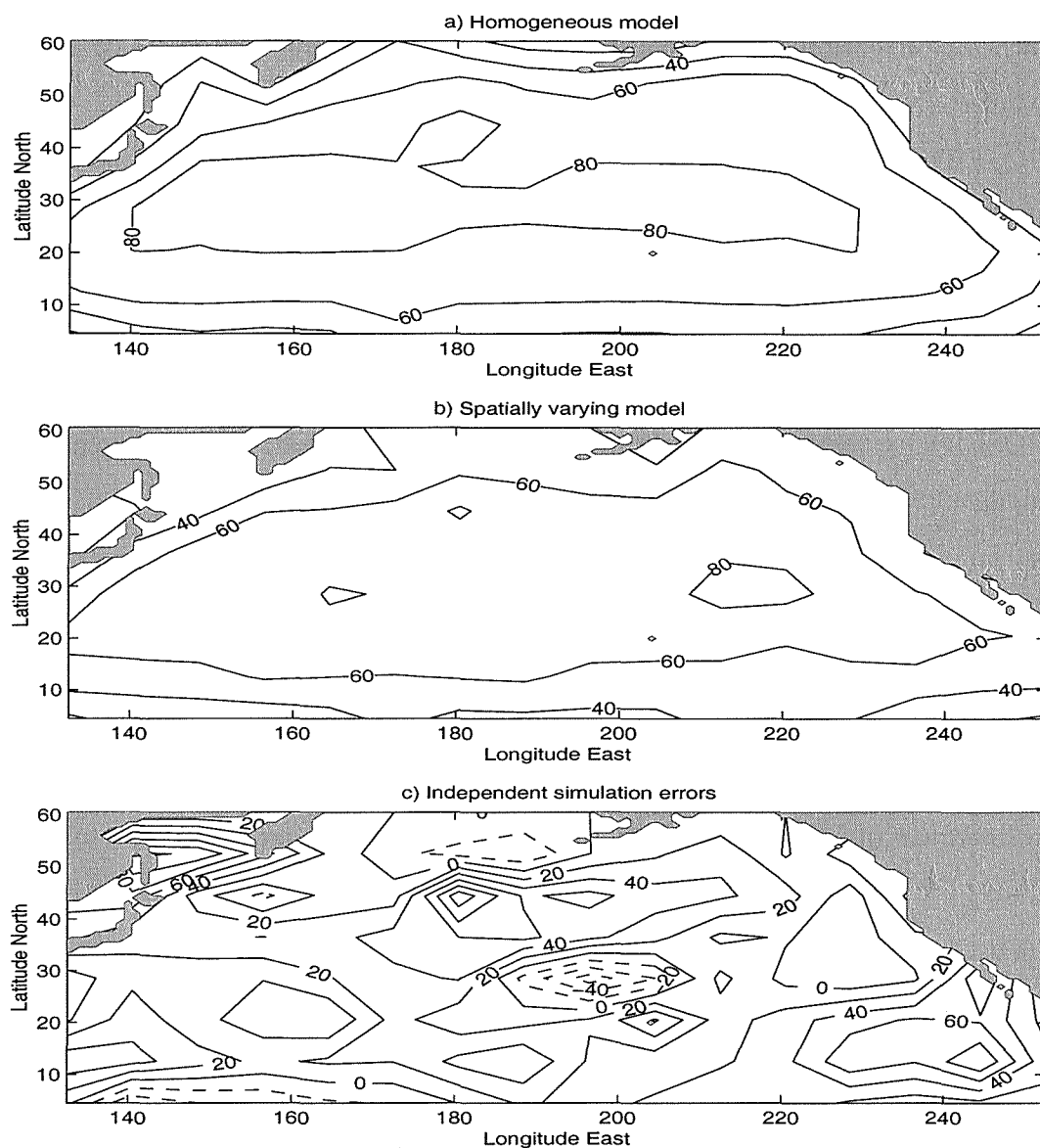
Figure 4.5: Prior estimate for percent variance of GCM-TOPEX/POSEIDON residual which is explained by system error, that is, ratio of diagonal elements of $\mathbf{HPH}^T$ in (3.6) to those of $\mathbf{Y}$. The estimates are obtained using the covariance matching method for a) a homogeneous model for the errors, b) a spatially varying model, and c) the model proposed by Fu eta al. (1993) which assumes that the GCM simulation errors are independent from the ocean state. Contour intervals are 20%. Spurious negative regions in c) (dashed contours) result from the large uncertainty of the sample covariance matrices used in the analysis.

## 4.4.2 ATOC Data

We now turn our attention to the acoustic data. Fig. 4.6 compares the GCM-ATOC residual, converted to an equivalent sea level anomaly, to the range-averaged GCM-TOPEX/POSEIDON residual along each acoustic path, after removing trends and annual cycles. The acoustic data is used to estimate the vertical structure of the errors and to test noise model #2 from above, that is, $\mathbf{Q} = 0.047\mathbf{Q}_1$. We model $\mathbf{Q}$ as a diagonal matrix with four parameters, $\alpha_1 \ldots \alpha_4$, each representing system error variance associated with each of the four vertical EOFs, and with a spatial structure proportional to that of the GCM-TOPEX/POSEIDON residual variance (Fig. 4.4c). Measurement and representation error are modeled as $\mathbf{R} = \alpha_5\mathbf{I}$.

The cost function (3.20) is minimized assuming a priori estimates of $0.047 \pm 0.047$ for $\alpha_1 \ldots \alpha_4$, that is, the estimate obtained using TOPEX/POSEIDON data but allowing for a larger uncertainty in order to test the vertical equipartition hypothesis. The a priori estimate for $\alpha_5$ is taken to be $0.28 \pm 0.28$, that is, the variance of the acoustic data with a corresponding uncertainty. A conservative estimate for the prior sample covariance uncertainty is $\mathbf{R}_\varepsilon = 0.28\mathbf{I}$ (Section 3.3.1). The resulting estimates for system noise variance are $\alpha_1 = 0.15 \pm 0.04$, $\alpha_2 = 0.00 \pm 0.04$, $\alpha_3 = 0.11 \pm 0.04$, and $\alpha_4 = 0.00 \pm 0.04$. These estimates differ from the altimetric estimate of $0.047 \pm 0.006$, indicating that the vertical equipartition hypothesis is not valid.

A solution that is simultaneously consistent with both TOPEX/POSEIDON and ATOC data can also be obtained: $\alpha_1 = 0.04 \pm 0.03$, $\alpha_2 = 0.01 \pm 0.02$, $\alpha_3 = 0.06 \pm 0.03$, and $\alpha_4 = 0.12 \pm 0.02$. This solution differs from that using ATOC data alone in that it predicts less error variance associated with vertical EOF 1 and more with vertical EOF 4, that is, larger model errors above the seasonal thermocline (see Fig. 2.2). The differences are likely caused by different spatial and temporal extents for the ATOC and TOPEX/POSEIDON data and by inaccuracies in the assumed statistical models. All three covariance matching solutions, however, whether from TOPEX/POSEIDON data alone, from the ATOC data, or from their combination predict that about 60% of the
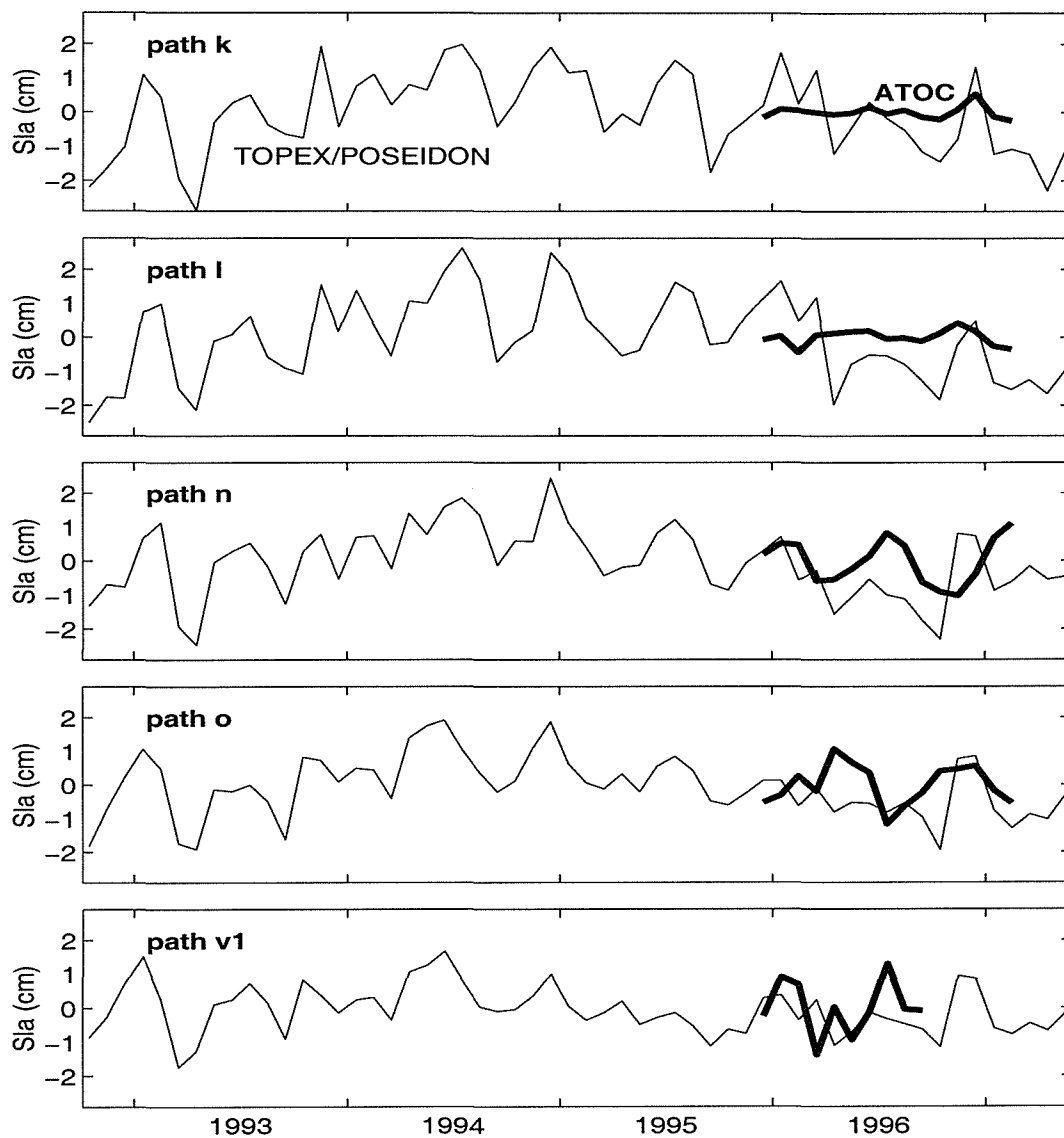
138

Figure 4.6: GCM-ATOC residual, along the five sections shown on Fig. 2.4, converted to an equivalent sea level anomaly for comparison with the TOPEX/POSEIDON data. Annual cycles and trends have been removed.

139
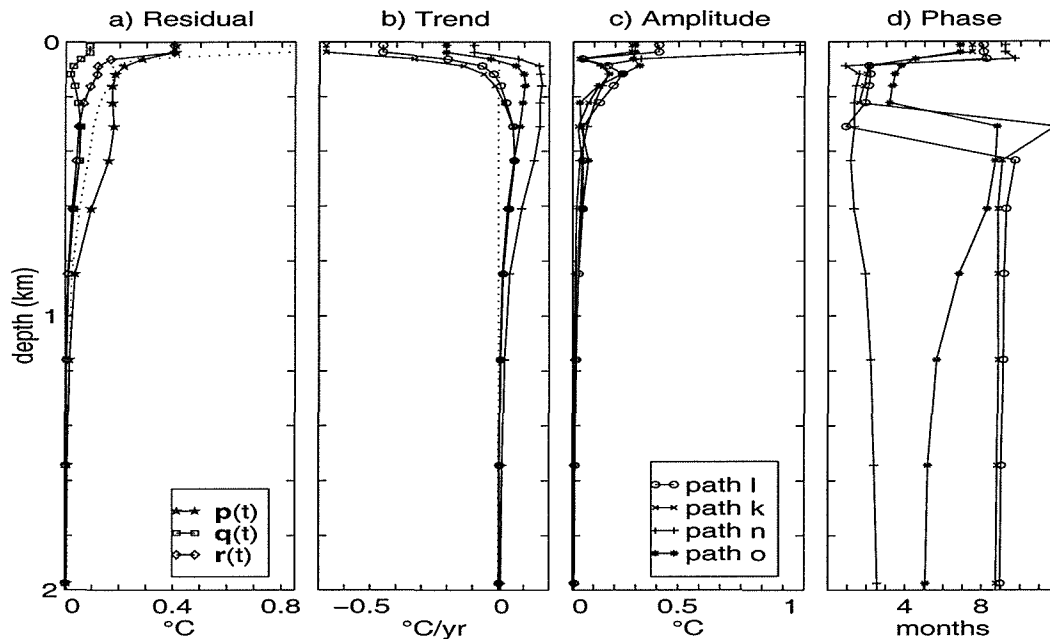
Figure 4.7: Vertical structure of the errors along the ATOC sections: a) standard error in $°C$ excluding trend and annual cycle, b) trend in $°C\ yr^{-1}$, c) annual cycle amplitude in $°C$, and d) annual cycle phase in months. The pentagrams, squares, and diamonds in a) correspond to estimated GCM, system, and measurement standard errors, respectively. The dotted line is the mean standard uncertainty of the acoustic inversions: it can be compared to that estimated using covariance matching (diamonds) and it provides an approximate measure of statistical significance. Trends and annual cycles are displayed for acoustic sections k, l, n, and o of Fig. 2.4. Positive trends correspond to warming of the GCM relative to the acoustic data. Annual cycle phase indicates the month of maximum positive anomaly for the GCM relative to the data.

GCM-TOPEX/POSEIDON residual variance is explained by system error.

Figure 4.7a displays the mean vertical structure of residual errors along the ATOC acoustic sections. The dotted line indicates the mean standard uncertainty of the acoustic inversions (The ATOC Consortium 1998, ) and can be compared to the covariance matching estimate of $\alpha_5 = 0.31 \pm 0.03$ (diamonds). Also displayed are the estimated GCM and system standard errors, $\mathbf{p}(t)$ and $\mathbf{q}(t)$, respectively. The acoustic data has limited depth resolution, being better suited to the measurement of top-to-bottom averages. Nevertheless, the data indicates significant errors in the GCM variability from about 100 m to 1000 m depth, with a maximum of $0.2°C$ at 300 m.

140
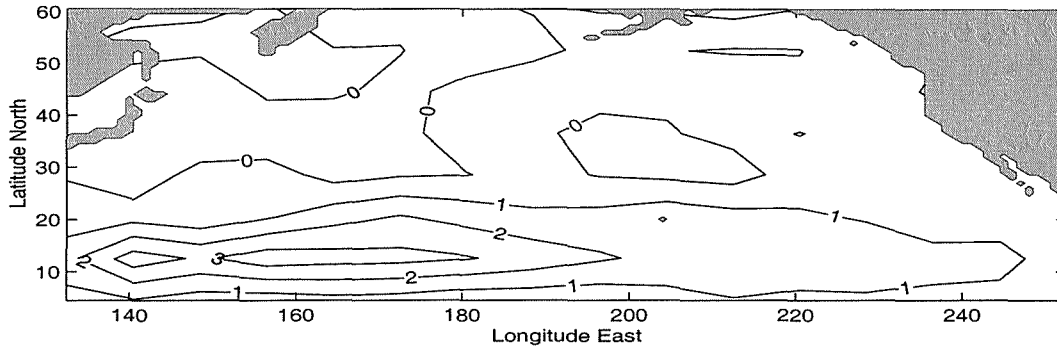
Figure 4.8: Trend in the GCM-TOPEX/POSEIDON residual. Contour intervals are in cm yr$^{-1}$ of sea level anomaly. Positive contours indicate a gradual warming of the GCM relative to TOPEX/POSEIDON.

### 4.4.3 Trend and annual cycle

Trends and annual cycles of the GCM-data residuals, which were excluded from the previous analysis, are discussed next. In the tropical Pacific, the GCM exhibits a warming trend relative to TOPEX/POSEIDON data of up to 3 cm yr$^{-1}$ (Fig. 4.8). The acoustic data indicate that most of the warming occurs between the seasonal and main thermoclines, 50–1000 m depth, with a peak warming of 0.1 to 0.2° C yr$^{-1}$, depending on location (Fig. 4.7b).

For most of the subtropical gyre, both the GCM and TOPEX/POSEIDON exhibit maximum sea level anomaly in September (month 9), but the TOPEX/POSEIDON amplitude is about 2 cm larger than that of the GCM (Fig. 4.9). As a result, the peak GCM-TOPEX/POSEIDON residual occurs in March (month 3), six months out of phase with the GCM or TOPEX/POSEIDON annual cycle. Excluding the surface layer, where resolution is poor, the acoustic data suggest that the annual cycle error is confined to a depth range shallower than 200 m, the phase-locked range in Fig. 4.7d, with a peak of 0.3° C at 120 m depth (Fig. 4.7c).

141

Figure 4.9: Annual cycle peak amplitude for a) GCM output, b) TOPEX/POSEIDON data, and c) GCM-TOPEX/POSEIDON residual. Contour intervals are 2 cm. The corresponding phase is displayed in d), e), and f), respectively, with 2 month contour intervals.

d) GCM phase

e) TOPEX/POSEIDON phase

f) GCM–TOPEX/POSEIDON phase

Figure 4.9: Continued.

143

## 4.5   Summary

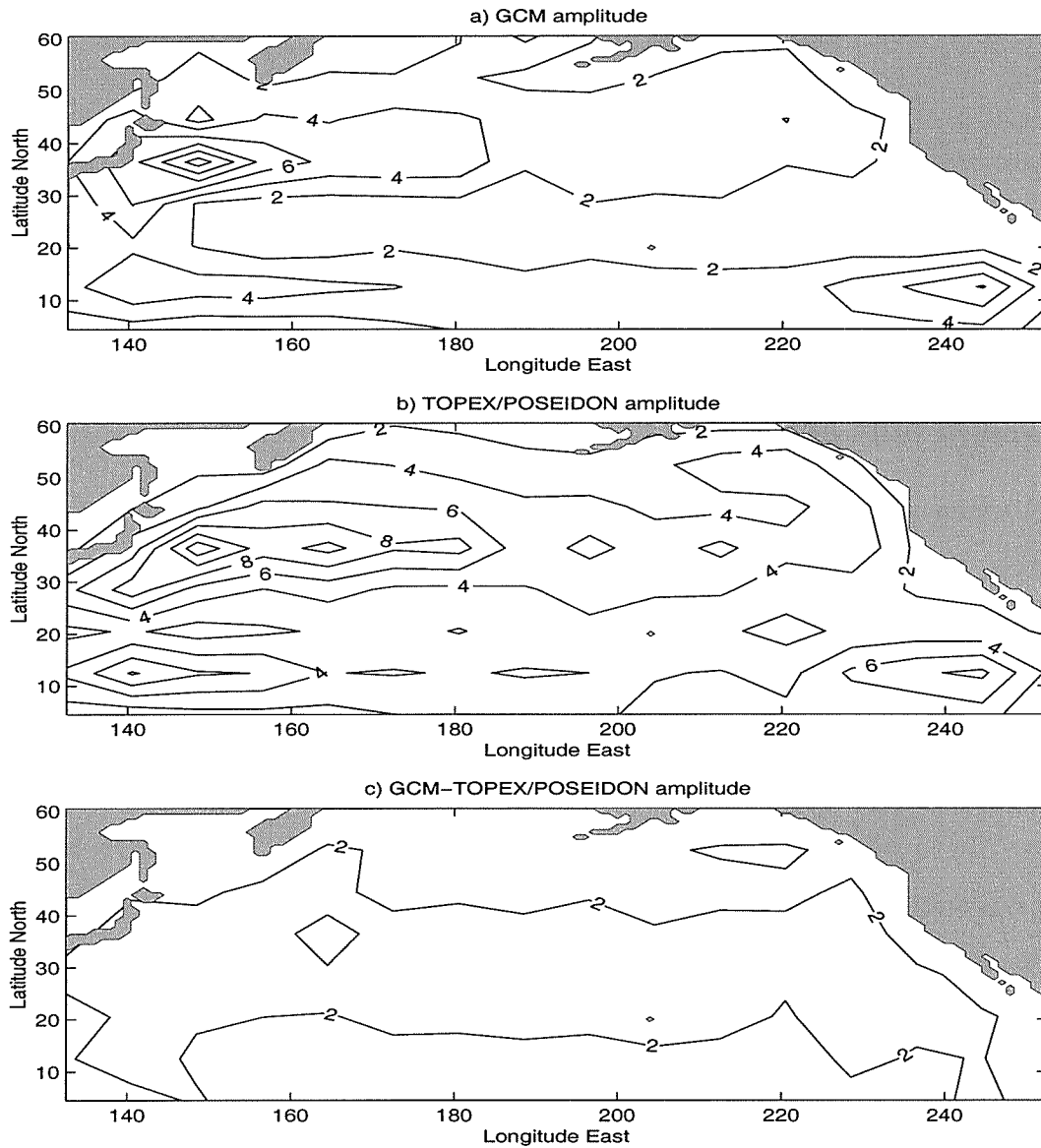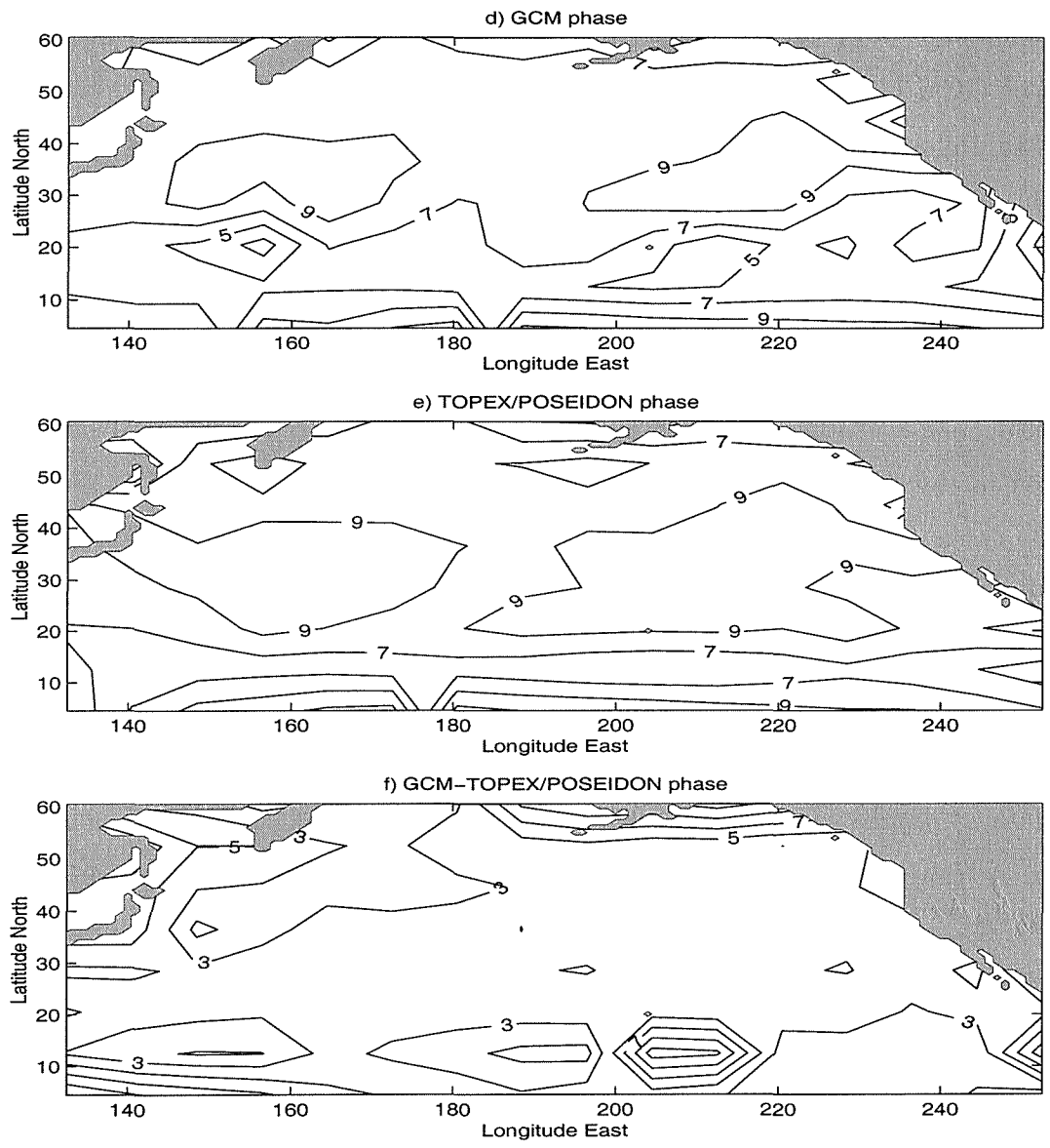In this chapter we illustrate the CMA, developed in Chapter 3, by applying it to a particular integration of the Marshall et al. (1997a, 1997b) GCM, 56 months of TOPEX/POSEIDON sea level anomaly data, and 14 months of acoustic tomography data from the ATOC project. The GCM is forced with observed meteorological conditions at the surface and integrated in a global configuration with $1°$ horizontal grid spacing and 20 vertical levels. A reduced state linear model that describes internal (baroclinic) error dynamics is constructed for the study area ($5°$–$60°$ N, $132°$–$252°$ E).

