

## INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

# U·M·I

University Microfilms International  
A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
313/761-4700 800/521-0600



Order Number 9002938

**Data assimilation and parameter estimation in oceanographic  
models**

Smedstad, Ole Martin, Ph.D.

The Florida State University, 1989

**U·M·I**  
300 N. Zeeb Rd.  
Ann Arbor, MI 48106



THE FLORIDA STATE UNIVERSITY  
COLLEGE OF ARTS AND SCIENCES

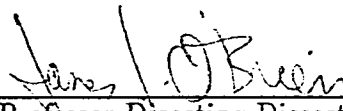
DATA ASSIMILATION AND PARAMETER ESTIMATION  
IN OCEANOGRAPHIC MODELS

by

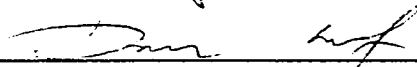
OLE MARTIN SMEDSTAD


A Dissertation submitted to the  
Program in Geophysical Fluid Dynamics  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Approved:

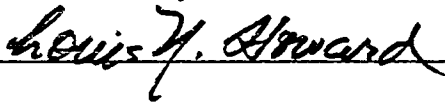


Professor Directing Dissertation









Summer Semester, 1989

## Abstract

A variational data assimilation method for a reduced gravity model is developed. The method is applied to the equatorial Pacific Ocean. In the variational formalism a cost function measuring the “distance” between the model solution and the observations is minimized. The phase speed of the model is used as a control parameter and the optimal spatial structure giving the best fit of the model to the observations is determined. In the minimization algorithm a conjugate gradient descent direction is used. The method is computationally effective, and for the experiments considered convergence is achieved in ten iterations or less.

Several experiments are performed using the model solutions as observations. It is shown that the assimilation algorithm is able to determine the spatial structure of the phase speed, even if observations are available at only three stations. The estimated phase speed is not sensitive to errors in the observations, and the algorithm gives a unique solution to the problem.

Real sea level observations from three stations are assimilated for two different periods. The year 1979 was chosen to represent a year without an El Nino, while 1982/83 was chosen to represent an El Nino year. For 1979 the assimilation gave a phase speed with higher values in the west and lower values in the east compared to the initial guess of a constant phase speed. Assimilation of observations in 1982/83 gave the opposite picture, with lower values in the west and higher values in the east. This result is consistent with observations. The phase speed is

proportional to the depth of the thermocline, and during normal conditions the basic stratification consists of a deep thermocline in the west and a shallow thermocline in the east. During an El Nino the picture is reversed. Calculating the correlation coefficient between the model results and the observations shows that the correlation increased for all the stations during the assimilation, even at stations which were not a part of the assimilation.

## Acknowledgement

This work was supported by NASA Oceanic Processes Branch and in part by the Florida State University through time granted on its Cyber 205 supercomputer.

I wish to express my deep gratitude to Dr. James J. O'Brien for his ideas and support as my major professor. His motivation and encouragement during the course of this research are sincerely appreciated. I am also grateful for the time taken by Drs. S. Blumsack, L. Howard, I. M. Navon and D. Nof while serving on my doctoral committee.

I would like to thank my colleagues at the Mesoscale Air-Sea Interaction Group for their comments and suggestions. Special thanks go to Tommy Jensen and Brian Kelly for many helpful discussions.

Assistance in the form of facilities and technical support were provided by the Supercomputer Computations Research Institute (SCRI) at Florida State University. I wish to thank Mimi Burbank for her assistance using  $\text{\TeX}$  during the typing of the dissertation.



## TABLE OF CONTENTS

LIST OF FIGURES	viii
<b>1. Introduction</b>	1
<b>2. Variational method</b>	12
2.1 <i>Description of the method</i>	12
2.2 <i>The cost function <math>J</math></i>	16
2.3 <i>Outline of solution procedure</i>	18
2.4 <i>Computational aspects</i>	20
<b>3. Descent methods</b>	22
3.1 <i>Introduction</i>	22
3.2 <i>The conjugate gradient method</i>	23
3.3 <i>Newton, quasi-Newton</i> <i>and limited-memory quasi-Newton methods</i>	28
3.4 <i>Computing the step length</i>	30
<b>4. Parameter estimation</b>	34
<b>5. Specification of the problem</b>	38
5.1 <i>The numerical model</i>	39
5.2 <i>The adjoint equations</i>	40
5.3 <i>The wind forcing</i>	47