Twin experiments, using the reduced state model, suggest that altimetric data are ill-suited to the estimation of internal GCM errors, but that such estimates can in theory be obtained using the acoustic data (Figures 4.2 and 4.3). These conclusions must however be qualified in the following way. First, the vertical modes used here are EOFs, not dynamical modes, and second, the tests were conducted using linearized GCM dynamics. We do not exclude the possibility that dynamical modes or fully nonlinear dynamics could enhance the resolution of internal GCM errors from altimetric data.

The GCM exhibits a warming trend relative to TOPEX/POSEIDON data of order 1 cm yr$^{-1}$ (Figure 4.8) corresponding to a peak warming of up to $0.2°$C yr$^{-1}$ in the acoustic data at depths ranging from 50 to 200 m (Figure 4.7b). This trend measures GCM drift. At the annual cycle, GCM and TOPEX/POSEIDON sea level anomaly are in phase, but GCM amplitude is 2 cm smaller (Figure 4.9). The acoustic data suggest that the annual cycle error is confined to the top 200 m of ocean (Figure 4.7c and d). These differences result from errors in the surface boundary conditions and in the dynamics of the GCM.

After removal of trends and annual cycles, the low frequency/wavenumber (periods > 2 months, wavelengths > $16°$) TOPEX/POSEIDON sea level anomaly is order 6 cm$^2$. The GCM explains about 40% of that variance (Figure 4.4). The CMA suggests that 60% of the GCM-TOPEX/POSEIDON residual variance is consistent with the reduced state dynamical model (Figure 4.5b). The remaining residual variance is attributed to

measurement noise and to barotropic and salinity GCM errors which are not represented in the reduced state model. The ATOC array measures significant GCM temperature errors in the 100–1000 m depth range with a maximum of 0.3° at 300 m (Figure 4.7a). The remaining GCM-TOPEX/POSEIDON residual variance is attributed to measurement noise, to barotropic and salinity GCM errors, and to vertical modes of temperature variability which are not represented by the reduced state model.

This chapter demonstrates that it "is" possible to obtain simple statistical models for GCM errors that are consistent with the available data. For practical applications, however, the GCM error covariance estimation problem is in general highly underdetermined, much more so than the state estimation problem. In other words there exist a very large number of statistical models that can be made consistent with the available data. Therefore, methods for obtaining quantitative error estimates, powerful though they may be, cannot replace physical insight. But used in the right context, as a tool for guiding the choice of a small number of model parameters, covariance matching can be a useful addition to the repertory of oceanographers seeking to quantify GCM errors or to carry out data assimilation studies.

# Chapter 5

# Application of the Covariance Matching Approach to a Linearized GFDL GCM

In this chapter we present an application of the covariance matching approach (CMA) to a linearized global ocean model. We use the same model and data that were used in the global data assimilation study of Fukumori et al. (1999) (F99, hereafter). The model is a linearization of the GFDL GCM and has only two vertical modes – the barotropic and first baroclinic modes. The observations are the T/P measurements of the sea surface height anomalies. Only two vertical modes are chosen because they explain most of the sea surface height variability and have very distinct projections onto the sea level height anomaly. A 3 year time series of the GCM–data residuals is sufficiently long for the CMA to provide statistically significant estimates for a small but carefully chosen set of parameters, $\alpha$, which describe the model and measurement error covariances.

Unlike the previous chapter, where we only derived the model and measurement error covariances, in this work we use the estimates with a global data assimilation scheme to obtain estimates of the state of the global ocean. This work supports the conclusions of F99 that it is possible to improve estimates of the ocean state using an approximate data

assimilation scheme (Section 5.2) with a global data set and a global GCM. Furthermore, it strengthens these conclusions by testing them with several parameterizations of the error statistics.

On the other hand, the results of data assimilation demonstrate that in this case adaptively tuned error covariances make very little difference for the estimates of the ocean state. This is explained by the fact that the measurement error, dominated by the representation error, is much larger than the model error. The representation error is due to the coarse resolution of the GCM (and to a smaller extent even coarser resolution of the reduced-state model). The coarse resolution means that the GCM is unable to adequately simulate the mesoscale eddies, and it is the mesoscale eddies dominate the variability of the sea surface height in the T/P altimetric data set. Large measurement error leads to very small Kalman gain, i.e. small weight on the measurements. Accordingly, the measurements are hardly used in the assimilation and the resulting estimate with adaptively tuned error statistics is improved very little (Section 5.4.3).

In addition, we use the CMA approach to study the vertical partitioning of the model error, where by model error we mean the differences on largest spatial scales between the sea surface height estimated by the model and that observed by T/P, for example one possible source of model error could be errors in the wind forcing which drive the model. Difference on small spatial scales are part of the representation error because we do not even attempt to model them with the coarse resolution GCM. The results show that most of the model error is explained by the barotropic mode. This can be understood by noting that the model-data residual has largest variance in the high latitudes where the barotropic mode has a larger projection. Thus, not only the sea level (Fukumori et al. 1998) but also the errors are dominated by the barotropic mode in high latitudes.

We start by describing the model and data. The data assimilation method used in this study is described in Section 5.2. We then provide a short overview of the CMA approach in section 5.3. Derivation of the error covariances for the data assimilation is presented in Section 5.4. The data assimilation based on the parametrization of F99 is presented in

Section 5.4.3. Section 5.5 discusses results with a different error parametrization. Vertical partitioning of the model error is studied in Section 5.6. The conclusions are given in Section 5.7.

## 5.1 Description of the Model and Data

The GCM used in this chapter is based on the Modular Ocean Model developed at the Geophysical Fluid Dynamics Laboratory of the National Oceanic and Atmospheric Administration. The model is a nonlinear primitive equation model extending over the world ocean from 80°S to 80°N with a uniform spatial resolution of 2° longitude and 1° latitude. There are 12 vertical levels which are based on the first 12 baroclinic modes corresponding to the mean temperature-salinity profiles of the global ocean, Levitus (1982). The model has realistic coast lines. For additional details on the model one may consult Pacanowski et al. (1991).

The reduced-state model is described in F99. Below we provide a short summary of its derivation. The recent study of Fukumori et al. (1998) investigated the nature of large-scale sea level variability, and the two main processes which dominate sea-level variability have been identified. In the tropics (latitude < 20°), low-frequency (>100 days) wind-driven baroclinic modes are dominant with the first baroclinic mode contributing most of the variance. In high latitudes (latitude > 40°), high-frequency wind-driven barotropic motions dominate the sea surface variability. In mid-latitude (between 20° and 40°), near-surface steric effects due to thermal heating/cooling dominate sea level variance. However, they have relatively little dynamic effect, see Gill and Niiler (1973). Therefore, dynamics of global large-scale sea level change can be approximated in terms of the barotropic and the first baroclinic mode and it is assumed that the steric effects have negligible contribution to the error.

The reduced-state is computed relative to the time-mean state of the model simulation. The equations of motion are non-separable due to nonlinearity and variable

148

bathymetry. This results in the coupling of dynamic vertical modes. However, locally dynamic vertical modes form a complete set of orthogonal basis functions, which are then used for expansion in terms of horizontal coefficients. The model's prognostic variables, namely zonal and meridional velocities, $(u, v)$, temperature, $T$, and salinity, $S$, are approximated by

$$(u, v, T, S) \approx (\bar{u}, \bar{v}, \bar{T}, \bar{S}) + \left( -\frac{1}{H}\frac{\partial \psi}{\partial y} + a_u \hat{p}, \frac{1}{H}\frac{\partial \psi}{\partial x} + a_v \hat{p}, a_h \hat{h}\frac{\partial \bar{T}}{\partial z}, a_h \hat{h}\frac{\partial \bar{S}}{\partial z} \right) \qquad (5.1)$$

where overbar denotes the reference state; $\hat{p}$ and $\hat{h}$ are the structures of the first baroclinic modes of velocity and displacement respectively; $\psi$ is the transport stream function; $x$ and $y$ are meridional and zonal coordinates; and $a_u, a_v, a_h$ are first baroclinic mode amplitudes of zonal velocity, meridional velocity, and vertical displacement, respectively. The stream function defines the contribution from the barotropic mode. Vertical displacements result in changes in the baroclinic structure of the temperature and salinity fields because of the adiabatic nature of wind-driven sea level change. Derivation of the vertical modes for the linear equations of ocean dynamics is given in Gill (1982).

A coarse grid was defined with a uniform 10° by 5° zonal and meridional spacing, respectively, which was sufficient to resolve the dominant scales. Thus, the total number of DOF was reduced to 691 (number of points on the coarse horizontal grid) times 4 (number of vertical variables, $a_u, a_v, a_h, \psi$).

### 5.1.1 Observations

Observations used in this study are given by T/P global sea level anomalies from January 1, 1993 to December 31, 1995 (see section 2.2 for a more detailed discussion of the T/P data). The original data set was averaged along-track in 2.5° latitudinal bins. Linear trends were computed and removed from the T/P data. This is done because trends are not adequately reproduced by the GCM, see Fukumori et al. (1998).

## 5.2 The Data Assimilation Scheme

The assimilation in this study uses an approximate Kalman filter based on a reduced-state and a time-asymptotic approximation (Fukumori and Malanotte-Rizzoli, 1995)[1]. The reduced-state approximation estimates the error covariances with a smaller number of degrees of freedom, which are chosen to resolve only the dominant spatial structures of the error (Section 5.1). The time-asymptotic approximation uses the time-asymptotic limit of the Ricatti equation instead of estimating the time-dependent error covariances (Appendix C). In practice the suboptimal character of the two approximations is of little importance given the uncertainties in the estimates of the error covariances, as discussed in detail in Chapter 2.

In order to compute the time-asymptotic limit for the error covariances, the data distribution during one particular subcycle of T/P is taken as a representative data distribution. Furthermore, it is assumed that all data is available instantaneously every 3 days. That is, to derive the limit we assume that the observations are given on the same grid every 3 days. The time-asymptotic limits for the forecast and update error covariances, $\mathbf{\Pi}_s(-)$ and $\mathbf{\Pi}_s$, are computed using the doubling algorithm (Appendix C) and stored. However, for the actual assimilation the *time-varying* distribution of data is used while using the *time-invariant* error covariances. Namely, the Kalman gain is derived at every time step of the assimilation using a modified equation (2.29):

$$\mathbf{K}(t) = \mathbf{\Pi}_s(-)\mathbf{H}(t)^{\mathrm{T}} \left( \mathbf{H}(t)\mathbf{\Pi}_s(-)\mathbf{H}(t)^{\mathrm{T}} + \mathbf{R}(t) \right)^{-1}. \tag{5.2}$$

## 5.3 Overview of the Covariance Matching Approach

In this section we present a brief overview of the covariance matching approach (CMA). The full description is given in Chapter 3. We start by parameterizing the model and

---

[1] In contrast to F99, the smoother is not used, because, as shown in the next Section, the KF estimates are very similar to the ones derived in F99. It can be shown that when the KF estimates are similar, the smoother estimates are very close as well. Accordingly, we restrict our attention to the KF forecast and update fields.

150

measurement error covariances:

$$\mathbf{Q} = \sum_{k=1}^{K} \alpha_k \mathbf{Q}_k, \text{ and } \mathbf{R} = \sum_{k=1}^{L} \alpha_{K+k} \mathbf{R}_k. \tag{5.3}$$

Next, for each element matrix, $\mathbf{Q}_k$, we obtain its corresponding reduced-state covariance matrix, $\mathbf{P}_k$, by solving the Lyapunov equation,

$$\mathbf{P}_k = \mathbf{A}\mathbf{P}_k\mathbf{A}^{\mathrm{T}} + \mathbf{Q}_k. \tag{5.4}$$

We then solve a simultaneous set of linear equations for the coefficients $\alpha_k$:

$$\begin{bmatrix} \mathbf{Y}(:) \\ \mathbf{D}_1(:) \\ \vdots \\ \mathbf{D}_S(:) \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{\mathbf{Y},1}(:) & \cdots & \mathbf{G}_{\mathbf{Y},K+L}(:) \\ \mathbf{G}_{\mathbf{D}_1,1}(:) & \cdots & \mathbf{G}_{\mathbf{D}_1,K+L}(:) \\ & \vdots & \\ \mathbf{G}_{\mathbf{D}_s,1}(:) & \cdots & \mathbf{G}_{\mathbf{D}_s,K+L}(:) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{K+L} \end{bmatrix}, \tag{5.5}$$

where we use the covariance of the model-data residual, $\mathbf{y}(t)$, and the covariances of the temporal lag– $s$ residuals:

$$\mathbf{Y} \equiv \mathrm{cov}\,\mathbf{y}, \text{ and } \mathbf{D}_s \equiv \mathrm{cov}[\mathbf{y}(t+s) - \mathbf{y}(t)], \tag{5.6}$$

and the Green's functions defined as:

$$\mathbf{G}_{\mathbf{Y},k} = \mathbf{H}\mathbf{P}_k\mathbf{H}^{\mathrm{T}}, \qquad \mathbf{G}_{\mathbf{Y},K+k} = \mathbf{R}_k, \tag{5.7}$$

$$\mathbf{G}_{\mathbf{D}_s,k} = \mathbf{H}(\mathbf{A}^s - \mathbf{I})\mathbf{P}(\mathbf{A}^s - \mathbf{I})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} + \sum_{i=1}^{s} \mathbf{H}\mathbf{A}^{s-i}\mathbf{Q}\mathbf{A}^{(s-i)\prime}\mathbf{H}^{\mathrm{T}}, \text{ and } \mathbf{G}_{\mathbf{D}_s,K+k} = 2\mathbf{R}_k.$$

Acknowledging the fact there are errors in the sample estimates of the covariance matrices on the left hand side of equation (5.5), and that the parameterizations (5.3) may be incomplete, we append an error to the set of equations (5.5),

$$\mathbf{d} = \mathcal{G}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}. \tag{5.8}$$

The parameter vector $\boldsymbol{\alpha}$ in equation (5.3) is determined by minimizing the weighted least-squares cost function,

$$J(\boldsymbol{\alpha}) = \boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{R}_{\varepsilon}^{-1}\boldsymbol{\varepsilon} + (\boldsymbol{\alpha} - \boldsymbol{\alpha}_\circ)^{\mathrm{T}}\mathbf{R}_{\alpha}^{-1}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_\circ), \tag{5.9}$$

151

where $\boldsymbol{\alpha}_o$, $\mathbf{R}_\alpha$, and $\mathbf{R}_\varepsilon$ represent prior knowledge of $\langle\boldsymbol{\alpha}\rangle$, cov $\boldsymbol{\alpha}$, and cov $\boldsymbol{\varepsilon}$, respectively. For a discussion on how to estimate the prior $\mathbf{R}_\varepsilon$ see Section 3.3.1. This completes the description of the basic CMA algorithm.

## 5.4   The Error Covariances of F99

The goal of this section is to establish whether we can find consistent estimates of the error covariance matrices using the model–data differences and to understand the impact of the adaptively estimated error statistics on the data assimilation experiment using real data.

We start by describing the global assimilation presented in F99 which was done with a Kalman filter and smoother using asymptotic and reduced-state approximations with three years, (1993–1995), of T/P along track sea level anomaly data (Section 5.2). The error covariances for the approximate Kalman filter were derived by the method of Fu et al. (1993, F93, hereafter). To check the quality of the resulting estimates of the state, F99 compared statistics of the assimilated estimates against their respective theoretical expectations. In addition, comparison of the estimated fields against independent in situ observations, e.g. the sea level as obtained from tide gauges, the currents as measured by moorings, and the pressure as obtained from bottom pressure gauges, was performed. Overall, the assimilated estimates were shown to be in better agreement with in situ measurements than the simulated estimates (i.e. a GCM run with atmospheric forcing over the 3 year period without any data assimilation). On the other hand, inaccuracies in estimates were identified in some regions. This points to violations of some of the assumptions used in assimilating the observations. Here we investigate whether some of these violations were due to the misspecification of the model and measurement error covariances, and whether we can further improve the data assimilation estimates.