5.4	<i>The observations</i>	47
<b>6.</b>	<b>Data assimilation in the Pacific Ocean model</b>	<b>52</b>
6.1	<i>The Pacific Ocean model initialized by a Kelvin wave</i>	53
6.1.1	<i>Experiment 1 : Constant phase speed and observations everywhere in space</i>	56
6.1.2	<i>Verification of the gradient of the cost function</i>	68
6.2	<i>Assimilation in the Pacific Ocean model forced by real winds</i>	70
6.2.1	<i>Experiment 2 : Constant phase speed and observations everywhere in space</i>	70
6.2.2	<i>Experiment 3 : Variable phase speed and observations everywhere in space</i>	73
6.2.3	<i>Experiment 4 : Variable phase speed and observations at 3 stations</i>	75
6.2.4	<i>Uniqueness, identifiability and stability of the solution</i>	81
6.3	<i>Assimilation of real sea level observations in the Pacific Ocean model</i>	82
6.3.1	<i>Experiment 5 : Real sea level observations from 3 stations in 1979</i>	82
6.3.2	<i>Experiment 6 : Real sea level observations from 3 stations in 1982/83</i>	87
<b>7.</b>	<b>Summary and conclusions</b>	<b>93</b>
	<b>References</b>	<b>99</b>

<b>Appendix A : Derivation of the continuous adjoint equations</b>	<b>108</b>
<b>Appendix B : Derivation of the finite difference adjoint equations</b>	<b>114</b>

## LIST OF FIGURES

- Figure 1 : The geometry of the reduced gravity model 41
- Figure 2 : The Pacific Ocean model domain. The southern and the northern boundaries are open. The position of the sea level stations used in the assimilation is also shown. 42
- Figure 3 : The Arakawa C grid 44
- Figure 4 : Pseudostress for the Pacific Ocean model for June 1979 48
- Figure 5 : Time series of upper layer thickness (ULT) for the station at Santa Cruz. ULT in meters is plotted as a function of time. a) Daily observations of ULT. b) The filtered time series using a pass period of 60 days. 50
- Figure 6 : Plot of the upper layer thickness (ULT) and velocity field from the Pacific Ocean model. A phase speed of  $c^2 = 6.0\text{m}^2\text{s}^{-2}$  was used. The initial ULT was

300 meter. Contour interval is 10 meter. The Figure shows the results at a) day 0, b) day 60, c) day 90, d) day 180. 54

Figure 7 : Data misfits during the first iteration for the model gridpoint corresponding to the island station at Santa Cruz. Misfits in meters are shown along the vertical axis, while time in days is plotted along the horizontal axis. 57

Figure 8 : Plot of the Lagrangian multipliers for the case where observations are available everywhere. The Lagrangian multipliers for the h equation is shown using contours, while the Lagrangian multipliers for the U and V equations are represented by vectors. The scale of the vectors are shown in the lower left corner of each figure. Contour interval is 10.0. The Figure shows the results at a) day 150, b) day 120, c) day 60, d) day 0. 58

Figure 9 : Plot of the longitudinal and time variation of the integral in equation (5–13) after the first iteration. Dashed lines represents negative values of the gradient, while solid lines represents positive values. The gradient has been scaled by its maximum value. Contour interval is 0.15. 61

Figure 10 : The case when the gradient of the cost function is not a function of longitude. a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\| \mathbf{g} \|}{\| \mathbf{g}_0 \|}$  is plotted as a function of the number of iterations. c) The value of  $c^2$  is plotted as a function of the number of iterations. 63

Figure 11 : The results for the case when the gradient of the cost function is a function of longitude. a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\| \mathbf{g} \|}{\| \mathbf{g}_0 \|}$  is plotted as a function of the number of iterations. 65

Figure 12 : The results for the case when the gradient of the cost function is a function of longitude. The longitudinal variation of the phase speed during the iterative process, a) after one iteration, b) after two iterations, c) after five iterations and d) after seven iterations. 67

Figure 13 : Plot of the function  $f(\alpha)$  in (6-2). 69

Figure 14 : The results from experiment 2 are shown.

a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function

of the number of iterations. b) The normalized gradient of the cost function  $\frac{\| \mathbf{g} \|}{\| \mathbf{g}_0 \|}$  is plotted as a function of the number of iterations.

71

Figure 15 : The longitudinal variation of the phase speed during the iterative process in experiment 2, a) after one iteration, b) after three iterations, c) after seven iterations.

72

Figure 16 : The longitudinal variation of the phase speed used to create the observations

74

Figure 17 : The results from experiment 3 are shown.

a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\| \mathbf{g} \|}{\| \mathbf{g}_0 \|}$  is plotted as a function of the number of iterations.

76

Figure 18 : The longitudinal variation of the phase speed during the iterative process in experiment 3, a) after one iteration, b) after three iterations, c) after six iterations, d) after ten iterations.

77

Figure 19 : The results from experiment 4 are shown.

a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\| \mathbf{g} \|}{\| \mathbf{g}_0 \|}$  is plotted as a function of the number of iterations.

79

Figure 20 : The longitudinal variation of the phase speed during the iterative process in experiment 4, a) after one iteration, b) after three iterations, c) after six iterations, d) after ten iterations.

80

Figure 21 : The results from experiment 5 are shown.

a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\| \mathbf{g} \|}{\| \mathbf{g}_0 \|}$  is plotted as a function of the number of iterations.

83

Figure 22 : The longitudinal variation of the phase speed during the iterative process in experiment 5, a) after one iteration, b) after three iterations, c) after six iterations.

84

Figure 23 : The correlation coefficient for different stations as a function



of the number of iterations in experiment 5. 88

Figure 24 : The results from experiment 6 are shown.

a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\|g\|}{\|g_0\|}$  is plotted as a function of the number of iterations. 90

Figure 25 : The longitudinal variation of the phase speed during the iterative process in experiment 6, a) after one iteration, b) after three iterations, c) after five iterations. 91

Figure 26 : The correlation coefficient for different stations as a function of the number of iterations in experiment 6. 92

## 1. Introduction

The oceanographers will experience an explosion in the number of observations available in the next decade. Several new observational techniques are being developed. Satellite measurements (altimetry, scatterometry) around 1990 will give the oceanographers a large amount of new data. New techniques for observing the interior of the ocean (such as tomography) will also create a new data set. It will be important to be able to extract as much information as possible from this new data set. The use of numerical models will play a crucial role in this work, and methods for four-dimensional data assimilation have to be developed. From a meteorologist's point of view, data assimilation is the process through which observations distributed in time and space are treated in order to specify the initial conditions of a numerical forecast. The large number of stations taking meteorological observations at synoptic times make the initializing process possible. Even with the new observational techniques in oceanography mentioned earlier, the amount of data will not be large enough for an initialization of the oceanographic models in the same sense as the meteorological models.

Oceanographers have until now had little experience with four-dimensional data assimilation. Meteorologists have on the other hand worked with this problem for a long time. For a review of data assimilation methods used in meteorology, see e. g. Bengtsson et al. (1981), Lorenc (1986), Navon (1986) or Le Dimet and Navon (1989). Over the last decade, the development and implementation of four-dimensional data assimilation techniques have dramatically improved the accuracy

and dynamical consistency of meteorological analysis. Over the past 30 years of development of numerical analysis and assimilation schemes, most presented schemes belong to one of basically three different classes of algorithms:

- (1) Local polynomial interpolation methods
- (2) Statistical (optimal) interpolation methods
- (3) Variational numerical analysis methods

In the polynomial interpolation method, polynomial functions are adjusted to the observed data in the close neighborhood of each gridpoint. These methods were first introduced by Panofsky (1949) and also by Gilchrist and Cressman (1954), and have the great advantage of being both very simple and economical to use. The polynomial interpolation methods have been used for a number of years and are still often used for various purposes.

The choice of mathematical functions used to approximate the variation of meteorological variables in the vicinity of the gridpoint is quite arbitrary, and past experience on e. g. atmospheric scales does not enter into the analysis. In the second class of assimilation schemes, the statistical (optimal) interpolation schemes, past experience about the behavior of the atmosphere is used as the main source of information for determination of the interpolation weights. Eliassen (1954) first proposed the method based on spatial autocorrelation functions of the pressure field, while Gandin (1965) developed the method for operational use. Weights are assigned to a linear combination of observed departures from a guess field in an attempt to minimize the mean square analysis error. This requires knowledge of the statistical properties of the observed and unknown variables. The statistical information needed is the first and second moments of these variables. These methods are now widely used, especially as part of assimilation procedures for large-scale

prediction models, see e. g. Lorenc (1981). One major advantage of statistical interpolation is that it produces a practical and internally consistent approach for treating a large set of heterogenous observations, and it is at present the technique which produces the best results for operational weather forecasting. However, there are certain difficulties in using statistical interpolation. Several problems occur when this method is extended to the time dimension or is used in conjunction with an explicit dynamical model for temporal evolution of the atmospheric flow. In fact the procedures now in operational use are already in effect a heuristic generalization to the time dimension, but the successive analyses are still performed almost independently of the evolution equations. The question remains open of how much could be achieved by making the assimilation process more consistent with these equations. There are also other defects with statistical interpolation. One is that it tends in some cases to smooth excessively the analyzed fields. Excessive smoothing may be particularly troublesome for mesoscale forecasting models, since it may inhibit developments which are unlikely from a statistical point of view, but are very important to predict correctly, just because of the rarity of their occurrences.