a) [−21.63 ,731.4 ]

b) [9    ,786.3 ]
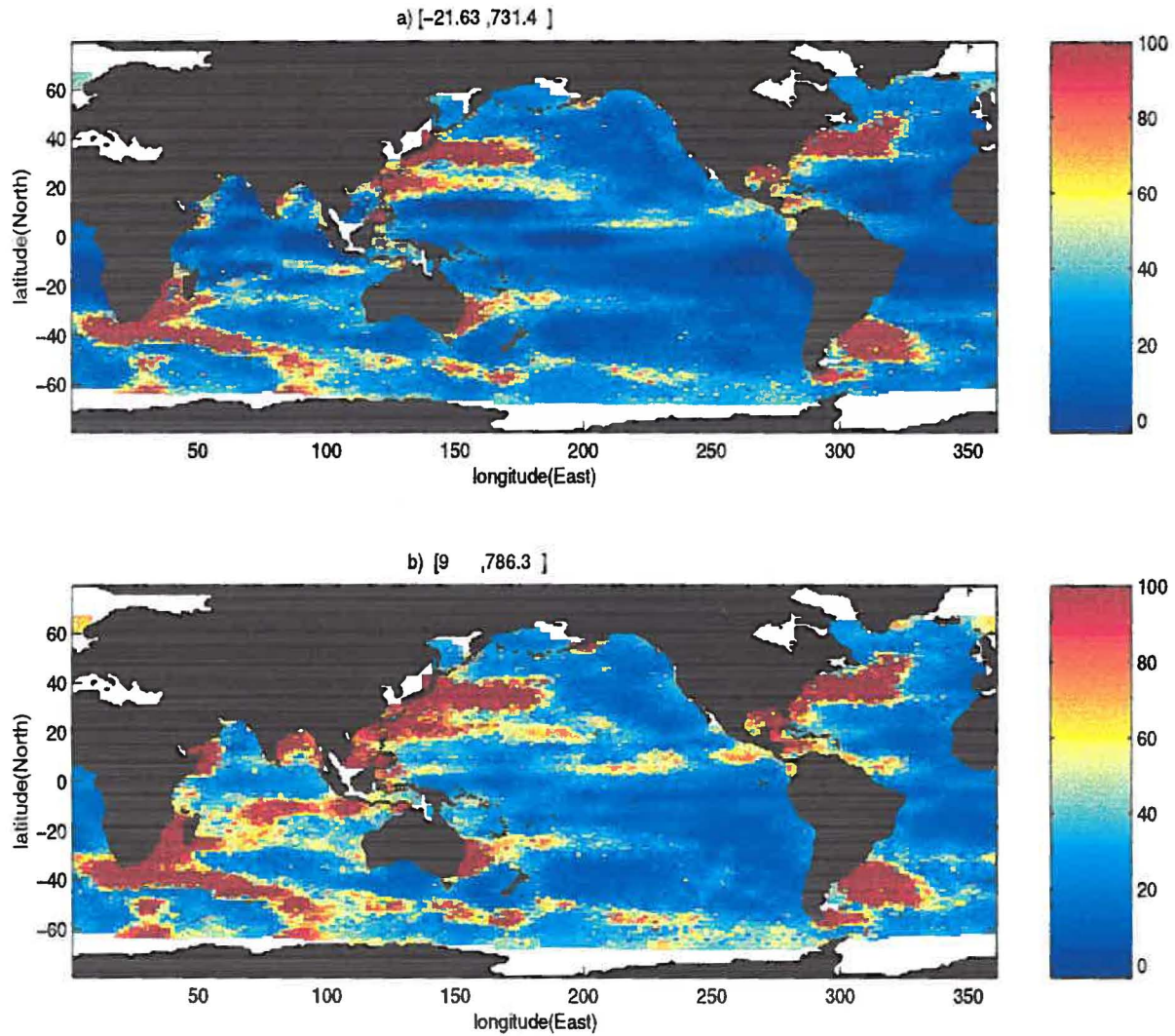
Figure 5.1: Spatial distribution of the variance of the data error in cm$^2$ for a) **R** obtained from equation (5.11) and b) **R** used in F99. The correlation between the two fields is 0.90. The minimum and maximum values are given in the square brackets.
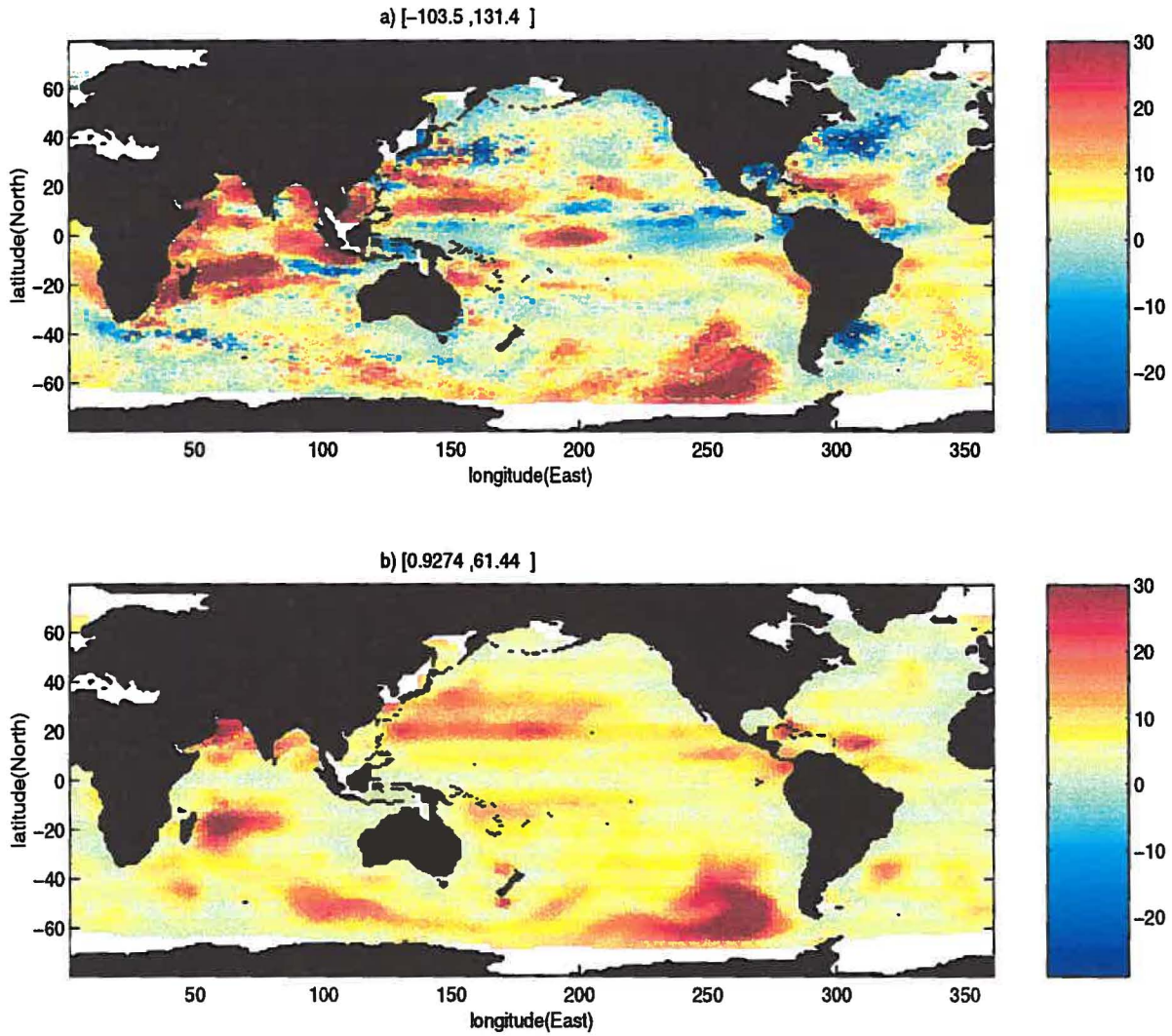
Figure 5.2: Variance of the projection of the model error onto the sea level in cm$^2$ (diag $\mathbf{HPH}^\mathrm{T}$) a) from equation (5.10) and b) for the $\mathbf{Q}$ actually used in F99. The correlation between the two fields is 0.56. The minimum and maximum values are given in the square brackets.
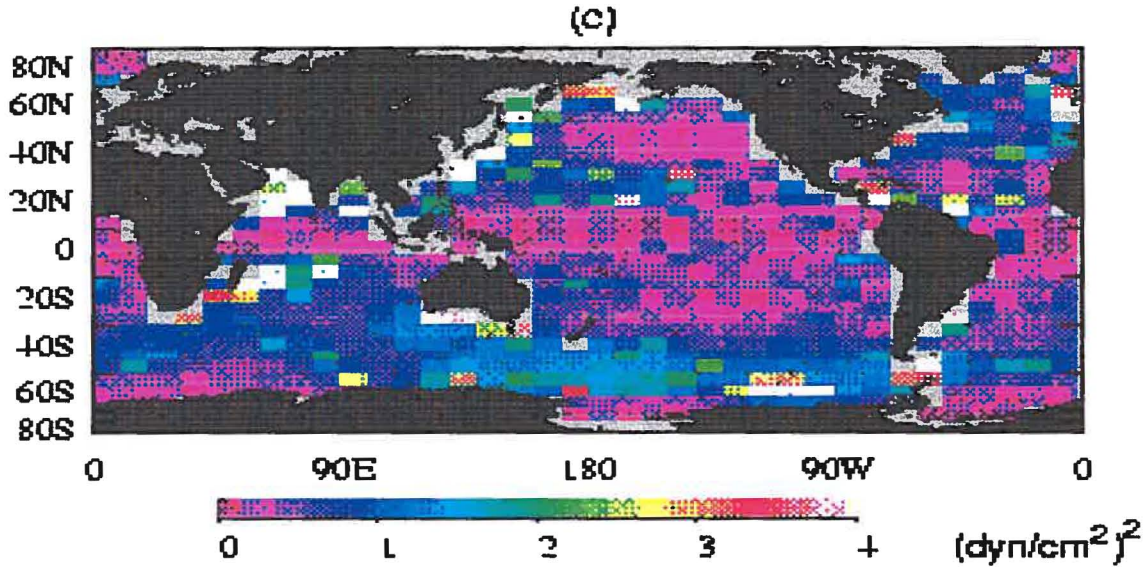
Figure 5.3: Variance of the NCEP wind stress(from F99).

ance at each spatial location was unknown. The values of the local variance were chosen to give the best local fit with the deduced $\mathbf{HPH}^T$ by using the Lyapunov equation (5.4). While it should not be literally interpreted as model errors being dominated by the wind errors, it is an assumption that the structure of the covariance matrix $\mathbf{Q}$ is similar to that of the wind stress.

Figure 5.2b shows the diagonal of $\mathbf{HPH}^T$ for the model error covariance $\mathbf{Q}_{F99}$ used in F99. The total correlation between the two fields (Figures 5.2a and 5.2b) is 0.56. This is surprisingly high considering the assumption that the model error is proportional to the covariance of the NCEP winds, i.e. to forcing fields themselves rather than to the errors in the forcing fields (as is often done, e.g. Miller and Cane 1989, Cane et al. 1996). However, it is substantially less than one because of the large negative values in the sample estimate (equation 5.10). The variance of the model error projected onto the sea level has strong peaks in the South Pacific west of South America and near Madagascar, and is very small in the equatorial region, as expected since its spatial structure roughly corresponds to the spatial distribution of the wind variance (Figure 5.3).

In F99 most of the residual is attributed to the measurement error, which includes

156

both instrument and representation errors (Section 2.2). This is attributed to the absence of mesoscale eddies in the model, and the fact the variance of the measurement errors is largest in the areas of intense eddy activity (western boundary currents, Figure 5.1b). Figure 5.4a shows the fraction of the model–data residual variance explained by the model error. Over most of the domain it is less than 20 per cent, and it implies that the measurement error is much greater than the model error in these regions. The only large areas where the model error accounts for more than a half of the residual variance is in the South Pacific west of the South America, in the south-west Indian ocean and in the west equatorial Atlantic. This is due to the fact that the covariance of the model errors was assumed to be proportional to the covariance of the NCEP winds (Figure 5.3) and these are the regions of the strongest wind activity.

Recapitulating, the method of F93 is constrained to partition the variance of the residual, $\mathbf{y}(t) \equiv \boldsymbol{\eta}_{\text{ocean}}(t) - \mathbf{E}(t)\boldsymbol{\zeta}_{\text{GCM}}(t)$, between the model and measurement error covariances, as can be easily seen by summing up the equations (5.10) and (5.11). This is exactly the first equation used in the CMA (equation 3.6). One of the differences between the two methods is that in the CMA algorithm one has to assume the parameterizations of both $\mathbf{Q}$ and $\mathbf{R}$, while in the method of F93 the measurement error covariance $\mathbf{R}$ is given directly (equation 5.11), by making the additional assumptions discussed Appendix D. The model error covariance $\mathbf{Q}$ is estimated in exactly the same way as in the CMA, i.e. through the Lyapunov equation (5.4). However, the projection of the model error covariance onto the sea level is given by the difference of the covariance of the model–data residuals and the covariance of the measurement errors, equation (5.10), instead of the matching the covariance of the model–data residuals of the CMA, equation (5.5). Overall, F99 estimate more than 17,000 parameters, that is, 1362 coefficients for the model error covariance (NCEP wind correlation matrix with variances of both zonal and meridional winds adjusted at every location to fit the sample estimate of $\mathbf{HPH}^{\text{T}}$) and the full diagonal of the measurement error covariance. Estimating such a large number of parameters from relatively short time series is known to be unstable, see the discus-
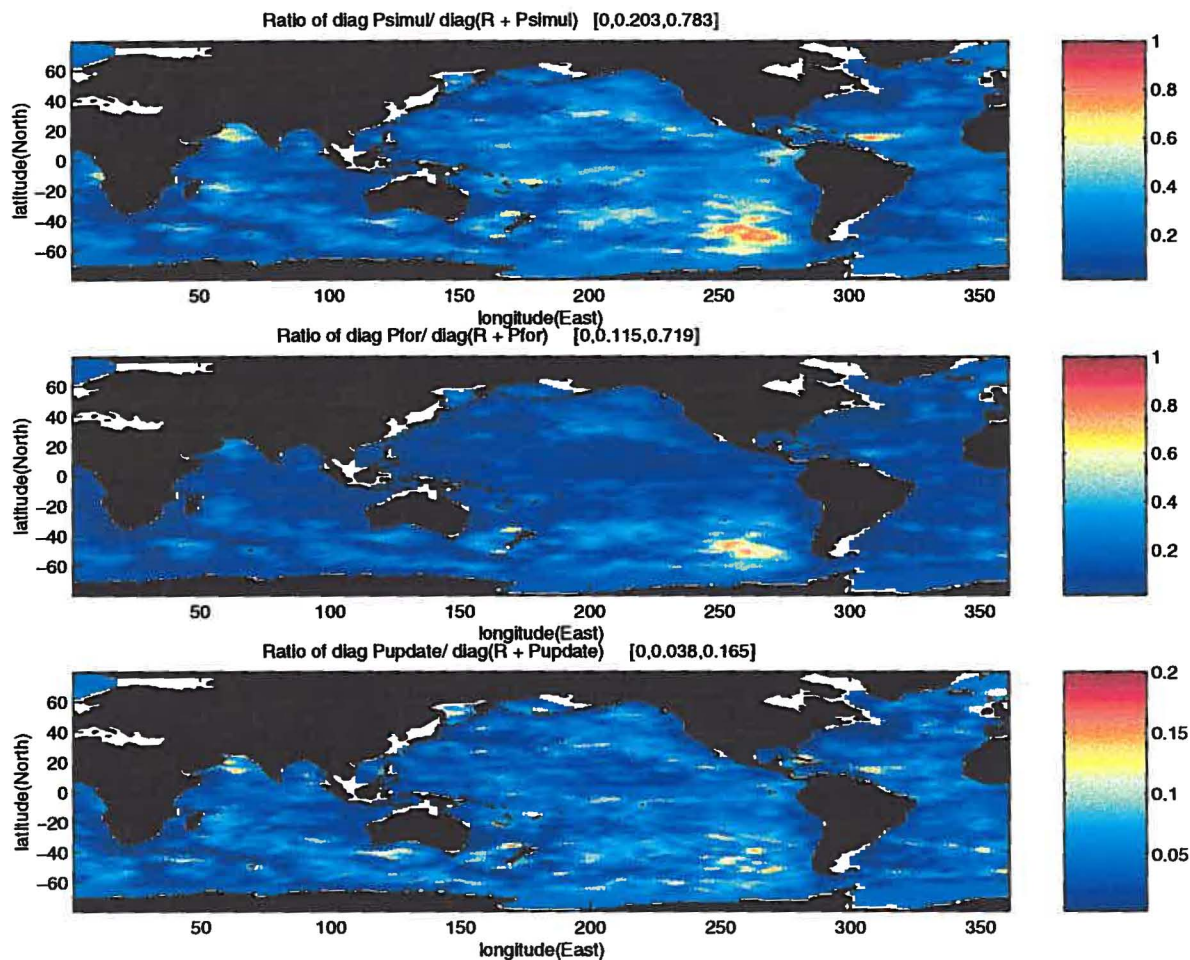
Figure 5.4: Relative ratio of the model and measurement errors for the covariance model used by F99 for : a) the model error (simulation error) $\text{diag}\mathbf{HPH}^T/(\text{diag}\mathbf{HPH}^T+\text{diag}\mathbf{R})$; b) forecast error forecast $\text{diag }\mathbf{H\Pi}_s(-)\mathbf{H}^T/(\text{diag }\mathbf{H\Pi}_s(-)\mathbf{H}^T + \text{diag }\mathbf{R})$; c) update error $\text{diag }\mathbf{H\Pi}_s(-)\mathbf{H}^T/(\text{diag }\mathbf{H\Pi}_s(-)\mathbf{H}^T + \text{diag }\mathbf{R})$. Values greater than one half indicate that most residual variance is explained by the model errors, and less than one half by the measurement error (which includes representation error). The minimum, mean and maximum values are given in the square brackets.

sion in Dee (1995). Other differences between the two methods include use of time-lag information and more consistent application of the least-squares machinery in the CMA. Estimates of the model and measurement error covariances obtained in F99 are such that most of the variance of the model–data residual is explained by the measurement error.

## 5.4.2 The CMA Test of the Error Covariances of F99

The CMA does not use the assumption of decorrelation between the true state of the ocean and the model errors, which is a possible source of the wrong partitioning between the model and measurement error covariances. Thus, to check the parametrization of F99, we parameterize the error covariances (equation 5.3) as

$$\mathbf{Q} = \alpha_1 \mathbf{Q}_{\text{F99}}, \text{ and } \mathbf{R} = \alpha_2 \mathbf{R}_{\text{F99}}, \tag{5.12}$$

where the two matrices $\mathbf{Q}_{\text{F99}}$ and $\mathbf{R}_{\text{F99}}$ are the model and measurement error covariances used in F99 (described above). The coefficients $\alpha_1$ and $\alpha_2$ should be equal to 1 if the estimates used in F99 are consistent with the CMA.

To obtain estimates of the coefficients $\alpha_1$ and $\alpha_2$ we need to choose the range of the column operator (:) and the subset of the covariances $\mathbf{Y}, \mathbf{D}_1, \ldots, \mathbf{D}_S$ which we use in the CMA. The method of F93 uses only covariances of the data, $\mathbf{Y}$, and does not use covariances of the lag-differences, $\mathbf{D}_s$, as in the CMA. We start by matching only the covariance of the residuals, matrix $\mathbf{Y}$ (Figure 5.5). We obtain the following estimates using the diagonal elements of the residual matrix $\mathbf{Y}$:

$$\alpha_1 = 0.68 \pm 0.12, \text{ and } \alpha_2 = 0.71 \pm 0.03. \tag{5.13}$$

Including off-diagonal elements does not change the estimates in a significant way. Both coefficients are less than one, that is, the estimates for the error covariances used in F99 are overly conservative. It may be surprising, since we are matching a single equation:

$$\text{diag} \left( \mathbf{Y} \right) = \text{diag} \left( \mathbf{HPH}^{\text{T}} + \mathbf{R} \right), \tag{5.14}$$

159

Figure 5.5: Log10 of the variance of the residual between T/P data and the GCM simulation over the 3 years, 1993-1995, i.e. diagonal of **Y**. The highest values are in the regions of intense western boundary currents. The minimum and maximum values are given in the square brackets.

and the matrices $\mathbf{Q}_{F99}$ and $\mathbf{R}_{F99}$ were chosen based on the equation (5.14). However, as explained above, the model error covariance was based on the sample covariance of the NCEP wind (forcing field for the GFDL GCM). In fact, large negative values on the diagonal of $\mathbf{HPH}^T$ (equation 5.10) had to be discarded, since **P** represents a covariance matrix and has to have positive values on the diagonal (Figure 5.2). In other words, the magnitude of $\mathbf{Q}_{F99}$ was chosen to be consistent with the large positive values of $\mathbf{HPH}^T$ disregarding the fact that there was a significant number of spurious negative elements. The CMA, on the other hand, does not use the sample estimate of $\mathbf{HPH}^T$ obtained through equation (5.10), and gives smaller estimates for the parameters $\alpha_1$ and $\alpha_2$ to match the LHS of equation (5.14).

Performing data assimilation with these estimates of the model and measurement error covariances would produce nearly identical estimates of the state, since the parameters $\alpha_1$ and $\alpha_2$ are nearly equal and the Kalman filter (and smoother) estimates of the state do not change when we multiply both **Q** and **R** by a constant (section 2.4). How-
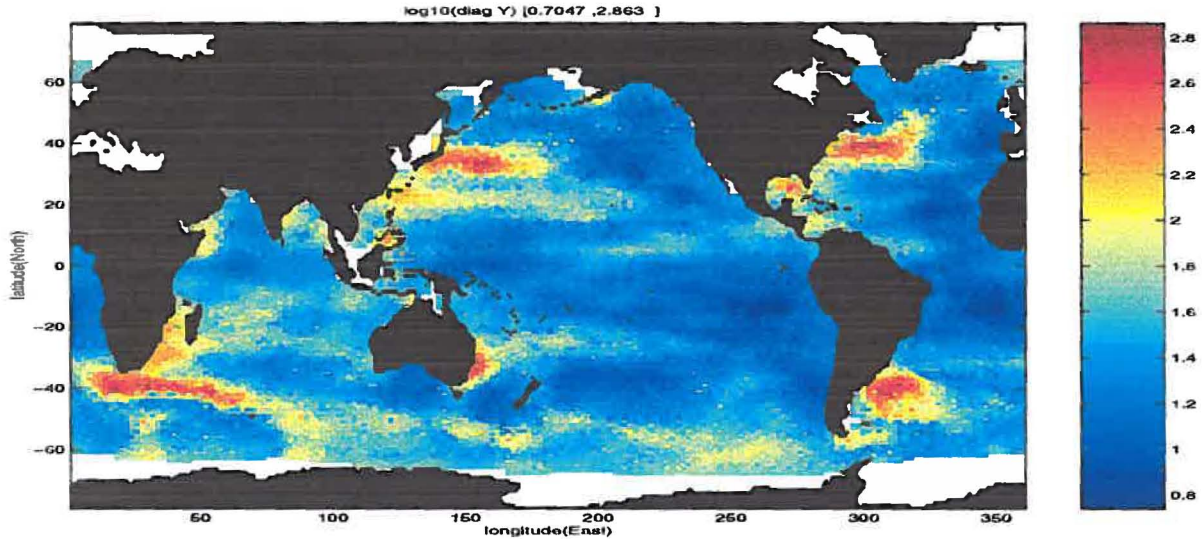
Figure 5.5: Log10 of the variance of the residual between T/P data and the GCM simulation over the 3 years, 1993-1995, i.e. diagonal of $\mathbf{Y}$. The highest values are in the regions of intense western boundary currents. The minimum and maximum values are given in the square brackets.

and the matrices $\mathbf{Q}_{F99}$ and $\mathbf{R}_{F99}$ were chosen based on the equation (5.14). However, as explained above, the model error covariance was based on the sample covariance of the NCEP wind (forcing field for the GFDL GCM). In fact, large negative values on the diagonal of $\mathbf{HPH}^{T}$ (equation 5.10) had to be discarded, since $\mathbf{P}$ represents a covariance matrix and has to have positive values on the diagonal (Figure 5.2). In other words, the magnitude of $\mathbf{Q}_{F99}$ was chosen to be consistent with the large positive values of $\mathbf{HPH}^{T}$ disregarding the fact that there was a significant number of spurious negative elements. The CMA, on the other hand, does not use the sample estimate of $\mathbf{HPH}^{T}$ obtained through equation (5.10), and gives smaller estimates for the parameters $\alpha_1$ and $\alpha_2$ to match the LHS of equation (5.14).