The statistical (optimal) interpolation methods are computationally expensive to use. Large systems of linear equations have to be solved to find the interpolation weights. Successive correction methods introduced by Bergthorsson and Döös (1955) and Cressman (1959) and modified by Bratseth (1986) may be looked upon as an empirical approximation to the statistical interpolation method. In these methods the interpolation weights are computed explicitly without solving a system of linear equations, so the number of computations is relatively low compared with the statistical interpolation method.

For linear models an extension of the optimal interpolation methods is the Kalman or Kalman-Bucy (K-B) filtering (Kalman, 1960; Kalman and Bucy, 1961).

The K-B filter is even more computationally expensive to use than the optimal interpolation methods. In the latter method an estimate of the error covariance function (the covariance of the model/data differences) is used to compute a correction to the model solution. In the K-B filter the estimate of the covariance function is updated each time step. If a sophisticated model is used, the computational requirements may be unaffordable. The K-B filter has appeared in meteorology, e. g. Ghil et al. (1981). Lately it has been discussed in the oceanographic literature: Miller (1986), Budgell (1986, 1987) and Bennett and Budgell (1987). In all these applications a simple model has been used.

Miller (1986) used a Kalman filter to assimilate data into an eddy resolving open ocean model. He used the full Kalman filter, but in order to implement the method, the filter was applied to simplified systems designed to capture some of the properties of open ocean modelling. Results from Bennet and Budgell (1987) show that the K-B filter with regular time and space sampling at a certain period/wavelength, will not converge for waves of shorter periods/wavelengths. Kindle (1986) got similar results using an eddy resolving numerical model. He found that the model integration would not converge given observed data unless the data had a time/space sampling rate equal to the time/space decorrelation scale of the model eddy activity. The method Kindle (1986) used was a direct replacing of observations into the numerical model, and he got essentially the same results as Bennett and Budgell (1987). So one conclusion from these studies is that use of the K-B filter does not overcome the problem of resolution, but it does allow for a more rapid convergence for the periods/wavelengths which can be resolved.

The third class of assimilation schemes consists of variational analysis methods. In these schemes a given measure of the "distance" between the analysis and the observations is minimized. The analyzed field must at the same time (ap-

proximately) satisfy an explicit dynamical constraint. The constraint will normally be expressed by one (or more) differential equation(s). Variational algorithms have the great advantage of being able to provide exact consistency between the analysis and the dynamics. In this respect they appear clearly superior to the other assimilation techniques. On the other hand, variational schemes have a high mathematical technicality and also a high computational cost. Although the usefulness of variational methods for meteorological problems was pointed out very early by Sasaki (1955, 1958, 1970), and in spite of a fairly large number of various studies, e. g. Thompson (1969), Lewis and Bloom (1978) and Bloom (1983), these methods have not been fully utilized until recently. This is especially true for assimilation studies which address problems containing explicit time derivatives, see e. g. Lewis and Bloom (1978).

Le Dimet and Talagrand (1986) studied different variational algorithms for analysis and assimilation of meteorological observations. They discussed two different approaches to data assimilation. In the first approach the original constrained problem is transformed into an unconstrained problem or a sequence of unconstrained problems. Three different classical algorithms are presented, the penalty, the duality and the augmented Lagrangian algorithm. The latter is a generalization of the penalty and the duality algorithms. The second approach uses optimal control techniques, and is based on the use of an adjoint dynamical equation. This algorithm seems to work well in the assimilation of observations distributed in time. The main advantage of looking at unconstrained problems is that these problems can be solved by classical descent algorithms such as the steepest descent or the conjugate-gradient method.

In the last few years the variational approach has been studied for meteorological application by several authors, e. g. Derber (1985), Lewis and Derber

(1985), Harlan and O'Brien (1986), Hoffman (1986), Lorenc (1986, 1988), Talagrand and Courtier (1987) and Courtier and Talagrand (1987). In many of these investigations the models used are similar to the oceanographic models and from the success reported in these investigations it may be inferred that the variational techniques should be considered for oceanographic data assimilation.

A few remarks about the different classes of data assimilation should be made. Although they were conceived and developed independently, they do have mutual relationships. For example the interpolation method of Gilchrist and Cressman (1954) can be described as a simple case of statistical interpolation. Kimeldorf and Wahba (1970) have shown that the statistical interpolation method produces fields which are the solution of a variational problem in which the function to be minimized is the sum of two terms with one term representing the distance to the observations and the other term some measure of smoothness of the fields. The forward-backward data assimilation introduced by Morel et al. (1971) can be related to the variational assimilation. In their approach, the model is integrated forward and backward repeatedly over time to obtain an adjustment of the model to the observations. In the variational approach the model itself is integrated forward, but the adjoint of the model is used in the backward integration. Talagrand (1981) has shown that a sufficient condition for convergence of a forward-backward assimilation scheme as described by Morel et al. (1971), is that the linearized perturbation equations are antisymmetric. The adjoint of an antisymmetric equation is identical to the equation itself, so this explains the success reported by Morel et al. (1971) in the case of an antisymmetric equation. The assimilation procedures found from the variational approach are not restricted to antisymmetric equations. Thacker (1986) discussed the connection between Kalman filtering and the varia-

tional approach for data assimilation using a linear model. Kalman filtering can be thought of as an algorithm to solve the variational equations.

As mentioned earlier oceanographers have not had much experience with data assimilation. If the phrase data assimilation is used in a broader sense, the first attempts of assimilating data into models for the ocean were the dynamical methods used to produce maps of the currents from hydrographic observations as in Sandstrøm and Helland-Hansen (1903). Geostrophic shear can be directly established from the density data. These early attempts of data assimilation were followed by more sophisticated models e. g. Stommel and Schott's (1977) beta spiral or Wunsch's (1977,1978) inverse method. Recently there have been some developments in the area of data assimilation in oceanographic numerical models. Marshall (1985) used estimation theory to assimilate simulated data from satellite altimetry into an ocean model, and at the same time improving the geoid. Webb and Moore (1986) made use of the projection methods of estimation theory, but they assumed that the measurements were error free, and that altimeter measurements were available everywhere. They also assumed that the ocean currents were weak so that the ocean could be represented by a superposition of linear Rossby waves. A result of this study was that the determination of the deeper structure of the ocean was limited by the phase separation that develops over each assimilation cycle between modes of the ocean with the same horizontal wavenumber but differing vertical structure. Hurlburt (1986) used a two-active-layer, free surface primitive equation model on a  $\beta$ -plane to investigate the dynamic transfer of surface to subsurface information. Perfect altimeter data were simulated by the free surface of the two layer model. The results showed that the maximum update interval that provided success was about half the shortest major time scale in the model.



Bennett and McIntosh (1982) used a variational method in the investigation of tidal motion. The method is an infinite-dimensional generalization of the finite-dimensional method used by Wunsch (1978) to determine barotropic circulation in the North Atlantic. Their results show that the choice of data weights are of great importance. Provost (1983) and Provost and Salmon (1986) have used a variational technique to assimilate hydrographic station data to estimate the three-dimensional field of geostrophic velocities. The method of weak constraints (Sasaki, 1970) was used. They determined the smoothest velocity field which was consistent with the data and at the same time approximately satisfied the dynamical constraints. The way they assured a smooth solution was to penalize kinetic energy as well as enstrophy. Malanotte-Rizzoli and Holland (1986) used a quasi-geostrophic general circulation model to investigate the effect of data insertion into a numerical model, and how data insertion in local, limited regions effect the dynamics of the model, and also how the dynamics spread the inserted information to different regions. Their model is steady, weakly nonlinear and highly frictional. The method they use is a naive approach with direct insertion of “observations” in the model. The “observations” used in this study are identical-twin data, i.e. data generated by the model. The results show that the region influenced by the insertion and the improvement of the results depend upon the location of the “observed” data, the orientation of the section inserted and the length of the section. Malanotte-Rizzoli and Holland (1988) extended their results from the 1986 paper to the transient eddy-resolving case. One of their major results is that in the transient case a single data section is very ineffective to drive the model towards the reference ocean over time scales of  $\sim 100$  days. All the different sections they tried were equally ineffective. If instead the data are inserted over a period longer than the model equilibrium time, a single section is quite effective in driving the model to the reference ocean.

Holland and Malanotte-Rizzoli (1989) studied the effect of along-track assimilation of altimeter data into a three layer eddy-resolving quasi-geostrophic model. Again identical twin data were used. They reached the conclusion that if perfect data are available (i. e. data at every gridpoint and every  $\frac{1}{2}$  day with linear interpolation in time), a simple (nudging) technique can be very successful in driving the model towards the control solution. If altimetric data are available only along actual satellite tracks with repeat periods of either 10 or 20 days, the results are not as promising. The residual root mean square (rms) errors are close to 70% at the end of the assimilation period. In Malanotte- Rizzoli et al. (1989) a primitive equation model was used to study a series of idealized initialization/assimilation experiments. The first question they ask is whether the initialization shocks are equally important in ocean models as they are in the atmospheric models. Their conclusion is that the ocean models do not seem to be as sensitive to unbalanced initializations. All that is necessary to ensure a smooth evolution, is a geostrophically balanced initialization. They also studied data assimilation and the effect of different types of data. A simple insertion technique was used as the assimilation scheme, and identical twin data represented the observations. The results of this investigation show that the knowledge of the interior density field is the most effective data in reducing the rms errors.