Performing data assimilation with these estimates of the model and measurement error covariances would produce nearly identical estimates of the state, since the parameters $\alpha_1$ and $\alpha_2$ are nearly equal and the Kalman filter (and smoother) estimates of the state do not change when we multiply both $\mathbf{Q}$ and $\mathbf{R}$ by a constant (section 2.4). How-
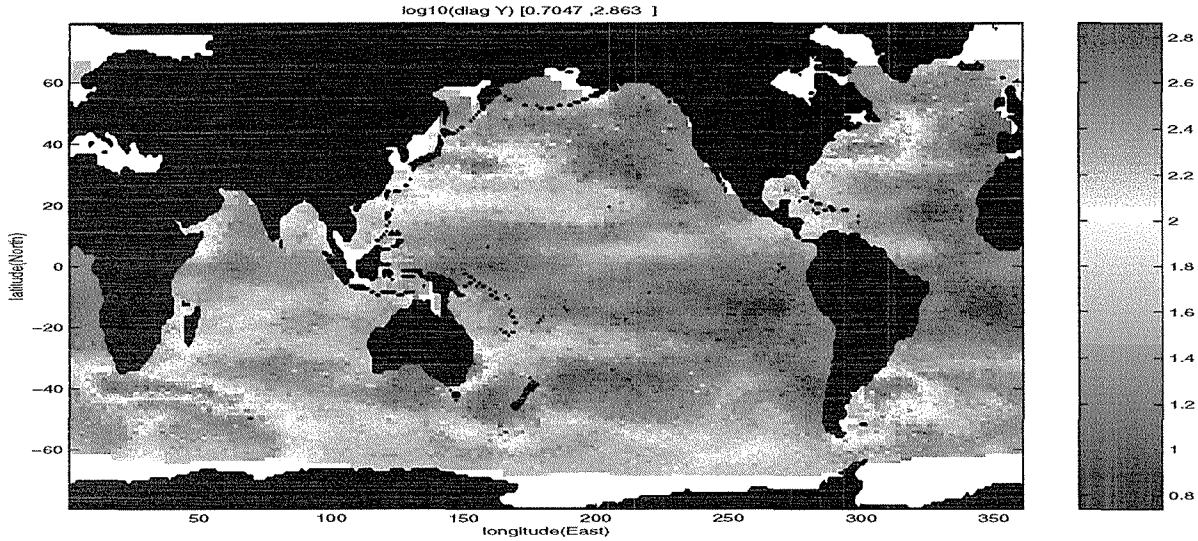
160

ever, the estimates of theoretical uncertainties need to be reduced by roughly 17 percent. This is consistent with the fact that the theoretical estimates of the uncertainty in F99 are slightly larger than the sample ones (Figure 5.6).

The CMA does not use the assumption of independence between the true state of the model and the model errors, and therefore provides an estimate of the cross-covariance between the true state and the model errors (Figure 5.7). Over most of the domain the cross-covariance, and thus the correlation, is very small, but in the areas of intense western boundary currents it is higher than 100 $cm^2$. In fact, over most of the area with high model–data residuals the cross-covariance is as important as the model and measurement error covariances. Since the cross-covariance cannot be neglected the estimates obtained by using the method of F93 are biased. The fact that the cross-covariance is mostly positive suggests that either the model error is underestimated or the measurement error is overestimated[2]. This conclusion is supported by the additional analysis presented below.

To check the above estimates, we used the CMA with the same parametrization as above (equation 5.12), but adding time-lagged covariances into the CMA (equation 5.5). First we add covariances of the lag 3 difference (9 days), $\mathbf{D}_3$. the T/P has cycle at 9.9 days. Since lag–3 corresponds to 9 days (time step of the reduced-state model is 3 days) we can obtain significant estimates of the lag–3 difference covariance. In addition, the higher lag reduces the effect of neglected time-correlations in the model and measurement errors. The estimates of the parameters change to

$$\alpha_1 = 1.15 \pm 0.14, \text{ and } \alpha_2 = 0.42 \pm 0.03. \tag{5.15}$$

These estimates are significantly different from those obtained above (equation 5.13) with $\mathbf{Y}$ only. That is, use of additional time-correlation information gives estimates which are outside the uncertainty range of the earlier ones, equations 5.13. This is explained by the fact that now we have to fit not only the variances of the residuals but also the

---

[2]This can be seen by observing that $\mathbf{R}$ makes a positive contribution to the cross-covariance, while $\mathbf{HPH}^T$ makes negative contribution to the cross-covariance in equation (D.11).
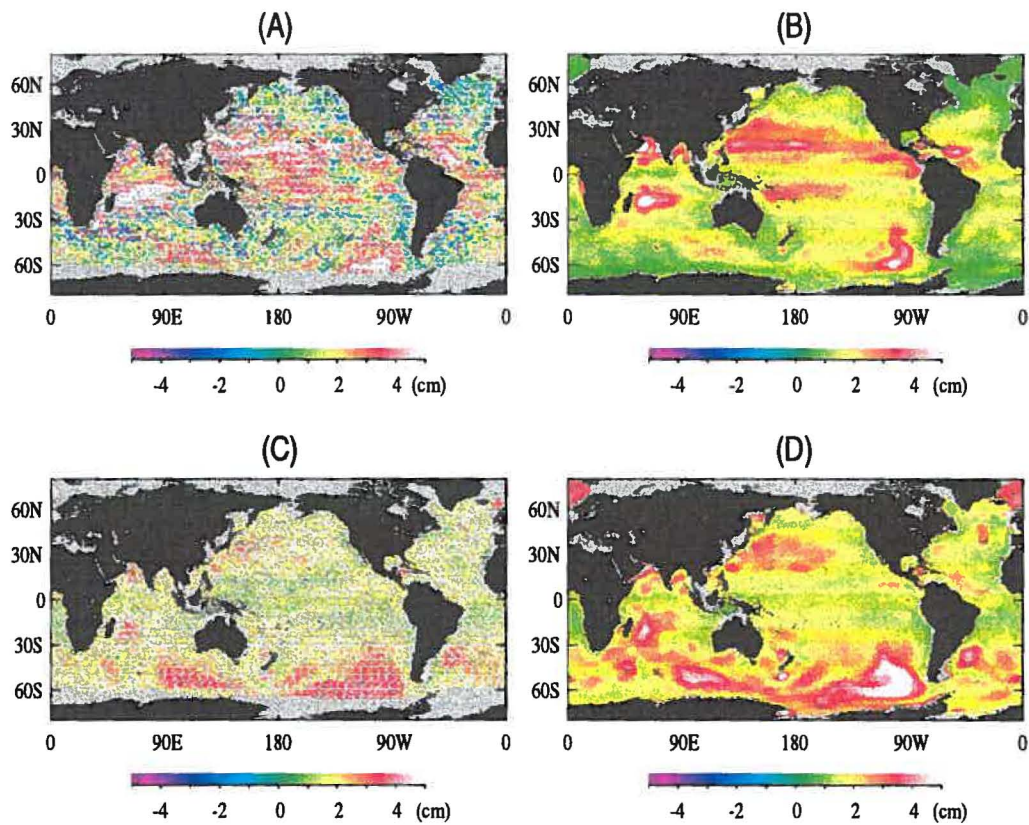
Figure 5.6: RMS differences of the model-data residuals for the data assimilation done in F99; (a) simulation minus KF forecast, (b) expected values of (a), (c) KF forecast minus KF update, (d) expected values of (c). Sign as indicated above, e.g. in (a) positive values indicate larger simulation residual than KF forecast residual (Reproduced from F99).
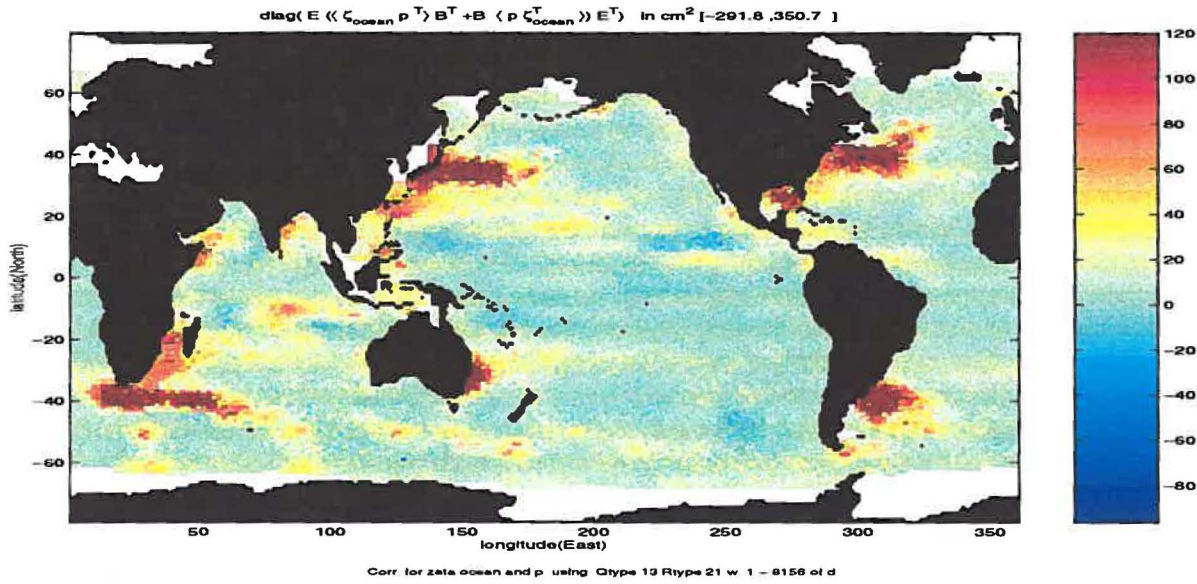
Figure 5.7: Estimate of the point cross-covariance between $\mathbf{H}\mathbf{p}$ and $\mathbf{H}\hat{\zeta}_{GCM,r}$ using equation (D.11) for parametrization of F99 and estimates given in equation (5.13). The minimum and maximum values are given in the square brackets.

time-lag correlations, and the lower estimates of the mean measurement error variance are in better agreement with the time-correlation structure in the residuals. However, it has to be pointed out that this change in estimates could be also due to neglected time-correlations in the errors (section 3.2.1).

With these estimates, the model error on average accounts for 40 percent of the model–data residual variance (Figure 5.8a), and explains most of the variance in large regions of the global ocean. We obtained similar estimates adding higher lag-differences covariances.

## 5.4.3 Data Assimilation with Parametrization of F99

In this section we present results of data assimilation with the error covariances derived in the previous section. The question we are going to address is whether we can improve quality of the state estimates. We use the data assimilation scheme of F99, but present only results with the approximate KF, omitting the smoother (Section 5.2).

163
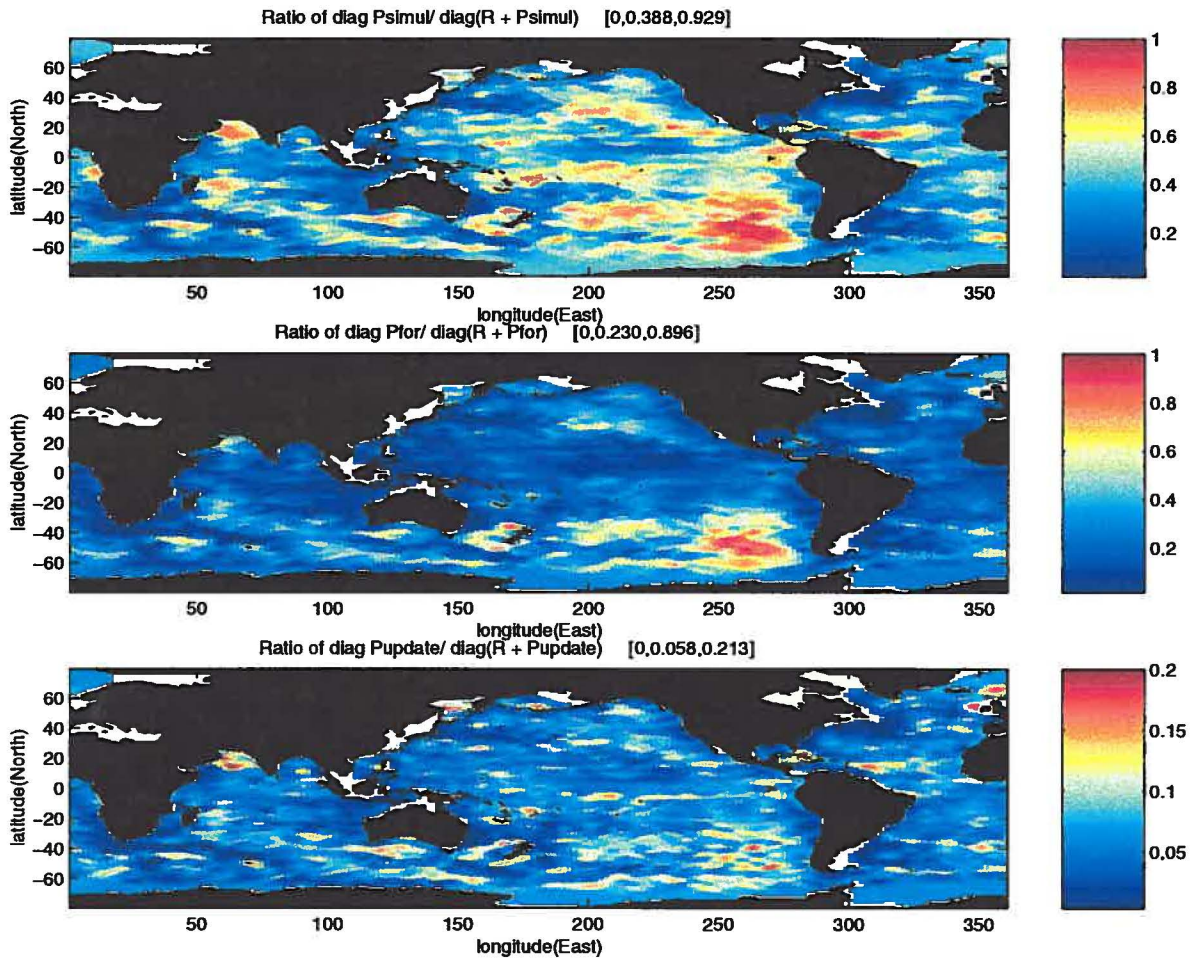
Figure 5.8: Relative ratio of the model and measurement errors for the rescaled covariance model used by F99 for : a) the model error (simulation error) $\mathrm{diag}\mathbf{HPH}^T/(\mathrm{diag}\mathbf{HPH}^T+\mathrm{diag}\mathbf{R})$; b) forecast error forecast $\mathrm{diag}\mathbf{H\Pi_s}(-)\mathbf{H}^T/(\mathrm{diag}\mathbf{H\Pi_s}(-)\mathbf{H}^T+\mathrm{diag}\mathbf{R})$; c) update error $\mathrm{diag}\mathbf{H\Pi_s}(-)\mathbf{H}^T/(\mathrm{diag}\mathbf{H\Pi_s}(-)\mathbf{H}^T+\mathrm{diag}\mathbf{R})$. Values greater than one half indicate that most residual variance is explained by the model error, and less than one half by the measurement error (which includes representation error). The minimum, mean and maximum values are given in the square brackets.

164

We use the rescaled error covariances of F99 (Section 5.4.2) with the parameters $\alpha$ derived using time-correlation information:

$$\mathbf{Q} = 1.15 \, \mathbf{Q}_{F99}, \text{ and } \mathbf{R} = 0.42 \, \mathbf{R}_{F99}. \tag{5.16}$$

First, we discuss time-asymptotic solutions for the forecast and update error covariances, $\mathbf{\Pi}_s(-)$ and $\mathbf{\Pi}_s$ (the reader is referred to Section 2.4 for the KF terminology). These error covariances provide a theoretical expectation of the uncertainties for the forecast and update fields. In Figure 5.8b we display the local ratio of the forecast error variance projected onto the T/P grid and the sum of the forecast and the measurement error variances,

$$\text{diag} \left( \mathbf{H} \mathbf{\Pi}_s(-) \mathbf{H}^T \right) / \text{diag} \left( \mathbf{H} \mathbf{\Pi}_s(-) \mathbf{H}^T + \mathbf{R} \right). \tag{5.17}$$

In Figure 5.8c we display the local ratio of the update error variance projected onto the T/P grid and the sum of the update error and the measurement error variances,

$$\text{diag} \left( \mathbf{H} \mathbf{\Pi}_s \mathbf{H}^T \right) / \text{diag} \left( \mathbf{H} \mathbf{\Pi}_s \mathbf{H}^T + \mathbf{R} \right). \tag{5.18}$$

Figure 5.8a shows the fraction of the total error variance explained by the simulation error covariance[3] derived in Section 5.4.2. The KF algorithm guarantees that the uncertainty is reduced once the assimilation takes place, i.e.

$$||\mathbf{P}|| \geq ||\mathbf{\Pi}_s(-)|| \geq ||\mathbf{\Pi}_s||. \tag{5.19}$$

These plots show that, as a result of the assimilation, the uncertainty of the forecast should be significantly reduced compared to the simulation, with a similar reduction from the forecast to the update. Furthermore, the update error is much less than the measurement error everywhere (note that the scale in Figure 5.8c is from 0 to 0.2 and from 0 to 1 in the other plots). A similar plot for the original simulation of F99 (unscaled

---

[3]i.e. the solution to the Lyapunov equation, or time-asymptotic approximation of the error covariance for simulation of the model without any data assimilation.

**Q** and **R**) is shown in Figure 5.4. While the simulation and forecast error covariances are quite different for the two experiments, the update uncertainties are very similar. The simulation model error variance explains on average 39 and 20 percent of the total error variance for the rescaled and original cases respectively; the forecast error explains on average 23 and 12 percent of the total error variance for the rescaled and original cases respectively; and the update error explains on average 6 and 4 percent of the total error variance for the rescaled and original cases respectively. Furthermore, the pattern of the fraction of the total variance explained by the model state uncertainty becomes similar as more data is assimilated. This shows that the rescaling of the measurement error covariance would be only significant for the forecasts of the state, but the update estimates should be nearly identical.

The error covariances influence estimates of the state through the weighting matrix, the Kalman gain (equation 2.28). In Figure 5.9a we display the projection of the Kalman gain onto the measurement grid, i.e. the diagonal of **HK**. This is equivalent to the local sea level anomaly associated with Kalman filter changes in model state corresponding to data–model difference (here the data denotes observations used in the KF, i.e. GCM-T/P residuals) of 1 cm. The value of zero implies that the weighting on observations is zero, and corresponding anomaly is 0, i.e. no assimilation takes place. The value of one means that the forecast is completely discarded, and the measurements are projected onto the model grid. The maximum projection of the Kalman gain is 0.2 and the average is 0.04. In other words, the average update is 96 percent forecast and 4 percent observations. For comparison, in Figure 5.9b we show a similar picture with the original assimilation of F99. Although the measurement error covariance has been rescaled by 0.4, the original Kalman gain is hardly changed. Accordingly, the estimates of the state are changed very little when we run the approximate KF with rescaled error covariances.

We now turn our attention to the estimates of the state. The resulting estimate of the ocean state is a three-dimensional $(x, y, z)$ time series for each of the GCM variables. To facilitate comparison among different data assimilation experiments (i.e. experiments

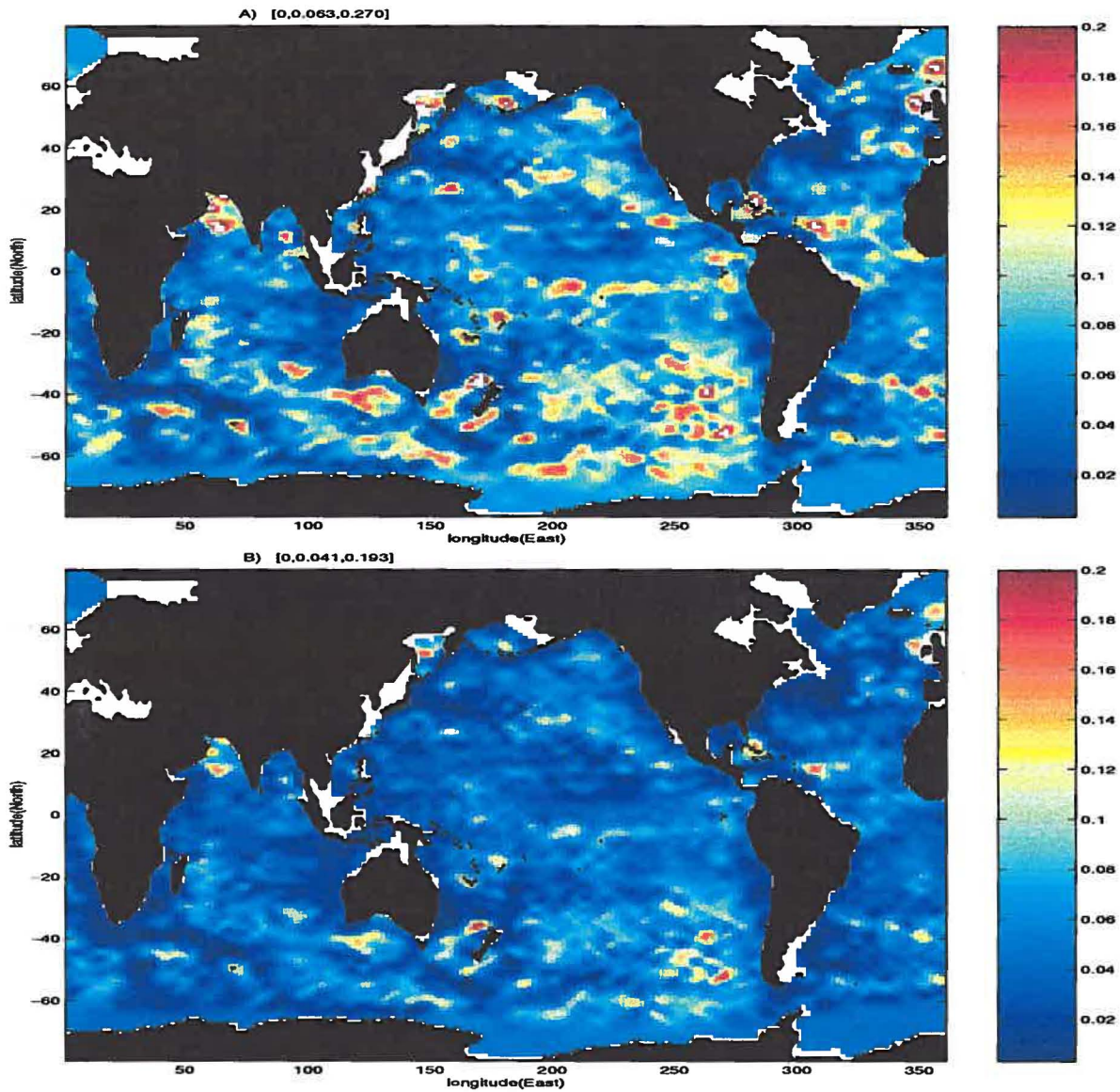Figure 5.9: Sea level anomaly (cm) associated with Kalman filter changes in model state corresponding to an instantaneous 1 cm model–data difference for a) the rescaled error covariances of F99, b) the original error covariances of F99. The estimates are strictly local reflecting sea level difference at each separate grid point and assuming the instantaneous data distribution used to derive the time-asymptotic limit of the KF.

with different choices of the error covariances) we only show estimates of the projection of the model state onto the sea surface height anomaly on the T/P grid.

To gain insight into the KF estimates, we compare the skill of different estimates of the model state against the T/P observations. Namely, we computed the following differences:

$$\text{cov} \left( \boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM, simul}} \right) - \text{cov} \left( \boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM, forecast}} \right), \quad (5.20)$$

$$\text{cov} \left( \boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM, forecast}} \right) - \text{cov} \left( \boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM, update}} \right), \quad (5.21)$$

where the subscript *simul* stands for the simulation of the state of the GCM, i.e. without any data assimilation; the subscript *forecast* stands for the forecast state of the GCM, i.e. the estimate immediately prior the recursive assimilation of observations; and the subscript *update* stands for the estimate of the state immediately after the recursive assimilation. The former, equation (5.20), quantifies the ability of the model in propagating data information consistently in time, whereas the latter, equation (5.21), quantifies the measurements' effect on the model at each instant of the assimilation and is directly dependent on the Kalman gain.

In Figure 5.10 we show these two quantities for the assimilation with the error covariances of F99, a) and b), and for the assimilation with the rescaled error covariances of F99, c) and d). The most remarkable feature of this figure is that the plots of the model–data differences for the simulation minus the forecast, a) and c), are very similar for the two assimilations. This is in agreement with the earlier estimates of the Kalman gain, Figure 5.9, which predicted that the purely local effect of the data assimilation is very small, and comparable, for the two assimilations. The average improvement of the RMS of the forecast over the RMS of the simulation is 1.77 cm and 1.79 cm for the run with original error covariances of F99 and the rescaled error covariances of F99, respectively. The lower plots, Figure 5.10b and Figure 5.10d, show that the assimilation is greater in the rescaled case since there is larger difference between the forecast and the update fields. However, greater amount of assimilation at time $t$ does not give a better estimate of the field, since the one step forecast is not on average closer to the observations.
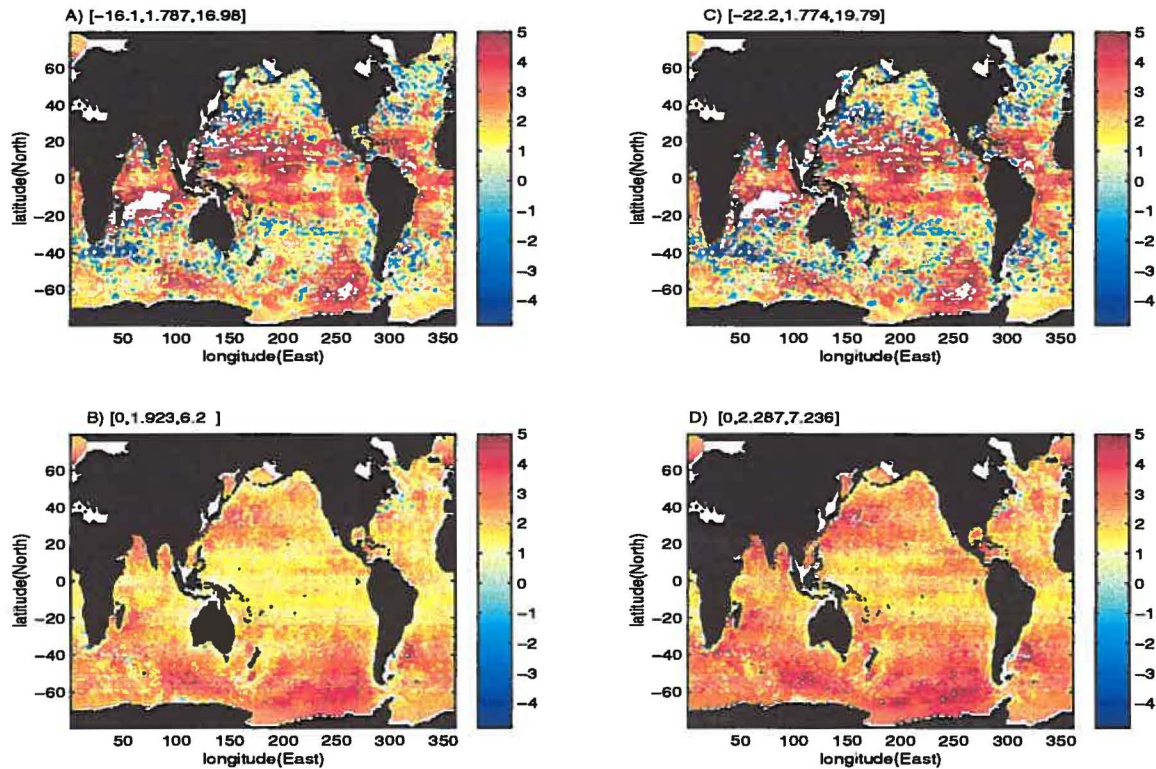
Figure 5.10: Differences (in cm) of model-data residuals; a) simulation minus forecast for the KF with the error covariances of F99; b) forecast minus update with the error covariances of F99; c) simulation minus forecast for the KF with the rescaled error covariances of F99; d) forecast minus update with the rescaled error covariances of F99. Values are RMS differences of model-data residual variances. The sign is as defined above, e.g. positive values in a) indicate larger simulation residual than forecast residual. Values above 5 cm are shown in white, while values less than -5 cm are shown in the deep blue. Note that the lower plots, b) and d), are strictly positive indicating that the update is always closer than the data, as required by the KF algorithm. In contrast, the simulation, a) and c), are often closer to the data than the forecast, indicating poor skill of the data assimilation.

Next, we consider sample variance of the innovations (equation 2.72), i.e. the difference between the observations and one-step forecast of the model. The innovations make a good check on the quality of the assimilation as they contrast a model forecast (the estimate immediately prior the recursive assimilation of observations) with independent data, i.e. the data not yet used in the analysis. If the model and the data were perfect, the model provides a perfect forecast and the innovations are zero, i.e. there is no new information in the observations. However, since the model and the data are not perfect, the innovations are different from zero. A more optimal data assimilation scheme on average provides a better estimate (update) of the state: a better update at time $t$ is used as the initial condition for the model forecast at time $t + 1$ and thus gives a better approximation of the true trajectory. This gives a better estimate of the update, and so on recursively (Figure 2.5 demonstrates the process graphically).

In Figure 5.11 we show an estimate of the differences of the RMS of innovations for the run of F99 and the run with rescaled error covariances. A positive difference corresponds to areas where the RMS of the innovations of the run with the original error covariances of F99 is higher, i.e. the areas where the run with the rescaled error covariance gives a better data assimilation estimate. Areas where the RMS of innovations for the run of F99 is smaller, i.e. the difference is negative, correspond to regions where the run with the rescaled error covariances gives a worse data assimilation estimate. The average difference is 0.02 cm, i.e. the rescaled run is on average better by 0.4 per cent (the average variance of the innovations for the run of F99 is 5.8 cm). This is a very small difference, predicted by analysis of the Kalman gain (Figure 5.9). There are several regions where we have significant improvement in the quality of the data assimilation estimates, e.g. in the North East Pacific. On other hand there are several areas where the estimates are worse, e.g. east of South Africa.

This chapter demonstrated that data assimilation with rescaled error covariances of F99 is very similar to the one presented in the original paper of F99. This is shown both by the analysis of the theoretical uncertainties and the sample estimates of the state. This is
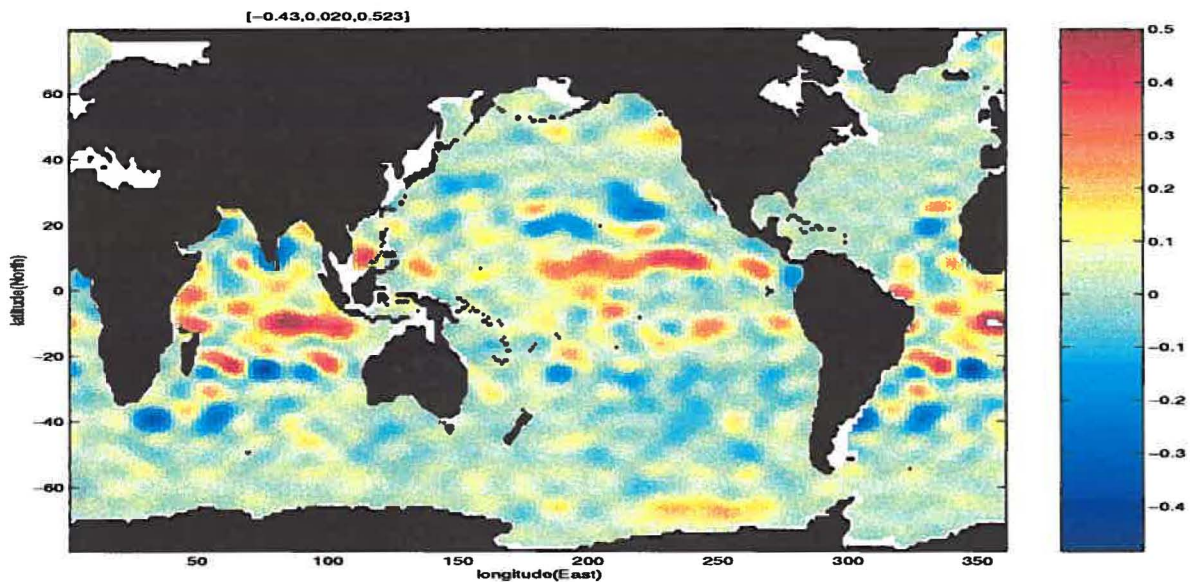
Figure 5.11: Difference (in cm) between the variance of the innovations (data–forecast residual) for the run of F99 and the run with the rescaled error covariances of F99. The fields has been extrapolated from the T/P grid onto the 2° by 1° grid by using a Fourier-based technique. Values less than 0 indicate areas where the assimilation of F99 is doing better, and values greater than zero indicate areas where the assimilation with the rescaled error covariances is doing better. The average is 0.02 cm.

explained by the fact that although the rescaled error covariances increase the respective weight on the model error as compared to the measurement error, the resulting change in the Kalman gain (the weighting matrix used in blending the recursive model forecast and the data) is very small. Accordingly, the change of variance of the innovations is very small.

## 5.5 Data Assimilation with an Independent Parametrization of the Error Covariances

In the previous section we presented results of data assimilation with the error covariances given in F99. There the model-data residuals were used to tune their parametrization of the error covariances, and thus the estimates derived in the previous section were not based on independent data. Thus the error $\varepsilon$ in the matching equation (5.8) was not completely independent of the sample variances $\mathbf{d}$, and the estimates of the rescaling parameters were possibly biased. In addition, we could not provide a proper derivation of the uncertainty of the sample estimates of $\boldsymbol{\alpha}$ as it is very difficult to account for possible dependence of the Green function $\mathcal{G}$ and the sample estimates $\mathbf{d}$ (equation 5.8). In this section we present analysis with a completely independent parametrization of the error covariances.

We parameterize the model error covariances as the sum of 691 (one for each point on $10°$ by $5°$ grid of the reduced model) delta matrices:

$$\mathbf{Q} = \sum_{i=1}^{691} \alpha_i \boldsymbol{\Gamma}_{3 \text{ days}} \boldsymbol{\Delta}_{i,i} \boldsymbol{\Gamma}_{3 \text{ days}}^{\mathrm{T}}, \tag{5.22}$$

where $\boldsymbol{\Gamma}_{3\text{days}}$ is a projection from the the wind stress field to the reduced-state model variables (a 2764 by 1382 matrix)[4], $\boldsymbol{\Delta}_{i,i}$ is a 1382 by 1382 diagonal delta matrix with the variance of the zonal wind at the $i$th and the variance of the meridional wind at the

---

[4]There are 691 spatial grid points for the reduced-state model and 4 variables (Section 5.1), i.e. 2764 model variables. Wind stress has two components, zonal and meridional, and thus 691 times 2, 1382 variables.

691 + $i$th positions on the diagonal, and zeros everywhere else. Thus, we assumed that the variance is scaled by the same constant for both zonal and meridional winds, but the scaling factor is changing in space. Note that because the projection matrix $\Gamma_{3days}$ is a full matrix, the resulting error covariance $\mathbf{Q}$ is a full matrix as well.

The measurement error covariance is assumed to be diagonal, i.e. measurement errors are assumed to be uncorrelated in space. The variance of the measurement errors is assumed be locally homogeneous in space. To define local areas of constant measurement error variance we divide the global ocean surface into 382 areas of equal variance by splitting each of the three oceans: Pacific, Atlantic and Indian, into 4 longitudinal areas for each 3° latitudinal band.

Firstly, we discuss the Kalman gain obtained with this error parametrization. The projection of the Kalman gain onto the sea level anomaly is shown in Figure 5.12. This is equivalent to the local sea level anomaly associated with Kalman filter changes in model state corresponding to 1 cm innovation, or data–model difference (here the data denotes observations used in the KF, i.e. GCM-T/P residuals). On average, the Kalman gain is smaller than that obtained with either the original or the rescaled error covariances of F99, Figure 5.9, i.e. less assimilation takes place with this error model. Furthermore, the structure of the local response to the innovations has changed dramatically. This choice of the error covariances disregards the observations completely in the tropics, and the Kalman gain is significantly reduced in the South East Pacific. On the other hand, there are high values in the East Indian ocean, and in the Kuroshio. In Figure 5.13 we show an estimate of the differences of the RMS of innovations for the run of F99 and the run with this new error covariances. A positive difference corresponds to areas where the RMS of the innovations of the run with the original error covariances of F99 is higher, i.e. the areas where the run with the new model for the the error covariances gives a better data assimilation estimate. Areas where the RMS of innovations for the run of F99 is smaller, i.e. negative difference, correspond to regions where the run with the new model for the the error covariances gives a worse data assimilation estimate. The average

173
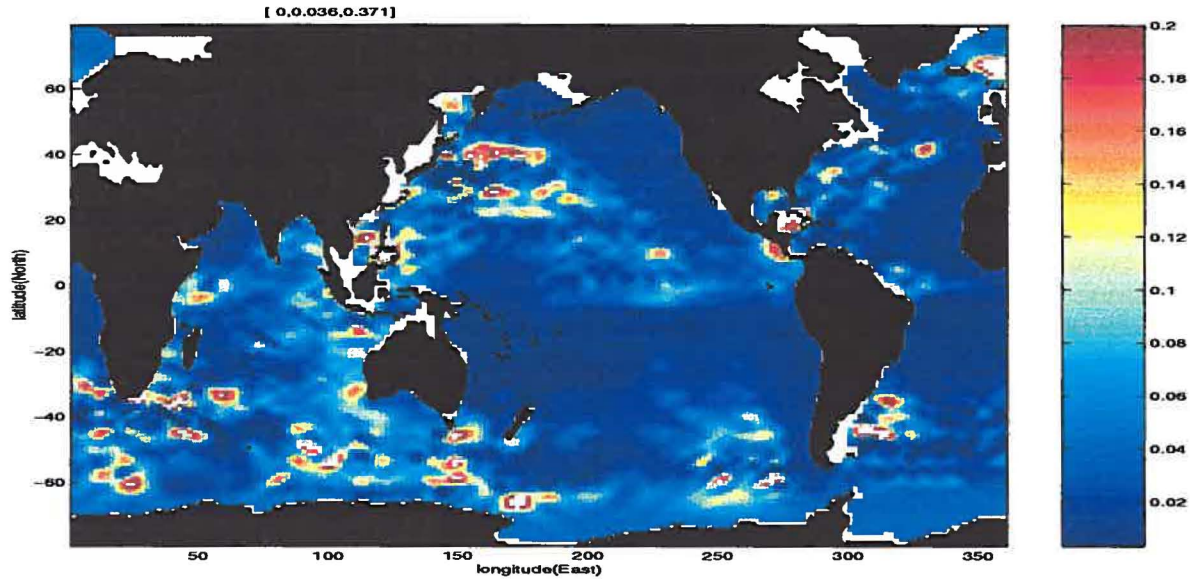
Figure 5.12: Sea level anomaly (cm) associated with Kalman filter changes in model state corresponding to an instantaneous 1 cm model–data difference for a new parametrization of the error covariances. The estimates are strictly local reflecting sea level difference at each separate grid point and assuming the instantaneous data distribution used to derive the time-asymptotic limit of the KF.

difference is $-0.3$ cm, i.e. the run of F99 is on average better by 5 per cent. The negative impact is due to very small values of the model error, and accordingly small values of the Kalman gain, in the tropical Pacific and in the West Indian ocean (Figure 5.12). In other words, this data assimilation run fails to extract information available in the observations in these regions, information that was successfully used in the assimilation of F99 (Figure 5.10a). On the other hand, in several regions the assimilation is improved, e.g. in the East Indian Ocean and the tropical Atlantic.

This section presents results of data assimilation with a different parametrization of the error covariances. A much larger, more than a thousand vs. two with the rescaled model of the F99, number of parameters is estimated using the CMA. The estimates of the parameters are sensitive to the choice of the uncertainty matrices used to invert the global matching equation (5.8). Nevertheless, we use one set of such estimates with an approximate KF. The results are negative, i.e. the resulting impact of the assimilation on
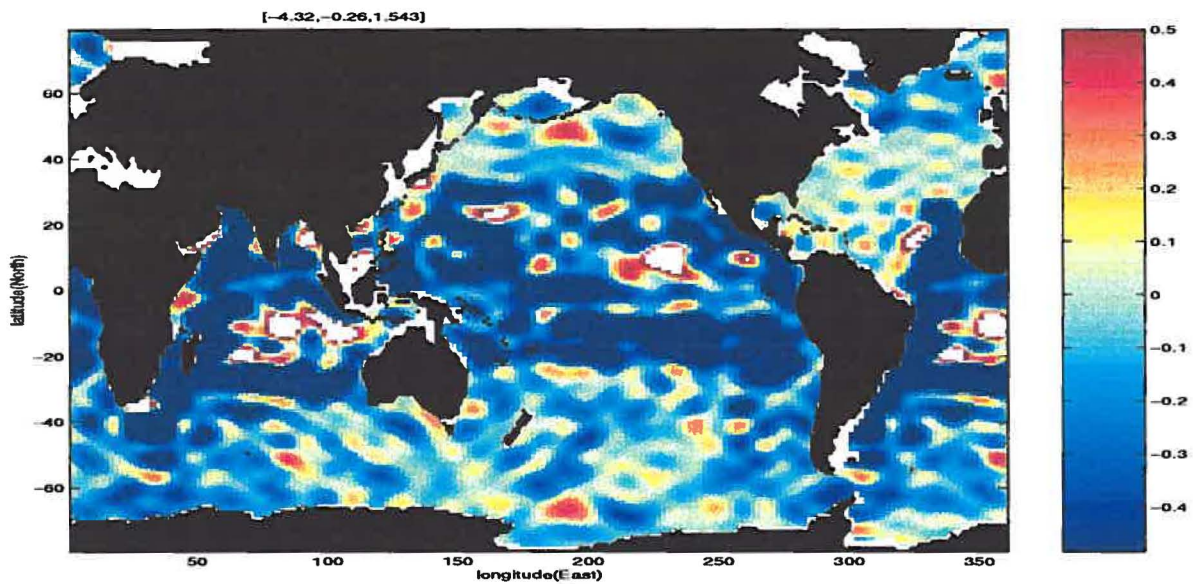
Figure 5.13: Difference (in cm) between the variance of the innovations (data–forecast residual) fo the run of F99 and the run with the new parametrization of the error covariances. The fields has been extrapolated from the T/P grid onto the 2° by 1° grid by using a Fourier-based technique. Values less than 0 indicate areas where the assimilation of F99 is doing better, and values greater than zero indicate areas where the assimilation with the rescaled error covariances is doing better. The average is -0.3 cm.

175
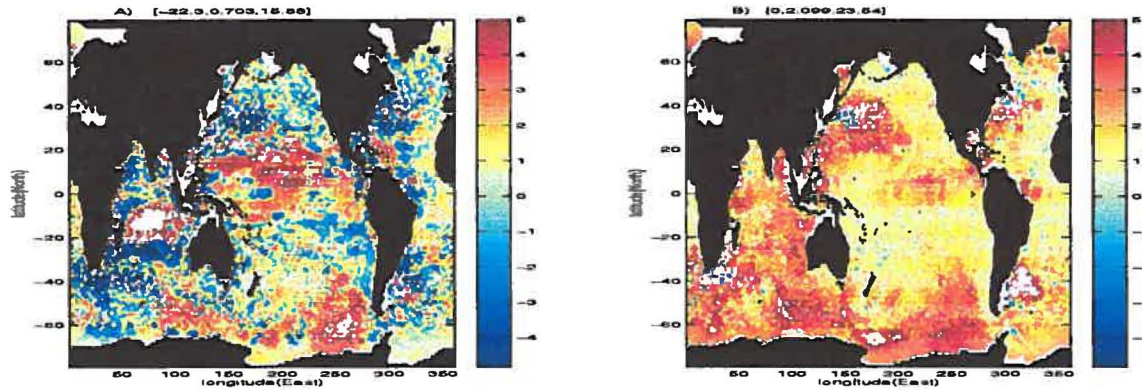
Figure 5.14: Differences (in cm) of model-data residuals with the new parametrization for the error covariances: a) simulation minus forecast for the KF ; b) forecast minus update with the error covariances of F99. The sign is as defined above, e.g. positive values in a) indicate larger simulation residual than forecast residual. Values above 5 cm are shown in white, while values less than -5 cm are shown in the deep blue. Note that the lower plots, b) and d), are strictly positive indicating that the update is always closer than the data, as required by the KF algorithm. In contrast, the simulation, a) and c), are often closer to the data than the forecast, indicating poor skill of the data assimilation.

the ocean state estimates is less than that in F99. This is explained by the fact that the estimate of the model error covariance is very small in many regions of the global ocean, and accordingly the data is not used in the assimilation in these areas, most notably the tropical Pacific. Therefore, the assimilation fails to improve the estimate of the ocean state in these areas. There are several smaller regions where the estimate is improved, but the cumulative impact is negative, and the average variance of the innovations is greater by 0.3 cm than that in F99. This demonstrates that while in principle the CMA can estimate a large number of parameters, the estimates are not very useful. Thus, while the estimate of the state is improved over the simulation (Figure 5.14a), the average difference between the update and the forecast field is smaller than in the data assimilation experiments with the error covariances of F99(Figure 5.14a).