In a study by Schröter and Wunsch (1986) the effect of observational errors in the driving of the models is investigated. The procedure is based upon nonlinear optimization methods. From their algorithm it is possible to calculate the qualitative sensitivity of the objective function to change in the data errors and finding an optimization technique capable of dealing with data uncertainty. Moore et al. (1987) studied the initialization and data assimilation in two different models of the Indian Ocean. The models they used were a reduced-gravity model and

the GCM of Bryan (1969) in the Semtner (1974) coding. The method they used was a direct updating of the whole temperature and/or the velocity fields. In an attempt to simulate assimilation of data from the proposed TOGA XBT network, they used the method of successive correction (Bergthorsson and Döös, 1955). All their tests were performed using an "identical twin" approach. The result of this study showed that the use of temperature data (mass) gave better results than if velocity observations were used. The effect of using velocity data could be increased if the ratio between the potential and kinetic energy in the model were changed. An increase of the kinetic energy made velocity data more useful in the assimilation process.

In two papers, Thacker and Long (1988) and Thacker (1988), a variational data assimilation procedure is described. The method is illustrated using a simple model of the wind-driven equatorial ocean. It is shown how surface elevation and wind stress observations can be used to recover the model state. In Thacker (1988) the process of fitting a model to inadequate data is discussed. Results show that for the simple three wave model of Thacker and Long (1988), a reasonable fit can be obtained even if the number of data is less than the number of degrees of freedom of the model.

The approach followed here is the variational data assimilation method. Sasaki's (1970) method of strong constraints is used. The assimilation procedure will be formulated for an equatorial Pacific Ocean model. The goal will be to determine the solution of the model which best fits the observations. The best fit solution could be determined in different ways. The meteorological approach of finding the initial conditions which give the best solution is one method. When the variational data assimilation method is used, it is possible to adjust other variables than the initial conditions. Parameters in the model can easily be incorporated

in the procedure, so that they can be used as control parameters. The variational algorithm is going to be formulated using the phase speed of the Pacific Ocean model as the control parameter. The phase speed is then going to be estimated so that the model solution is as close as possible to the observations. Observations of sea level from island stations are going to be assimilated into the model. Knowledge of the spatial structure of the phase speed will give information about the basic stratification of the Pacific Ocean.

The variational method is described in section 2. A general description of the variational formalism is given. Similarities between different approaches to the variational data assimilation method are discussed. The choice of the cost function and different ways of dealing with insufficient data are also discussed. An outline of the solution procedure and some computational aspects of the method are described. Section 3 gives a general description of the conjugate gradient method, while section 4 discusses some aspects of parameter estimation. In section 5 the problem of data assimilation in a Pacific Ocean model is specified. The model equations are described and the corresponding adjoint equations are derived. The wind stress used as forcing in the Pacific model is presented. A description of the sea level observations from the island stations in the Pacific Ocean is given. The method for calculating the data misfit used as forcing in the adjoint equations is also described. Section 6 presents results obtained using the data assimilation algorithm. Both simple experiments using the height and velocity field of a Kelvin wave as initial conditions and realistic results using observed wind stress as forcing, are discussed. A summary of the results and the conclusions are given in section 7. The derivation of the continuous adjoint equations are given in Appendix A, while the finite difference version of the equations are derived in Appendix B.

## 2. Variational formalism

### 2.1. Description of the method

The assimilation scheme that will be used is a variational method. As mentioned in the introduction, this scheme minimizes a given measure of the distance between the model and the observations. The primary field of interest is time dependent models of the ocean and a mathematical model may be written as

$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{F}(\mathbf{x}; \mathbf{c}) \quad (2 - 1)$$

where the components of the vector  $\mathbf{x}$  are the dependent variables,  $t$  is time and the vector  $\mathbf{c}$  contains the parameters of the model. The components of  $\mathbf{x}$  denote the various oceanographical fields ( velocity, height field, ....) under consideration.  $\mathbf{F}$  may be a linear or non-linear operator. It is assumed that the system (2 - 1) is not closed, i. e. additional information has to be provided in order to obtain a unique solution to (2 - 1). The additional information will be given by  $\mathbf{y}$  - a control variable.  $\mathbf{y}$  can for instance consist of the initial conditions, boundary conditions, some of the parameters in the problem, or a combination of these. When  $\mathbf{y}$  is defined, a unique solution  $\mathbf{x}(\mathbf{y})$  of (2 - 1) can be found. It is important to realize that the control variable  $\mathbf{y}$  must belong to a set of admissible control,  $\mathbf{y}_{ad}$ . To determine  $\mathbf{y}_{ad}$  one may use physical information about the initial conditions or the parameters used as a control variable.