## 5.6 Partitioning of the Model Error

In this section we demonstrate how the CMA method can be used to study partitioning of the model errors. To understand contributions of each of the vertical modes to the covariance model of F99, equation (5.12), we next choose to parameterize the model error covariance as a sum of four diagonal matrices, one $\mathbf{Q}_k$ for each of the four coarse state model variables (equation 5.1):

$$\mathbf{Q} = \sum_{k=1}^{4} \alpha_k \mathbf{Q}_{\mathrm{diag},k}, \mathbf{R} = \alpha_5 \mathbf{R}_{\mathrm{F99}}, \tag{5.23}$$

$$\mathbf{Q}_{\mathrm{diag},1} = \mathrm{diag}\,\mathbf{Q}_{\mathrm{F99}(u,u)}, \mathbf{Q}_{\mathrm{diag},2} = \mathrm{diag}\,\mathbf{Q}_{\mathrm{F99}(v,v)}, \mathbf{Q}_{\mathrm{diag},3} = \mathrm{diag}\,\mathbf{Q}_{\mathrm{F99}(h,h)},$$

$$\mathbf{Q}_{\mathrm{diag},4} = \mathrm{diag}\,\mathbf{Q}_{\mathrm{F99}(\psi,\psi)},$$

where $\mathrm{diag}\,\mathbf{Q}$ denotes a diagonal matrix with the diagonal equal to the diagonal of $\mathbf{Q}$, and $\mathbf{Q}_{\mathrm{F99}(u,u)}$ denotes a model error covariance for the zonal baroclinic velocity $u$, and so on. Thus, we are assuming that the model error covariance has zero off-diagonal elements, unlike the full covariance used in F99. Using only the variance of the residual, $\mathbf{Y}$, the resulting estimate of the projection for the model error onto the sea level variance is very similar to that obtained with the full model error covariance, $\mathbf{Q}_{\mathrm{F99}}$. In addition, we obtain similar estimates for the measurement error covariance, and thus, similar distribution of the fraction of model–data residual explained by the model error. Using the lagged differences the estimates of coefficients are

$$\alpha_1 = 0, \alpha_2 = 0, \alpha_3 = 1.99 \pm 0.05, \alpha_4 = 0.49 \pm 0.01, \alpha_5 = 0.50 \pm 0.01. \tag{5.24}$$

That is, although we are trying to estimate four parameters for $\mathbf{Q}$, only two, $\alpha_3$ and $\alpha_4$ are different from zero[5]. The projections of each of the delta model error covariances onto the sea level are shown in Figure 5.15.

They show that the model errors corresponding to the baroclinic velocities have a strong peak on the equator, but the spatial distribution of the associated sea level variance

---

[5]The least squares solution tries to set the other two to negative values, and we have to use constrained optimization instead.

is different for the zonal and meridional velocities. However, the structure of the model errors is inconsistent with the variance of the model–data residual, and therefore the estimates for the coefficients $\alpha_1$ and $\alpha_2$ are zero. The model errors corresponding to the vertical displacement have a strong peak between $20°$ and $40°$ latitude and explain half of the model variance. The model errors corresponding to the barotropic stream function have maximum in high latitudes and explain the maximum in the South-East Pacific of the South America. Fukumori et al. (1998) have found that most of the sea level variance in high latitudes is explained by the barotropic mode. Our results suggest that not only the sea level but the also the errors in high latitudes are dominated by the barotropic mode. Note that these results are very different from the results of Chapter 4 where we could not distinguish amplitudes of the errors of different internal modes with the T/P data. The difference is due to the fact that here we are trying to discriminate between the errors in the barotropic and baroclinic modes instead of the errors in four internal vertical modes.

To recapitulate, the sum of four diagonal matrices can explain the pattern given by the full matrix $\mathbf{Q}_{F99}$. That is, if we only use diagonals of the covariance and lagged difference covariances of the model–data residual, we cannot prefer one parametrization over the other (equations 5.12 and 5.23).

To check the robustness of these results we ran the CMA with the following parametrization of the model error covariance $\mathbf{Q}$

$$
\mathbf{Q} = \begin{bmatrix} \alpha_1 \mathbf{Q}_{T/P} & \mathbf{0}_{691} & \mathbf{0}_{691} & \mathbf{0}_{691} \\ \mathbf{0}_{691} & \alpha_2 \mathbf{Q}_{T/P} & \mathbf{0}_{691} & \mathbf{0}_{691} \\ \mathbf{0}_{691} & \mathbf{0}_{691} & \alpha_3 \mathbf{Q}_{T/P} & \mathbf{0}_{691} \\ \mathbf{0}_{691} & \mathbf{0}_{691} & \mathbf{0}_{691} & \alpha_4 \mathbf{Q}_{T/P} \end{bmatrix}, \tag{5.25}
$$

where $\mathbf{Q}_{T/P}$ is a diagonal matrix of 691 by 691 with the diagonal given by the variance of the T/P measurements and $\mathbf{0}_{691}$ is a zero matrix of 691 by 691. The resulting contributions of the model error of each of the four model variables are shown in Figure 5.16. The results are similar to those shown above, but the coefficients for the baroclinic velocities,
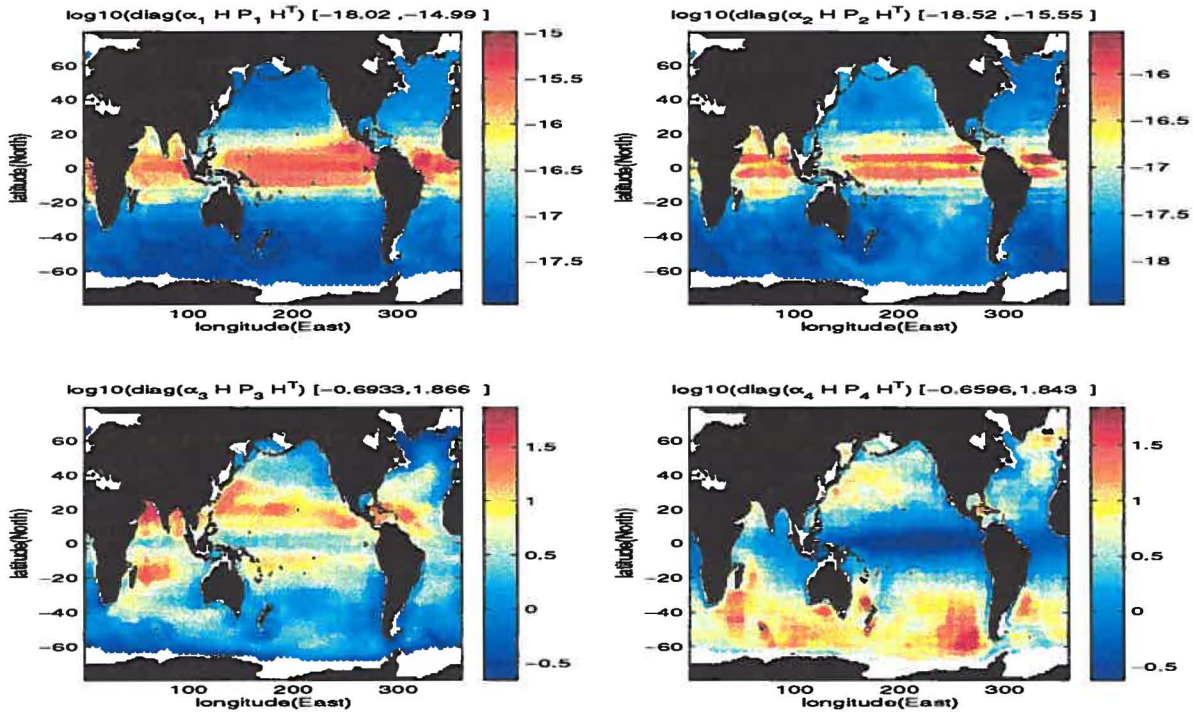
Figure 5.15: Log10 of diagonals of $\alpha_k \mathbf{H} \mathbf{P}_k \mathbf{H}^{\mathrm{T}}$. The covariance model consists of four diagonal matrices for $\mathbf{Q}$ and the measurement error covariance used in F99. The estimates were obtained using $\mathbf{Y}$ and $\mathbf{D}_3$. Note the top two plots are zero, and are presented only to show the signature for the baroclinic velocities $u$ and $v$. The minimum and maximum values are given in the square brackets.

$\alpha_1$ and $\alpha_2$ are different from zero. They are however very small, despite the fact that the projections onto the sea level have a significant signal off the equator. The model error corresponding to the vertical displacement has a strong peak in latitudes higher than 40°, but is smaller that the model error corresponding to the barotropic stream-function. It is worth noting that the model errors corresponding to the streamfunction are very similar to those in the previous parameterizations. That is, two very different diagonal $\mathbf{Q}_4$, first with the diagonal given by the NCEP winds variance and second with the diagonal given by the T/P measurement variances, have very similar projections on the sea surface height. This indicates that the results of data assimilation should not be very sensitive to the structure of the prior model error covariance $\mathbf{Q}$. The coefficient for
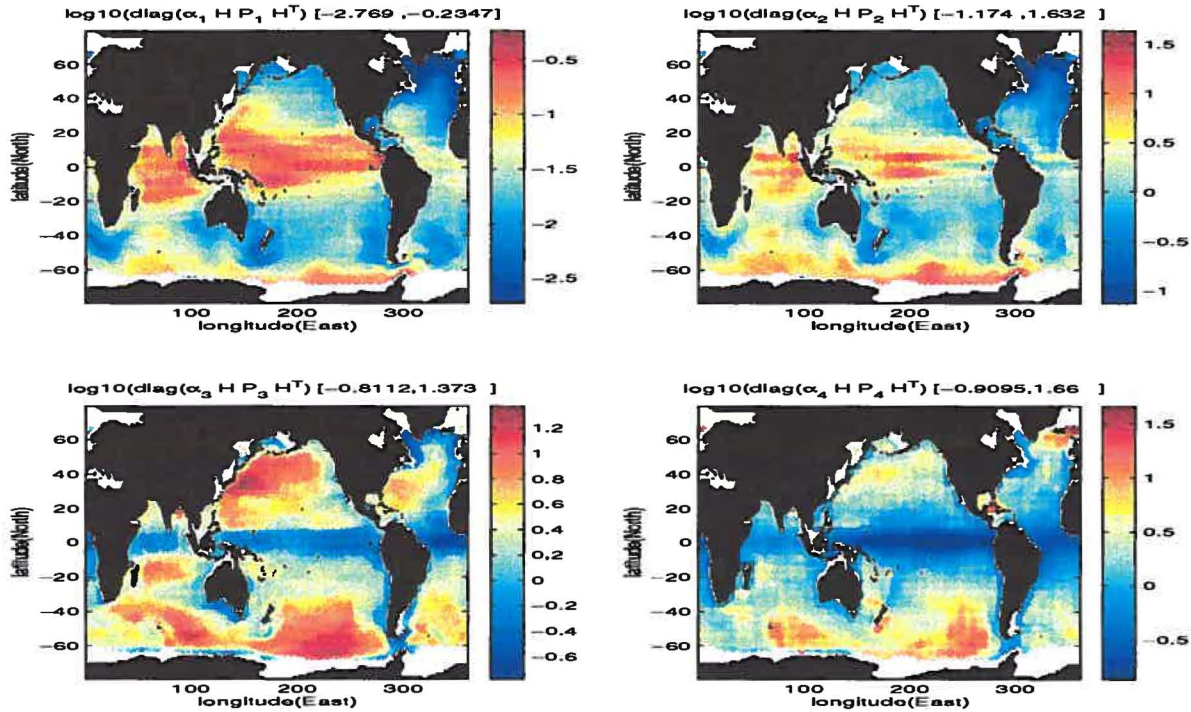
179

Figure 5.16: Log10 of diagonals of $\alpha_k \mathbf{HP}_k\mathbf{H}^T$. The model error covariance is parameterized as a sum of four diagonal matrices with the diagonal given by the variance of the T/P measurements (equation 5.25). The estimates were obtained using $\mathbf{Y}$ and $\mathbf{D}_3$. The minimum and maximum values are given in the square brackets.

the measurement error covariance is similar to the one obtained earlier in equation 5.15, $\alpha_5 = 0.50 \pm 0.01$. Thus, this shows that our estimates obtained earlier are robust.

## 5.7  Summary

In this chapter we applied the CMA to a global integration of the GFDL GCM and 3 years of TOPEX/POSEIDON data. We have shown that the estimates used in F99 overestimate the measurement error variance and that in order to match additional lag-difference covariances we need to increase the fraction of the model–data residual variance explained by the model error covariance. We tested this hypothesis with a variety of parameterizations for the model and measurement error covariances, and obtained similar

conclusions. The resulting estimate of the model error covariance on average explains 40 percent of the model–data residual variance, but the uncertainty is significantly reduced when the model is run with an approximate KF. In addition, we have demonstrated that most of the model error variance is explained by the barotropic mode, and that the model error corresponding to baroclinic velocities has a negligible contribution. This can be understood by noting that the model–data residual variance is much greater in the the mid and high latitudes than in the tropics (Figure 5.5). The baroclinic velocity contribution to the model errors is maximum in the tropics, and has a spatial pattern which is different from the pattern of the GCM–data residual variance. We tested this conclusion with two parameterizations of the spatial distribution of the model error variance: the first where the model error variance is proportional to the variance of the NCEP winds, and the second where it is proportional to variance of the sea level.

The CMA estimates of the error covariances are used with a global data assimilation scheme, but the quality of the data assimilation estimates is improved very little, as shown by the statistics of the innovations. While in principle the adaptively tuned error statistics should improve the data assimilation estimates, it is not necessarily achieved for each particular data assimilation. As pointed out in Chapter 3 the problem of error statistics estimation is very under-determined. To obtain statistically significant estimates of the error statistics it is crucial to have a good understanding of the structure of the error covariances, that is to have a good physical understanding of the model's shortcomings. The covariances used in F99, which are already tuned to the model-data residuals, use error structures which proved to be quite robust. Other statistical models give similar estimates. Ability of the CMA to provide estimates of the error covariances for other statistical models makes it possible to run several data assimilation experiments exploring the effect of different assumptions for the error statistics. The results of this chapter, which compares several data assimilation experiments which differ only by the choice of the error covariances demonstrate that data assimilation estimates are not very sensitive to a particular parametrization of the adaptively tuned error statistics.

tuned error statistics.

# Chapter 6

# Conclusions

## 6.1 Summary of the Thesis

Data assimilation is routinely used to study ocean processes, to test ocean model sensitivities, and to initialize ocean fields for forecasting. Data assimilation combines imperfect models with noisy observations to obtain the best possible estimates of the state of the model. The statistics of the model and measurement errors are prior information required to perform data assimilation. The measurement errors include not only the instrument noise but also representation error, i.e. processes which affect observations but that are missing from the model, and typically correspond to scales smaller than the model grid size. These missing processes aggravate the problem because poor knowledge of the shortcomings of the model translates into poor knowledge of both the model and the measurement error statistics. In addition, estimates of the state and the uncertainty of the state depend on the model and measurement error covariances. The major problem addressed in this thesis is adaptive estimation of the model and measurement error statistics for data assimilation with GCMs and global data sets. The term "adaptive" is used to stress that estimates of the error statistics are derived from the observations.

The principal contribution of this thesis has been to couch the error estimation problem in a familiar least squares context using the so-called covariance matching approach

(CMA). It then becomes possible to take advantage of a large number of tools from discrete linear inverse theory. The CMA is illustrated with two different models, the linearized MIT GCM and the linearized GFDL GCM which approximate large scale GCM dynamics. The data consist of TOPEX/POSEIDON (T/P) altimeter measurements of sea surface height and ATOC tomographic measurements which have been inverted to give anomalies of temperature. We show that the CMA can be used to obtain consistent and statistically significant estimates of the model and measurement error covariances. In addition, the CMA allows one to determine what components of the model and measurement errors can be resolved with a particular type of measurements. The method can also be extended to estimate other error statistics.

Following the introduction in the first chapter of the thesis, the second chapter starts by defining model and measurement equations for a reduced-state model. We show that in this setup, measurement errors include not only the instrumental errors, but also the representation errors. Representation errors correspond to processes which affect observations but that are missing from the model, and typically correspond to scales smaller than the model grid size. We then describe available methods of adaptive error estimation. These methods use new information available in observations at every time step (innovations) to update estimates of the error statistics. They are based on the ideas from control engineering literature and can be used online. Following the discussion in Blanchet et al. (1997) we focus first on the method of Myers and Tapley (1976) (MT). To introduce the method we first apply it to a scalar model. We then extend the results to a model with two variables, which allows for a thorough testing of the MT method. The method has several major drawbacks: (1) when we have fewer observations than the number of degrees of freedom in the model, it may be sensitive to the initial guess of the model error covariances, (2) it takes many iterations for the method to converge, and because the method requires running the Kalman filter, it is computationally very expensive, (3) simultaneous estimation of the model and measurement error statistics is unstable, (4) the method does not provide estimates of the uncertainties of the de-

rived error covariances, and (5) no information how much data is required, and which parameters can be estimated and which cannot.

We apply the method to a linearized version of the MIT GCM in combination with TOPEX/POSEIDON altimetric and ATOC acoustic tomography measurements in a twin experiment setup. As in the case with the low-dimensional models, the estimates depend on the initial choice of the error statistics and the type of observations used in the assimilation. In Section 2.9 we show that similar results are obtained with a maximum likelihood method. The conclusion is that neither of the adaptive data assimilation methods are suitable for quantifying large scale internal ocean model errors with the altimetric or acoustic tomography observations which are available at present.

In Chapter 3 we develop a new approach to adaptive error estimation which we call the Covariance Matching Approach (CMA). It is related to a method described in Fu et al. (1993) and Fukumori et al. (1999) who estimated the model and measurement error covariance by comparing the observations with the model forecast. Although related, the new approach relaxes some of the restrictive assumptions of the method used by Fu et al. (1993). It also utilizes information in a more efficient way, provides information on which combination of parameters can be estimated and which cannot, and allows the estimation of the uncertainty of the resulting estimates. Through a series of twin experiments, we show that the new covariance matching approach seems much better suited for the problem of estimating internal large scale ocean model error statistics with acoustic measurements, but not with altimetric measurements. In addition it allows the simultaneous estimation of measurement and model error statistics. This does not seem possible with the adaptive methods described in Chapter 2.

In Chapter 4 we apply the CMA to actual T/P and ATOC data to obtain estimates of the error statistics for the linearized MIT GCM. Based on twin experiments we conclude that the acoustic data but not the altimetric data can in principle provide reliable estimates of the vertical partitioning of the model error variance. We then use real data to show that the model error explains most of the GCM–data misfit variance and that

extensions of the CMA can be used to obtain information about the trends, the annual cycle amplitudes and the phases of the errors. However the limited duration of the ATOC time series and failure of T/P measurements to provide information about the vertical structure of baroclinic errors undermine the quality of the obtained estimates.

In Chapter 5 we apply the CMA to a second problem, one which involves estimating global ocean error statistics for a linearized GFDL GCM. The linearization has only two vertical modes, the barotropic and first baroclinic internal modes. This model has been recently used for a global data assimilation study using TOPEX/POSEIDON sea surface height measurements, Fukumori et al. (1999). In this setup the T/P measurements have sufficient information to differentiate between the barotropic and baroclinic error structures unlike the linearized MIT GCM which had four internal vertical modes based on temperature EOFs. Most of the model error is explained by the barotropic mode, and the success of the method is attributed to the fact that the barotropic and first baroclinic error modes have very different projections onto the GCM-data residuals.

The obtained estimates of error statistics are significantly different from those used in the study of Fukumori et al. (1999). However, the impact of this change on the estimates of the ocean state obtained with an approximate Kalman filter is very small. This is explained by the fact that the Kalman gain is very small regardless of which parametrization of model and measurement error statistics is used. This is due to the fact that the measurement errors, dominated by the representation error, are much larger than the model errors. In other words, the small-scale structure in the observations, , i.e. the mesoscale eddies, makes only a very small fraction of the data consistent with the model.

The thesis shows that the problem of adaptive error estimation of model and measurement error statistics can be addressed even with global GCMs and a few years of global ocean observations, such as the sea surface height measurements from T/P altimeter. When the error covariances can be parameterized by a few "delta" matrices, the CMA allows one to estimate the relative contribution of each matrix and the uncertainty

of the resulting parameters. However, it is important to stress that for large dimensional problems a clear physical understanding of the problem is crucial. Only through qualitative understanding of the model and the data can one choose reasonable parameterizations of the error covariances. Furthermore, the estimates are not guaranteed to remain valid if additional observations are included. This work has investigated some aspects of the error estimation problem with very large models and global data sets, but left many important issues untouched. Used in the right context the CMA can be useful to oceanographers facing a data assimilation challenge or trying to quantify the errors of a GCM, as exemplified by the two applications of the method with real data.

## 6.2 Future Work

The work presented in this thesis shows that adaptive estimation of error statistics can be done in an offline mode by using all available observations. There are several assumptions made in the CMA that might not hold, namely, the assumptions of zero correlation in time of model and measurement errors and zero correlation between model and measurement errors. Consistent biases in atmospheric forcing, such as sea surface winds, result in time-correlated model errors, which were neglected in the thesis.

Measurement error includes representation error, e.g. mesoscale eddies. Eddies can be correlated over several weeks, and therefore measurement errors would be correlated in time. It would be useful to investigate effect of these assumptions on estimates of the error covariances with realistic models.

Better tests of consistency of estimates would be helpful. A good understanding of the effects of misspecifying the delta error covariances, (for example, neglecting or over-simplifying spatial correlations in the errors) is also much needed. The second application presented in this thesis shows that with global datasets one needs to compress the information to be able to invert for the error statistics parameters. It would be very interesting to see whether the CMA approach can be coupled with large-scale estimation

techniques, such as the multi-resolution optimal interpolation method (Menemenlis et al. 1997).

An interesting test of the CMA would be whether it can improve data assimilation schemes used to initialize forecasting models, such as that used for El-Niño prediction. Using historic datasets one can do many hind-casts to investigate whether improved data assimilation can be indeed achieved.

The thesis lays the ground work for creation of new adaptive methods suitable for large scale problems with few and limited data sets. These efforts, and others like these, are needed to guide the selection of weight matrices in the cost function used for ocean estimation studies.