One is interested in obtaining a solution of (2 - 1) which is close to the observations,  $\mathbf{x}'$ . Closeness may be defined in different ways, and it is usual to define a cost function  $J$ , which measures the distance between the solution of (2 - 1) and

the observations. The definition of  $J$  will be discussed later. The observation,  $\mathbf{x}'$ , will consist of measurements of oceanographic fields, e. g. the height field or the velocity field. A variational problem can now be stated as a problem (**P**):

*Find  $\mathbf{y}^*$  which belongs to  $y_{ad}$  and minimizes the cost function  $J$ .*

where  $\mathbf{y}^*$  represents the optimal  $\mathbf{y}$ .

The problem stated above is a constrained minimization problem with the model equation (2 – 1) representing the constraint. The way this problem is going to be solved, is by redefining it so that it becomes a problem of unconstrained minimization. Standard procedures for solving problems of this type may be found in textbooks on optimization (e. g. Gill et al, 1981, Luenberger, 1984) or in the meteorological literature (e. g Le-Dimet and Talagrand, 1986, Navon and Legler, 1987).

Sasaki (1970) introduced two different methods in variational analysis, the method of weak constraint related to the penalty method and the method of strong constraint. The weak constraint formalism consists of minimizing the functional  $E$  defined by

$$E(\mathbf{x}, \mathbf{y}) = J(\mathbf{x}, \mathbf{y}) + \alpha \| \mathbf{G}(\mathbf{x}; \mathbf{c}) \|^2 \quad (2 - 2)$$

where  $\alpha$  is a prespecified weight and  $\mathbf{G}$  is given by

$$\mathbf{G}(\mathbf{x}; \mathbf{c}) = \frac{\partial \mathbf{x}}{\partial t} - \mathbf{F}(\mathbf{x}; \mathbf{c}) \quad (2 - 3)$$

It is important to note that in this formulation  $E$  is quadratic in  $\mathbf{G}$  and the weights are prespecified. It therefore follows that the constraint is only approximately verified. In some cases this may be justified since the model used as constraints is an approximation to the real world and should only be satisfied to its own accuracy.

The approach that will be followed here is based on the classical Lagrange multiplier technique. A Lagrangian  $L(\mathbf{x}, \lambda, \mathbf{y})$  can be constructed

$$L(\mathbf{x}, \lambda, \mathbf{y}) = J(\mathbf{x}, \mathbf{y}) + (\lambda, \mathbf{G}(\mathbf{x}; \mathbf{c})) \quad (2 - 4)$$

where the components of the vector  $\lambda$  are the Lagrangian multipliers.  $(\cdot, \cdot)$  is an inner product defined on the functional space to which  $\mathbf{G}(\mathbf{x}; \mathbf{c})$  belongs. This is the strong constraint formalism according to Sasaki (1970). The constrained minimization problem (**P**) is thus replaced by an unconstrained problem with respect to the variables  $\mathbf{x}$ ,  $\lambda$  and  $\mathbf{y}$ . Using this formalism it is insured that the observations will satisfy the constraints exactly. It can be shown that the problem of determining the stationary points of the functional  $J(\mathbf{x}, \mathbf{y})$  under the constraint  $\mathbf{G}(\mathbf{x}; \mathbf{c}) = 0$  is equivalent to the problem of determining the stationary points of (2 - 4) with respect to the variables  $\mathbf{x}$ ,  $\mathbf{y}$  and  $\lambda$  (see e. g. Bertsekas (1982)). The equations which express that the Lagrangian is stationary, are called the Euler-Lagrange equations of problem (**P**). The Euler-Lagrange optimality condition is given by

$$\frac{\partial L}{\partial \lambda}(\mathbf{x}^*, \lambda^*, \mathbf{y}^*) = 0 \quad (2 - 5)$$

$$\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}^*, \lambda^*, \mathbf{y}^*) = 0 \quad (2 - 6)$$

$$\frac{\partial L}{\partial \mathbf{y}}(\mathbf{x}^*, \lambda^*, \mathbf{y}^*) = 0 \quad (2 - 7)$$

and determines  $\mathbf{x}^*$ ,  $\lambda^*$  and  $\mathbf{y}^*$ . Equation (2 - 5) gives the original model equation back. The operator in (2 - 6) is the adjoint of the operator in (2 - 5). It is therefore customary to call equation (2-6) the adjoint equation of (2-5). The model equation will propagate information forward in time, while the adjoint equation is going to propagate information backward in time. When the appropriate model equations have been derived, the interpretation of the adjoint equation will be discussed more closely. The last equation, (2 - 7), gives the gradient of  $L$  with respect to the

control variable. The gradient plays an important role in determining the best fit solution, since it is one of the major constituents in the descent algorithms used to find the minimum of the cost function (the Lagrangian). As can be seen from these equations the whole variational analysis depends on the choices of the functional  $J$  and the constraint  $\mathbf{G}$ . The method of solution is also an important part of the analysis. For most choices of  $J$  and  $\mathbf{G}$  an iterative method has to be used and a goal will be to keep the number of iterations as low as possible.