# Appendix A

# Notation and Abbreviations

| Symbol | Definition | Equations |
|---|---|---|
| **bold face** | vectors and matrices | |
| normal face | scalars | |
| $\hat{\,}$; $\tilde{\,}$; $\bar{\,}$; $\langle \cdot \rangle$; $\lvert \cdot \rvert$ | true value; estimate; sample mean; expectation; norm | |
| $^{\mathrm{T}}$; cov; $(\vdots)$ | transpose; covariance; column operator | |
| $\sim$ | comes from a distribution | 2.36 |
| $\boldsymbol{\alpha}$; $\alpha_k$ | vector of and individual parameters for $\mathbf{Q}$ and $\mathbf{R}$ | 2.35, 3.19, 3.4 |
| $\boldsymbol{\Gamma}$ | projects the large scale model error $\mathbf{u}(t)$ onto the coarse state | 2.3, 2.18 |
| $\gamma$ | misspecification of the meas. error cov. $\lvert \hat{\mathbf{R}} \rvert / \lvert \mathbf{R} \rvert$ | 2.52 |
| $\delta_{t,t'}$ | 1 if $t = t'$; 0 otherwise | 2.22 |
| $\boldsymbol{\epsilon}$ | high frequency components of the GCM state | 2.5 |
| $\varepsilon$ | sample error | 3.19, 3.38 |
| $\hat{\boldsymbol{\zeta}}_{\mathrm{GCM}}$; $\boldsymbol{\zeta}_{\mathrm{GCM}}$ | true and model states of the GCM | 2.2 |
| $\boldsymbol{\zeta}_{\mathrm{ocean}}(t)$ | state of the real ocean | 2.11 |
| $\boldsymbol{\eta}_{\mathrm{ocean}}$ | data | 2.11, 2.15 |
| $\mathcal{N}$ | the real ocean to the GCM state projection | 2.12 |
| $\boldsymbol{\nu}$ | ocean data error | 2.14 |
| $\boldsymbol{\nu}_{\mathrm{ocean}}(t)$ | instrument error | 2.14 |
| $\boldsymbol{\Pi}$ | uncertainty of the GCM error | 2.20 |
| $\boldsymbol{\Pi}(t+1\lvert t)$; $\boldsymbol{\Pi}(t\lvert t)$ | uncertainty of the KF forecast and analysis | 2.27, 2.30 |

Table A.1: Summary of notation.

| Symbol | Definition | Equations |
|---|---|---|
| $\Pi^f(t)$, $\Pi^a(t)$ | uncertainties of the scalar model forecast and analysis (update) dvivided by $\mathbf{R}$ | 2.38 |
| $\Pi^a{}_S$ | steady state $\Pi^a(t)$ | 2.42 |
| $\rho$, $\sigma$ | correlation coefficient, variance | 3.23 |
| $\mathbf{A}$ | linear dynamic model | 2.3 |
| $\mathbf{B}^*$; $\mathbf{B}$ | state reduction operator and its pseudo-inverse | 2.2, 2.4, 2.5 |
| $\mathbf{B}_h^*$; $\mathbf{B}_v^*$; $\mathbf{B}_t^*$ | horizontal, vertical, and time reduction operators | 2.8 |
| $\mathbf{C}_j(t)$ | lag $j$ covariance of innovations | 2.73 |
| $\mathbf{D}_s$ | $\mathrm{cov}[\mathbf{y}(t+s) - \mathbf{y}(t)]$ | 3.9 |
| $\mathbf{d}$ | vector of elements from $\mathbf{Y}$ and $\mathbf{D}_s$ | 3.19 |
| $\boldsymbol{E}_{\mathrm{ocean}}$ | projection from the real ocean to the observations | 2.11 |
| $\mathbf{E}$ | projection from the GCM state to the observations | 2.16 |
| $\mathbf{G}_{\mathbf{Y},k}$, $\mathbf{G}_{\mathbf{D}_s,k}$ | Green's function | 3.7, 3.10 |
| $\mathcal{G}$ | Green's function kernel matrix | 3.10, 3.19 |
| $\mathbf{G}$ | maps the forcing of the linear model onto the state | 2.18 |
| $\mathbf{H}$ | "observation matrix", relates the coarse state to the model–data residual | 2.21,3.2 |
| $\mathbf{I}$; $\mathbf{I}_{128}$ | identity matrix; subscript indicates size | 2.4, 2.70 |
| $\mathbf{K}$, $\mathbf{K}_S$ | Kalman gain and steady Kalman gain | 2.28 |
| $\mathcal{M}$ | GCM model | 2.1 |
| $M$ | number of observations, length of vector $\mathbf{y}$ | 3.12 |
| $N$ | number of DOF in the model, length of vector $\mathbf{p}$ | 3.12 |
| $\mathbf{P}$ | Covariance of the GCM errors | 2.57 |
| $\mathbf{p}(t)$ | GCM error | 3.2 |
| $\tilde{\mathbf{p}}(t+1|t)$ | Kalman filter forecast | 2.26 |
| $\tilde{\mathbf{p}}(t+1|t+1)$ | Kalman filter analysis (update) | 2.29 |
| $p$ | decorrelation number | 3.27 |
| $\hat{\mathbf{Q}}$, $\mathbf{Q}$, $\tilde{\mathbf{Q}}_{\mathrm{mt}}$ | true, prior and MT model error covariance | 2.22, 2.23, 2.32 |
| $\mathbf{Q}_k$ | parametrizations of $\mathbf{Q}$ | 2.35, 3.4 |
| $\mathbf{Q} \leftrightarrow [\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ | representation for block diagonal form of $\mathbf{Q}$ | 2.70 |
| $q$ | ratio of magnitudes of $\mathbf{Q}$ and $\mathbf{R}$ | 2.38 |
| $\tilde{q}$ | MT estimate of $q$ | 2.52 |

Table A.1: Continued.

| Symbol | Definition | Equations |
|---|---|---|
| $\hat{\mathbf{R}}$, $\mathbf{R}$, $\tilde{\mathbf{R}}_{\mathrm{mt}}$ | true, prior and MT measurement error cov. | 2.23, 2.34 |
| $\mathbf{R}_k$ | parametrizations of $\mathbf{R}$ | 2.35,3.4 |
| $\mathbf{R}_\alpha$, $\mathbf{R}_\varepsilon$ | cov $\boldsymbol{\alpha}$; cov $\boldsymbol{\varepsilon}$ | 3.20 |
| $\mathbf{r}$ | total cumulative data error | 2.2 |
| $\tilde{\mathbf{r}}(t)$ | MT estimate of measurement error | 2.33 |
| $S$ | Length of MT averaging window | 2.32 |
| $S$ | maximum time lag | 3.10, 4.3, 5.5 |
| $T$ | number of time steps | 3.21 |
| $\mathbf{u}$ | system error (white noise forcing for GCM error | 2.17 |
| $\tilde{\mathbf{u}}(t)$ | MT estimate of the model and measurement error | 2.32 |
| $\mathbf{v}(t)$ | innovations | 2.72 |
| $\mathbf{w}_{\mathrm{GCM}}$ | GCM forcing | 2.1 |
| $\mathbf{Y}$ | Covariance of the measurements | 2.58 |
| $\mathbf{Y}_k$ | lag-covariance of the measurements | 2.57 |
| $\mathbf{y}$ | GCM–data residual | 2.16 |
| $\mathbf{0}$; $\mathbf{0}_{128}$ | zero matrix; subscript indicates size | 2.7, 2.70 |

Table A.1: Continued.

| Abbreviation | Definition |
|---|---|
| ATOC | Acoustic Tomography Ocean Experiment |
| BFC97 | (Blanchet et al. 1997) |
| CMIA | Covariance matching with innovations approach |
| CMOA | Covariance matching with observations approach |
| DOF | Degrees of freedom |
| ENSO | El-Niño Southern Oscillation |
| EOF | empirical orthogonal functions |
| F98 | Fukumori et al. (1999) |
| FU93 | Fu et al. (1993) |
| GFDL | Geophysical Fluid Dynamics Laboratory |
| GCM | General Circulation Model |
| LHS | Left Hand Side |
| KF | Kalman filter |
| LDEO | Lamont Doherty Earth Observatory |
| MLF | maximum likelihood function |
| MT (method) | Myers and Tapley (method) |
| RHS | Right Hand Side |
| RMS | root-mean square |
| SKF | simplified Kalman filter |
| T/P | TOPEX/POSEIDON |

Table A.2: Summary of abbreviations.

# Appendix B

# Analytics for the MT Method with a Scalar Model.

The value to which the adaptive algorithm converges is given by the solution of the equation

$$g(q; A, \hat{A}, \hat{q}, \gamma) = 0 \tag{B.1}$$

$$g(q; A, \hat{A}, \hat{q}, \gamma) \equiv \tilde{q}(q; A, \hat{A}, \hat{q}, \gamma) - q$$

This results holds in general for any stable model when $A = \hat{A}$ and $\gamma = 1$. To show this we need to prove that a) there is a unique solution of the equation (B.1); b) it is a stable solution, i.e. the estimates converges to this solution under successive application of the algorithm; c) and that a true value $\hat{q}$ is a solution.

To show that the first condition is satisfied one would need to prove that the function $g(q)$ is strictly monotonic and takes both negative and positive values. This is indeed the case as seen in Figure B.1.

The second condition is satisfied when

$$g(q) > 0, q < q_0,$$

$$g(q) < 0, q > q_0, \quad \text{where } g(q_0) = 0.$$

$$\tag{B.2}$$

Figure B.1: Graph of $g(q)$, $\hat{\mathbf{A}} = \mathbf{A} = 0.9$.
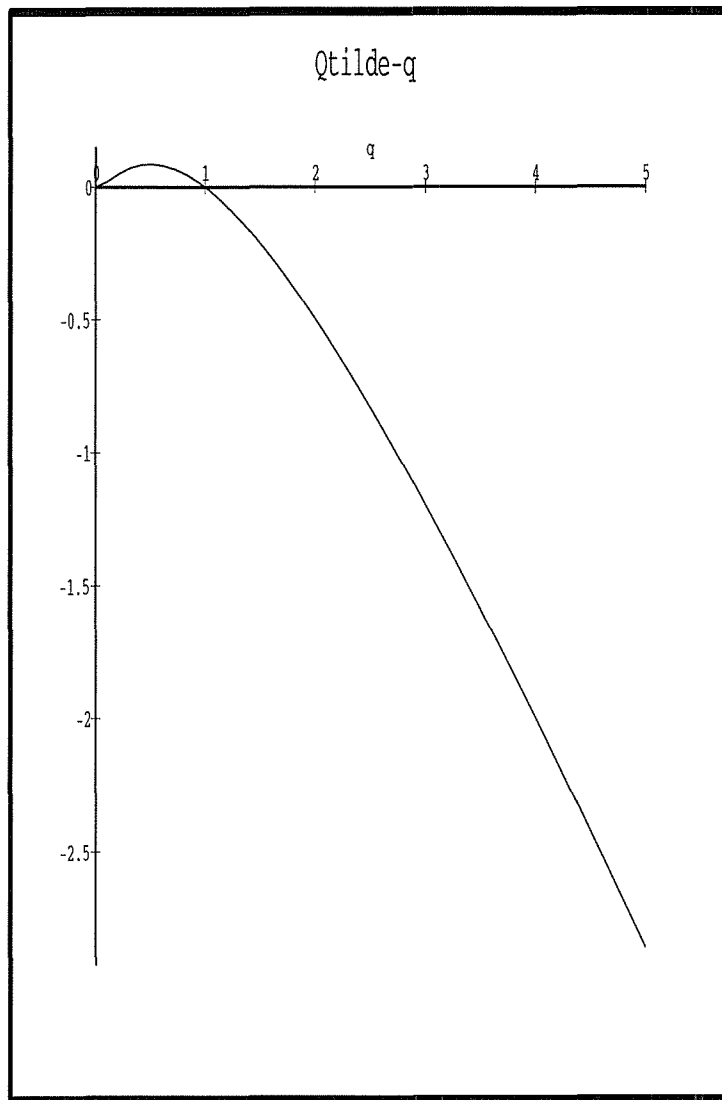
To show that the last condition is satisfied we show 3 dimensional plot of

$$g(\hat{q}; \hat{A}, \hat{A}, \hat{q}, 1) - \hat{q} \qquad\qquad (B.3)$$

for $\hat{A}$ between -1 and 1, and $\hat{q}$ between 0 and 10, see Figure B.2. We see that the expression in equation (B.3) is identically equal to zero, which proves that $\hat{q}$ is indeed a solution of (B.1). Note that because we are using a truncation retaining only low-lag correlations, cf. equation (2.50), we get non-zero values when the model is close to neutrally stable. To check that this is indeed the case we produced an analogous plot retaining higher lag correlations, up to order 10, and the maximum values diminished by 2 orders of magnitude.

Figure B.2: A plot of $g(\hat{q}; \hat{\mathbf{A}}, \hat{\mathbf{A}}, \hat{q}, 1) - \hat{q}$ for $\hat{\mathbf{A}}$ between -1 and 1, and $\hat{q}$ between 0 and 5.



Qtilde(qtrue,Atrue,qtrue,Atrue,1)-qtrue
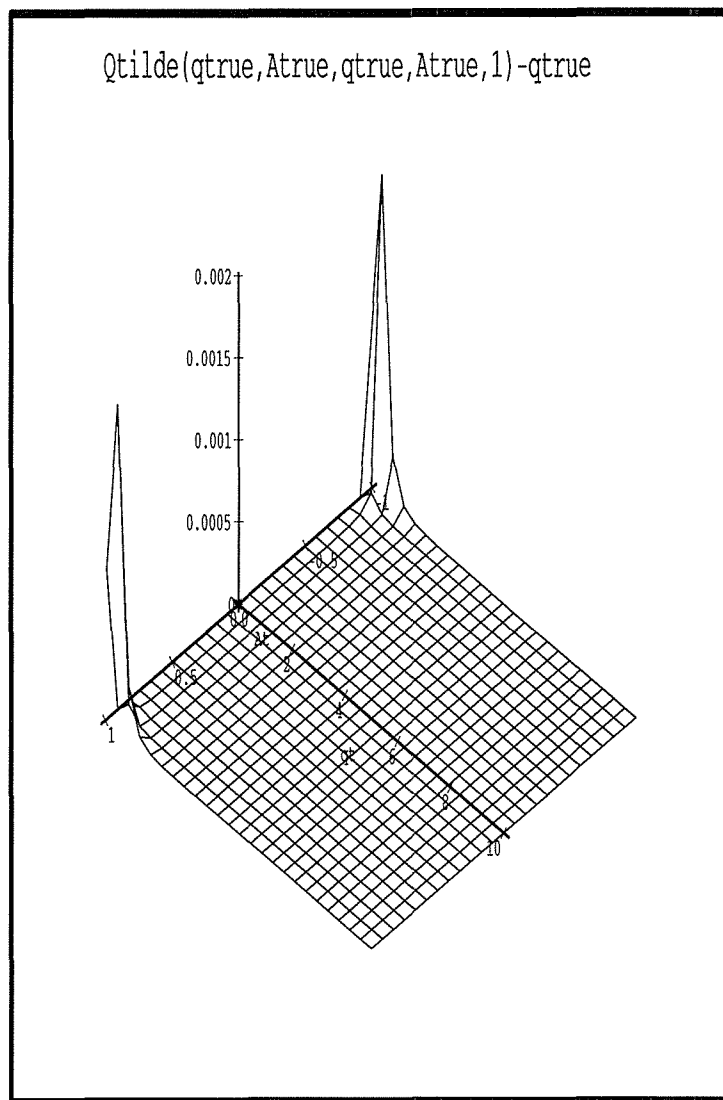
# Appendix C

# Time-Asymptotic Approximation of the Kalman Filter.

Here we present a "doubling algorithm", which provides a recursive solution to the time-asymptotic Kalman filter. The discussion is based on the paper of Fukumori et al. (1993).

The "doubling algorithm" allows one to compute $\Pi(2t|2t-1)$ from $\Pi(t|t-1)$ and can be written as:

$$
\begin{aligned}
\Phi(k+1) &= \Phi(k)\left[\mathbf{I} + \Psi(k)\Theta(k)\right]^{-1}\Phi(k), \\
\Psi(k+1) &= \Psi(k) + \Phi(k)\left[\mathbf{I} + \Psi(k)\Theta(k)\right]^{-1}\Psi(k)\Phi(k)^{\mathrm{T}}, \\
\Theta(k+1) &= \Theta(k) + \Phi(k)^{\mathrm{T}}\Theta(k)\left[\mathbf{I} + \Psi(k)\Theta(k)\right]^{-1}\Phi(k),
\end{aligned}
$$

where the recursion is started from

$$
\Phi(1) = \mathbf{A}^{\mathrm{T}}, \ \Psi(1) = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}, \ \Theta = \mathbf{Q},
$$

and

$$
\Pi(2t|2t-1) = \Theta(k).
$$

One pass of the algorithm requires eight matrix multiplications of the dimension of the state $N$, but the algorithm steps in power of two. The computational savings by the dou-

bling over the full Kalman filter (Section 2.4) which requires two matrix multiplications for one time step is exponential.

# Appendix D

# The Fu et al. (1993) Approach

The covariance matching approach of Fu et al. (1993) is derived from (2.5), (2.15), and (2.16):

$$\zeta_{\text{GCM}}(t) \;=\; \hat{\zeta}_{\text{GCM}}(t) - \mathbf{B}\mathbf{p}(t) - \boldsymbol{\epsilon}(t), \tag{D.1}$$

$$\zeta_{\text{GCM,r}}(t) \equiv \mathbf{B}^{*}\zeta_{\text{GCM}}(t) \;=\; \mathbf{B}^{*}\hat{\zeta}_{\text{GCM}} + \mathbf{p}(t), \tag{D.2}$$

$$\boldsymbol{\eta}_{\text{ocean}}(t) \;=\; \mathbf{E}\hat{\zeta}_{\text{GCM}}(t) + \boldsymbol{\nu}(t), \tag{D.3}$$

where $\boldsymbol{\epsilon}(t)$ and $\mathbf{p}(t)$ are small and large scale GCM errors and $\boldsymbol{\nu}(t)$ are measurement errors, respectively.

In addition, we split the true state and the GCM state into reduced space, or large scale (r), and null space, or fine scale (n), components,

$$\hat{\zeta}_{\text{GCM}}(t) \;=\; \mathbf{B}\hat{\zeta}_{\text{GCM,r}}(t) + \hat{\zeta}_{\text{GCM,n}}(t). \tag{D.4}$$

Substituting this into equations ( D.2-D.3), multiplying each expression by its transpose and taking the expectations yields

$$\text{cov}\,\zeta_{\text{GCM,r}} = \text{cov}\,\hat{\zeta}_{\text{GCM,r}} + \mathbf{P} + \langle \hat{\zeta}_{\text{GCM,r}}\,\mathbf{p}^{\text{T}}\rangle + \langle \mathbf{p}\hat{\zeta}_{\text{GCM,r}}^{\text{T}}\rangle\,, \tag{D.5}$$

$$\text{cov}\,\boldsymbol{\eta}_{\text{ocean}} = \mathbf{H}\,\text{cov}\,\hat{\zeta}_{\text{GCM,r}}\,\mathbf{H}^{\text{T}} + \mathbf{E}\,\text{cov}\,\hat{\zeta}_{\text{GCM,n}}\,\mathbf{E}^{\text{T}} + \text{cov}\,\boldsymbol{\nu}, \mathbf{H} = \mathbf{E}\mathbf{B}. \tag{D.6}$$

Next, we compute covariances of the the residuals (model data differences) on the GCM

grid assuming that all terms are uncorrelated from each other:

$$\text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{E}\boldsymbol{\zeta}_{\text{GCM}}\right) = \mathbf{H}\,\mathbf{P}\,\mathbf{H}^{\text{T}} + \mathbf{R}, \mathbf{R} = \text{cov}\left(\mathbf{E}\boldsymbol{\epsilon} + \boldsymbol{\nu}\right), \qquad (\text{D.7})$$

where the covariance of the data error of the reduced state, $\mathbf{R}$, includes the measurement error and the null space errors. The covariance of the residual on the coarse reduced space grid, is given by

$$\text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM,r}}\right) = \mathbf{E}\,\text{cov}\,\hat{\boldsymbol{\zeta}}_{\text{GCM,n}}\mathbf{E}^{\text{T}} + \mathbf{H}\mathbf{P}\mathbf{H}^{\text{T}} + \text{cov}\,\boldsymbol{\nu}, \qquad (\text{D.8})$$

Assuming that $\langle \mathbf{p}(t)\,\hat{\boldsymbol{\zeta}}_{\text{GCM,r}}(t)^{\text{T}}\rangle = \mathbf{0}$, as in F99, we manipulate linear equations (D.5-D.8) in five unknowns $\mathbf{E}\text{cov}\,\hat{\boldsymbol{\zeta}}_{\text{GCM,n}}\mathbf{E}^{\text{T}}$, $\mathbf{H}\text{cov}\,\hat{\boldsymbol{\zeta}}_{\text{GCM,r}}\mathbf{H}^{\text{T}}$, $\mathbf{H}\mathbf{P}\mathbf{H}^{\text{T}}$, $\mathbf{E}\text{cov}\,\boldsymbol{\epsilon}\mathbf{E}^{\text{T}}$, and $\text{cov}\,\boldsymbol{\nu}$, to obtain:

$$\mathbf{H}\mathbf{P}\mathbf{H}^{\text{T}} = \frac{1}{2}\left(\text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM,r}}\right) - \text{cov}\,\boldsymbol{\eta}_{\text{ocean}} + \mathbf{H}\,\text{cov}\,\boldsymbol{\zeta}_{\text{GCM,r}}\,\mathbf{H}^{\text{T}}\right) \qquad (\text{D.9})$$

$$\mathbf{R} = \text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{E}\boldsymbol{\zeta}_{\text{GCM}}\right) - \qquad\qquad\qquad\qquad\qquad (\text{D.10})$$
$$\frac{1}{2}\left(\text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM,r}}\right) - \text{cov}\,\boldsymbol{\eta}_{\text{ocean}} + \mathbf{H}\,\text{cov}\,\boldsymbol{\zeta}_{\text{GCM,r}}\,\mathbf{H}^{\text{T}}\right).$$

Because the CMA does not require $\langle \hat{\boldsymbol{\zeta}}_{\text{GCM,r}}\,\mathbf{p}^{\text{T}}\rangle\mathbf{B}^{\text{T}} = \mathbf{0}$, it is possible to evaluate the validity of this assumption:

$$\mathbf{H}\left(\langle\hat{\boldsymbol{\zeta}}_{\text{GCM,r}}\,\mathbf{p}^{\text{T}}\rangle + \langle\mathbf{p}\hat{\boldsymbol{\zeta}}_{\text{GCM,r}}^{\text{T}}\rangle\right)\mathbf{H}^{\text{T}} \approx \text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM,r}}\right) \qquad (\text{D.11})$$
$$+\mathbf{H}\text{cov}\,\boldsymbol{\zeta}_{\text{GCM,r}}\mathbf{H}^{\text{T}} + \mathbf{R} - \text{cov}\,\boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\,\mathbf{P}\,\mathbf{H}^{\text{T}} - \text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{E}\boldsymbol{\zeta}_{\text{GCM}}\right).$$

To compute a correlation coefficient between $\mathbf{H}\mathbf{p}$ and $\mathbf{H}\hat{\boldsymbol{\zeta}}_{\text{GCM}}$ we multiply both sides of equation (D.11) by $\text{diag}\left(\mathbf{H}\,\mathbf{P}\,\mathbf{H}^{\text{T}}\right)^{-1/2}$ on the left and $\text{diag}\left(\mathbf{H}\,\text{cov}\,\hat{\boldsymbol{\zeta}}_{\text{GCM,r}}\,\mathbf{H}^{\text{T}}\right)^{-1/2}$ on the right to obtain:

$$\text{diag}\left[\text{corr}\left(\mathbf{H}\,\mathbf{p},\,\mathbf{H}\,\hat{\boldsymbol{\zeta}}_{\text{GCM,r}}^{\text{T}}\right)\right] \approx \frac{1}{2}\,\text{diag}\left[\left(\mathbf{H}\,\mathbf{P}\,\mathbf{H}^{\text{T}}\right)^{-1/2}\left(\text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM,r}}\right)\right.\right.$$
$$\left. + \mathbf{H}\text{cov}\,\boldsymbol{\zeta}_{\text{GCM,r}}\mathbf{H}^{\text{T}} + \mathbf{R} - \text{cov}\,\boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\,\mathbf{P}\,\mathbf{H}^{\text{T}} - \text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{E}\boldsymbol{\zeta}_{\text{GCM}}\right)\right) \qquad (\text{D.12})$$
$$\left. \left(\text{cov}\,\boldsymbol{\eta}_{\text{ocean}} - \mathbf{R} - \text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{H}\boldsymbol{\zeta}_{\text{GCM,r}}\right) + \text{cov}\left(\boldsymbol{\eta}_{\text{ocean}} - \mathbf{E}\boldsymbol{\zeta}_{\text{GCM}}\right)\right)^{-1/2}\right].$$

The correlation coefficient given above can be greater than one by magnitude. This implies that either the terms neglected in (D.5-D.8) are significantly different from zero, or the estimates of the model and data error covariances are wrong.

The original paper of Fu et al. (1993) considered only the case of a full state model. In this case the null space of the reduction operator $\mathbf{B}^*$ vanishes and the equations are simplified, since $\zeta_{\mathrm{GCM}} \equiv \zeta_{\mathrm{GCM,r}}$, see equations (4.9) and (4.8).

# Appendix E

# Null Space of the Operator $\mathbf{H}\left(\mathbf{A}^l \cdot + \cdot (\mathbf{A}^l)^{\mathrm{T}}\right)\mathbf{H}^{\mathrm{T}}.$

Now, we prove that the dimension of the null space of $\mathbf{H}\left(\mathbf{A}^l \cdot + \cdot (\mathbf{A}^l)^{\mathrm{T}}\right)\mathbf{H}^{\mathrm{T}}$ is $(N - M)(N - M + 1)/2$. By performing singular value decomposition of $\mathbf{H}$, i.e.

$$\mathbf{H} = \mathbf{U}\tilde{\mathbf{H}}\mathbf{V}^{\mathrm{T}}, \tag{E.1}$$

$$\text{where } \tilde{\mathbf{H}} = \begin{bmatrix} \lambda_1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \lambda_M & 0 & \cdots & 0 \end{bmatrix}, \tag{E.2}$$

we can rewrite

$$\mathbf{F}_{l,k} = (\mathbf{U}\tilde{\mathbf{H}})(\tilde{\mathbf{A}}^l \tilde{\mathbf{S}}_k + \tilde{\mathbf{S}}_k (\tilde{\mathbf{A}}^l)^{\mathrm{T}})(\mathbf{U}\tilde{\mathbf{H}})^{\mathrm{T}}, \tag{E.3}$$

$$\text{where } \tilde{\mathbf{A}} = \mathbf{V}\tilde{\mathbf{A}}\mathbf{V}^{\mathrm{T}}, \tilde{\mathbf{S}}_k = \mathbf{V}\tilde{\mathbf{S}}_k\mathbf{V}^{\mathrm{T}} \tag{E.4}$$

Note, that the transformation $\mathbf{V} \cdot \mathbf{V}^{\mathrm{T}}$ represents rotation, and; therefore, in general, does not change any properties of either $\mathbf{A}$ or $\mathbf{S}_k$. Therefore, we can drop the tildes on $\mathbf{A}$ and $\mathbf{S}_k$. We can take $\tilde{\mathbf{S}}_k$ to be symmetric delta matrices, defined as

$$\mathbf{S}_{k_1,k_2} : s_{i,j} = 0, \text{for all} \quad i, j, \quad \text{except for} \quad s_{k_1,k_2} = s_{k_2,k_1} = 1, \tag{E.5}$$

since we can the rank of $\mathcal{F}$ does not change under linear operations on the columns, and since

$$c_1 \mathbf{F}_{l,i} + c_2 \mathbf{F}_{l,j} = \tilde{\mathbf{H}} \left( \mathbf{A}^l (c_1 \mathbf{S}_i + c_2 \mathbf{S}_j) + (c_1 \mathbf{S}_i + c_2 \mathbf{S}_j)(\mathbf{A}^l)^{\mathrm{T}} \right) \tilde{\mathbf{H}}^{\mathrm{T}}. \qquad (E.6)$$

The null space of the operator $\tilde{\mathbf{H}}(\cdot)\tilde{\mathbf{H}}^{\mathrm{T}}$ is spanned by all matrices which have zeros in the upper $M$ by $M$ corner. We note that if we apply $\mathbf{A}^l \cdot + \cdot (\mathbf{A}^l)^{\mathrm{T}}$ to $\mathbf{S}_{k_1,k_2}$ with $k_1 > M, k_2 > M$, and only those $\mathbf{S}_{k_1,k_2}$, we get matrices which have upper right corner of size $M$ identically equal to 0. Therefore all $\mathbf{F}_{l,k}$ matrices corresponding to $\mathbf{S}_{k_1,k_2}$ with $k_1 > M, k_2 > M$, and only those $\mathbf{S}_{k_1,k_2}$, are identically equal to zero. There are $(N - M)(N - M + 1)/2$ of them, and they span the null space of the operator $\mathbf{H} \left( \mathbf{A}^l \cdot + \cdot (\mathbf{A}^l)^{\mathrm{T}} \right) \mathbf{H}^{\mathrm{T}}$.

# Appendix F

# Covariance of a Covariance Toolbox.

The toolbox has been written for the covariance matching algorithm, but it implements the algorithm described above for the most general case. The software is written in Matlab, but most computations are done through MEX interface, both C and Fortran versions are available. See *README.covcov* for installation instructions. It is available via anonymous FTP to gulf.mit.edu, IP Address 18.83.0.149, from directory pub/misha/.

The front end program *covdiff.m* is written for the covariance matching applications. However, the main callable routine, *covcov.m*, is completely general, and can be modified as appropriate.

The function *covcov.m* computes the covariance of $\mathbf{Y}_{(i,j)}(q)$ and $\mathbf{Y}_{(k,l)}(r)$ using equation (3.25) for a vector of indices **ix, jx, kx, lx**. The function allows one to compute uncertainties for all possible combinations $(i, k)$ of pairs $(ix(i), jx(i)$ and $kx(k), lx(k)$, by setting ReFlag to 1. In addition, the user can specify the true covariance matrices (if known) instead of using the sample estimates. This is useful for testing the validity of the method, see example *demo-covcov.m*. Additional options are available, see help for the routine.

| YFlag | $\varepsilon$ includes | ReFlag | What part of $R_\varepsilon$ is computed | $N_\mathcal{R}$ |
|-------|------------------------|--------|------------------------------------------|-----------------|
| 0 | full covariance | 0 | diagonal | $Nlags \cdot nY \cdot (nY + 1)/2$ |
| 1 | diagonal of the covariance | 0 | diagonal | $(nY \cdot Nlags)$ |
| 0 | full covariance | 1 | full | $(nY \cdot Nlags) \cdot (nY \cdot Nlags + 1)/2$ |
| 1 | diagonal of the covariance | 1 | full | $nY \cdot Nlags \cdot (Nlags + 1)/2$ |

Table F.1: Number of elements of the covariance matrix R, depending on the inputs into covdiff.m or covcov.m (in which case $Nlags = 1$). $nY$ no. of elements in observation vector $\mathbf{y}(t)$.

# F.1  Computational time.

The routines *covcov.m* and *covdiff.m* can be very slow, since the number of "do" loops is of the order

$$8\, N_\mathcal{R} \cdot (T \cdot \text{MaxLag})^2 \tag{F.1}$$

where $N_\mathcal{R}$ denotes the number of elements of covariance $\mathcal{R}_T = \text{cov}\left[\mathbf{Y}_{(i,j)}(q)\, \mathbf{Y}_{(k,l)}(r)\right]$, see equation (3.25), being estimated. There are four possible cases for the number of elements of $\mathcal{R}_T$:  In addition, if the maximum significant lag (MaxLag) is much smaller than the length of the time series, one can use the following approximation:

$$\mathcal{R}_T = \frac{30 \cdot MaxLag}{T} \mathcal{R}_{30 \cdot MaxLag} \tag{F.2}$$

# References

Abramson, P. J., 1968: Simultaneous estimation of the state and noise statistics in linear dynamic systems, Ph.D. thesis, Mass. Inst. of Tech.

Anderson, B. D., and J. B. Moore, 1979: *Optimal Filtering*, Prentice-Hall, Englewood Cliffs.

Anderson, T. W., 1971: *The statistical analysis of time series.*, John Wiley and Sons.

Belanger, P. R., 1974: Estimation of noise covariance matrices for a linear time-varying stochastic process., *Automatica*, **10**, 267–275.

Blanchet, I., 1997: Assimilation de donnees a l'aide d'un filtre de Kalman adaptatif pour un modele d'ocean, Ph.D. thesis, L'Universite Pierre et Marie Curie (Paris VI).

Blanchet, I., C. Frankignoul and M. Cane, 1997: A Comparison of Adaptive Kalman Filters for a Tropical Pacific Ocean Model., *Mon. Weather Rev.*, **125**, 40–58.

Boning, C., W. Holland, F.O.Bryan, G. Danabasoglu and J. McWilliams, 1995: An overlooked problem in the model circulation of the thermohaline circulation and heat transport in the Atalantic ocean., *J. Climate*, **8**, 515–523.

Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. Hackert and A. J. Busallachi, 1996: Mapping tropical Pacific sea level: data assimilation via a reduced state space Kalman filter., *J. Geophys. Res.*, pp. 22,599–22,617.

Chen, D., M. A. Cane, S. Zebiak and A. Kaplan, 1998: The impact of sea level data assimilation on the Lamont model prediction of the 1997/1998 El Nino, *Geophys. Res. Lett*, Submitted.

Cohn, S. E., and D.F.Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions., *Mon. Weather Rev.*, **1190**, 1757–1785.

Daley, R., 1992a: The effect of serially correlatied observation and model error on atmospheric data assimilation., *Mon. Weather Rev.*, **120**, 165–178.

Daley, R., 1992b: The lagged innovation covariance: a performance diagnostic for atmospheric data assimilation., *Mon. Weather Rev.*, **120**, 178–196.

Dee, D. P., 1991: Simplification of the Kalman filter for meteorological data assimilation., *Quart. J. Roy. Meteor. Soc.*, **117**, 365–384.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation., *Mon. Weather Rev.*, **123**, 1128–1145.

Dee, D. P., S. Cohn, A. Dalcher and M. Ghil, 1985: An efficient algorithm for estimating noise covariances in distributed systems, *I.E.E.E. Trans on Autom. Contr.*, **30**, 1057–1065.

Dee, D. P., and A. M. DaSilva, 1997: Data assimilation in the presence of forecast bias., *Quart. J. Roy. Meteor. Soc.*, **124**, 269–295.

Evensen, G., D. Dee and J. Schroter, 1998: Parameter estimation in dynamical models., in *NATO ASI, Ocean Modeling and Parameterizations*, edited by E. P. Chassignet, and J. Verron.

Fu, L.-L., I. Fukumori and R. Muller, 1993: Fitting dynamics to the GEOSAT sea level observations in the tropical pacific ocean. Part II: a linear wind-driven model., *J. Phys. Oceanogr.*, **23**, 798–809.

Fukumori, I., J. Benveniste, C. Wunsch and D. Haidvogel, 1993: Assimilation of sea surface topography in to an ocean general circulation model using a steady-state smoother., *J. Phys. Oceanogr.*, **23**, 1831–55.

Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation., *J. Geophys. Res.*, **100**, 6777–6793.

Fukumori, I., R. Raghunath and L. Fu, 1998: Nature of global large-scale sea level variability in relation to atmospheric forcing: A modelling study., *J. Geophys. Res.*, **103**, C3, 5493–5512.

Fukumori, I., R. Raghunath, L. Fu and Y. Chao, 1999: Assimilation of TOPEX/POSEIDON altimeter data into a global ocean circulation model: are the results any good?, *J. Geophys. Res.*, submitted.

Gajić, Z., and M. T. J. Qureshi, 1995: *Lyapunov Matrix Equation in System Stability and Control*, Academic, San Diego, Calif.

Gates, W. (ed.), 1995: *Proceedings of the first international AMIP scientific conference.*

Gelb, A., 1979: *Applied Optimal Estimation*, MIT Press, Cambrdige, Mass.

Gill, A., and P. P. Niiler, 1973: The theory of the seasonal variability in the ocean, *Deep Sea Res.*, **20**, 141–177.

Gill, A. E., 1982: *Atmosphere-Ocean Dynamics*, vol. 30 of *International Geophysics Series*, Academic, San Diego, Calif.

Groutage, F. D., R. G. Jacquot and R. Kirlin, 1987: Techniques for adaptive state estimation through the utilization of robust smoothing., in *Control and Dynamic System*, vol. 23, pp. 273–308.

Hogg, N., 1996: Oceanographic data for parameter estimation., in *Modern approaches to data assimilation in ocean modelling.*, edited by P. Malanotte-Rizzoli., Elsevier Science, B.V.

Holland, W. R., and P. Malanotte-Rizzoli, 1989: Assimilation of Altimeter Data into an Oceanic Circulation Model: Space versus Time Resolution Studies, *J. Phys. Oceanogr.*, **19**, 1507–1534.

Jazwinsky, A. H., 1969: Adaptive filtering., *Automatica*, **5**, 475–485.

Kailath, T., 1968: An innovations control approach to least square estimation - Part I: linear filtering in additive white noise., *I.E.E.E. Trans on Autom. Contr.*, **12**, 646–655.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems., *J. Basic. Eng., ASME, Series D*, **82**(1), 35–45.

Kerr, R., 1998: Climate prediction - Models win big in forecasting El Nino, *Science*, **280**(5363), 522–523.

Large, W. G., J. McWilliams and S. Doney, 1994: Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization, *Reviews of Geophysics*, **32(4)**, 363–403.

Lermusiaux, P. F. J., 1997: Error subspace data assimilation for ocean field estimation: theory, validation, and applications., Ph.D. thesis, Harvard Univ.

Levitus, S., 1982: Climatological Atlas of the World Ocean, *Professional Paper 13*, , National Oceanic and Atmospheric Administration.

Macdonald, A. R., and C. Wunsch, 1996: An Estimate of the Global Ocean Circulation and Heat Flux, *Nature*, **382**, 436–439.

Malanotte-Rizzoli, P., and J. Levin, 1999: TOPEX/POSEIDON altimetry assimilation in the North Atlantic Ocean with a suboptimal Kalman filter, in preparation.

Malanotte-Rizzoli, P., and E. Tziperman, 1996: The oceanographic data assimilation problem:

overview, motivation and purposes., in *Modern approaches to data assimilation in ocean modelling.*, edited by P. Malanotte-Rizzoli., Elsevier Science, B.V.

Mardia, K. V., J. T. Kent and J. M. Bibby, 1979: *Multivariate Analysis*, Academic, San Diego, Calif.

Marshall, J., C. Hill, L. Perelman and A. Adcroft, 1997a: A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers, *J. Geophys. Res.*, **102**(C3), 5753–5766.

Marshall, J., C. Hill, L. Perelman and A. Adcroft, 1997b: Hydrostatic, quasi-hydrostatic and non-hydrostatic ocean modeling., *J. Geophys. Res.*, **102**(C3), 5733–5752.

Maybeck, P. S., 1982: *Stochastic models, estimation, and control.* , vol. 2, Academic Press, N.Y.

Mehra, R. K., 1970: On the identification of variances and adaptive Kalman filtering., *I.E.E.E. Trans on Autom. Contr.*, **AC-15**, 175–184.

Menemenlis, D., and M. Chechelnitsky, in print, 1999: Error estimates for an ocean general circulation model from altimeter and acoustic tomography data, *Mon. Weather Rev.*

Menemenlis, D. T., P. Fieguth, C. Wunsch and A. Wilsky, 1997a: Adaptation of a fast optimal interpolation algorithm to the mapping of oceanographic data, *J. Geophys. Res.*, **102**(C5), 10,573–10,584.

Menemenlis, D. T., T. Webb, C. Wunsch, U. Send and C. Hill, 1997b: Basin-scale ocean circulation from combined altimetric, tomographic and model data, *Nature*, **318**, 618–621.

Menemenlis, D., and C. Wunsch, 1997: Linearization of an Oceanic Circulation Model for Data Assimilation and Climate Studies, *J. Atmos. Oceanic Technol.*, **14**(6), 1420–1443.

Menke, W., 1989: *Geophysical Data Analysis: Discrete Inverse Theory*, vol. 45 of *International Geophysics Series*, Academic, San Diego, Calif., revised edition.

Miller, R. N., and M. A. Cane, 1989: A Kalman filter analysis of sea level height in the tropical Pacific., *J. Phys. Oceanogr.*, **19**, 773–790.

Moghaddamjoo, A. R., and R. L. Kirlin, 1993: Robust adaptive Kalman filtering, in *Approximate Kalman Filtering*, edited by G. Chen, pp. 65–85, World Scientific.

Munk, W., P. Worcester and C. Wunsch, 1995: *Ocean Acoustic Tomography, Cambridge Monographs on Mechanics*, Cambridge Univ. Press, New York.

Myers, K. A., 1974: Filtering theory methods and applications to the orbit determination problem for near-earth satellites., Ph.D. thesis, The Univ. of Texas, Austin.

Myers, M. A., and B. D. Tapley, 1976: Adaptive sequential estimation with unknown noise statistics, *I.E.E.E. Trans on Autom. Contr.*, **21**, 520–523.

Oberhuber, J. M., 1988: An atlas Based on the 'COADS' Data Set: The Budgets of Heat, Buoyancy and Turbulent Kinetic Energy at the Surface of the Global Ocean, *Report 15*, , Max-Plank-Institute for Meteorology, 2000 Hamburg 13, Bundesstrasse 55, FRG.

Pacanowksi, R., K. Dixon and A. Rosati, 1991: Modular Ocean Model User's Guide, Climatological Atlas of the World Ocean, *Ocean Group Tech. Rep. 2*, , GFDL, National Oceanic and Atmospheric Administration.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics.*, Springer-Verlag, 2nd edition.

Rauch, H. E., F. Tung and C. T. Striebel, 1965: Maximum Likelihood Estimates of Linear Dynamic Systems, *AIAA J.*, **3**, 1445–1450.

Sage, A. P., and G. Husa, 1969: Algorithms for sequential adaptive estimation of prior statistics., in *Proc. 8th IEEE Symp. on Adapt. Proc., Penn. State Univ, University Park*.

Shellenbarger, J., 1966: Estimation of Covariance Parameters for an Adaptive Kalman Filter, in , pp. 698–702., National Electronics Conference.

Shellenbarger, J., 1967: A Multivariate Learning Technique for Improved Dynamic System Performance, in , pp. 146–151., National Electronics Conference.

Stammer, D., R. Tokmakian, A. Semtner and C. Wunsch, 1996: How Well Does a 1/4° Global Circulation Model Simulate Large-Scale Oceanic Observations?, *J. Geophys. Res.*, **101**(C11), 25,779–25,811.

Stammer, D., and C. Wunsch, 1996: The Determination of the the Large-Scale Circulation of the Pacific Ocean from Satellite Altimetry Using Model Green's Functions, *J. Geophys. Res.*, **101**(C8), 18,409–18,432.

Stammer, D., C. Wunsch, R. Giering, Q. Zhang, J. Marotzke, J. Marshall and C. Hill, 1997:

The global ocean circulation estimated from TOPEX/POSEIDON altimetry and the MIT general circulation model, *Report 49*, , Massachusetts Institute of Technology, Cambridge.

The ATOC Consortium, 1998: Ocean Climate Change: Comparison of Acoustic Tomography, Satellite Altimetry, and Modeling., *Science*, **281**, 1327–1332.

Toddling, R., and S. E. Cohn, 1995: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter, *Mon. Weather Rev.*

Trenberth, K. E., J. G. Olson and W. G. Large, 1989: A Global Ocean Wind Stress Climatology Based on ECMWF Analyses, *Tech. Note NCAR/TN-338+STR*, , NCAR.

Visbeck, M., J. Marshall and H. Jones, 1996: Dynamics of isolated convective regions in the ocean, *J. Phys. Oceanogr.*, **26(9)**, 1721–1734.

Wergen, W., 1992: The effect of model errors in variational assimilation, *TELLUS*, **44A**:((4)), 297–313.

Wunsch, C., 1996: *The Ocean Circulation Inverse Problem*, Cambridge Univ. Press, New York.

Zebiak, S. E., and M. A. Cane, 1987: A model El Nino-Southern Oscillation, *Mon. Weather Rev.*, (115), 2262–2278.