There are other classical algorithms for solving the constrained problem (P). One algorithm introduces the penalized functional

$$J_\varepsilon(\mathbf{x}, \mathbf{y}) = J(\mathbf{x}, \mathbf{y}) + \frac{1}{\varepsilon} \|\mathbf{G}(\mathbf{x}; \mathbf{c})\|^2 \quad (2 - 8)$$

where  $\varepsilon$  represents what is called the penalty parameter and is a sequence of positive numbers tending to zero. It can be shown that under certain hypotheses the solution  $\mathbf{x}_\varepsilon^*$  tends to the solution  $\mathbf{x}^*$  of the original problem when  $\varepsilon$  tends to zero. The similarity between this method and what Sasaki (1970) called the weak constraint formalism (2 - 2) is evident. However, there is an advantage of using the penalty method. By varying  $\varepsilon$  instead of using only one constant prespecified weight  $\alpha$ , one can determine the dependence of the solution upon the choice of coefficient.

Numerical problems may arise when the penalty algorithm is used. For small values of  $\varepsilon$  the method may lead to ill-conditioning and inaccuracy in the determination of the minimum. This can be dealt with by introducing the augmented Lagrangian

$$L_\varepsilon(\mathbf{x}, \mathbf{y}) = J(\mathbf{x}, \mathbf{y}) + \frac{1}{\varepsilon} \|\mathbf{G}(\mathbf{x}; \mathbf{c})\|^2 + (\lambda, \mathbf{G}(\mathbf{x}; \mathbf{c})) \quad (2 - 9)$$

where as before  $\lambda$  represents the Lagrangian multipliers (see Bertsekas (1982)). The augmented Lagrangian method would in Sasaki's terminology be a combination between weak and strong constraints. The major advantage of the augmented



Lagrangian algorithm is that it is computationally more efficient than the penalty method. The augmented Lagrangian alleviates the problems connected with using the penalty or the multiplier method alone. Navon and de Villiers (1983) used an augmented Lagrangian algorithm to enforce integral invariants in a shallow water equation model of the atmosphere.

It is possible to formulate the method of adjoint equations without introducing the Lagrange multipliers. Optimal control techniques which in fact are techniques for minimizing a functional as the cost function, can be used to calculate the gradient of the cost function with respect to the control variable(s). Using these techniques it is necessary to solve the model equation (2 – 1) and then solve the adjoint of the linearized form of the same equation. This approach is somewhat more difficult to formulate. It is necessary to write a well-posed problem and carefully specify the functional framework of the variational problem. Since the optimal control techniques do not introduce the Lagrange multipliers it might look like this is a large reduction in the number of unknowns. But, in fact, in the calculation of the gradient of the cost function it is necessary to compute the same numbers of variables. The method of optimal control techniques is discussed in Le Dimet and Talagrand (1986) (see also Lions (1971)). A review of these techniques used in meteorology can be found in Le Dimet and Navon (1989).

## 2.2. *The cost function $J$*

The objective of the data assimilation is to find a solution which is close to the observations. The cost function  $J$  is constructed so it measures the distance (the misfit) between the model and the observations, and it must therefore have the property of a norm. There are several different functional forms of  $J$  which might be considered, and each one of them will give a different result for the best fit model solution. The variational method makes use of the derivative of  $J$ , and the cost

function must therefore be differentiable. It is usual to choose  $J$  as the square of a norm. Let  $\Sigma$  denote the spatial and/or temporal domain in which (2 – 1) is valid.  $J$  can then be defined as

$$J(\mathbf{x}, \mathbf{c}) = \int_{\Sigma} \frac{1}{2}(\mathbf{x} - \mathbf{x}')^T K_{\mathbf{x}}(\mathbf{x} - \mathbf{x}') d\sigma + \int_{\Sigma} \frac{1}{2}(\mathbf{c} - \mathbf{c}')^T K_{\mathbf{c}}(\mathbf{c} - \mathbf{c}') d\sigma \quad (2 - 10)$$

where as above  $'$  denotes an observed or estimated quantity and the  $K$ 's are specified validity matrices. They depend on the error variance of each observational point, and give information about the quality of the data. If the observational errors are uncorrelated, the validity matrices are diagonal. The  $K$ 's are always symmetric positive definite matrices. Observations may not be available everywhere. For points where there are no observations, the validity matrices are set equal to zero. The first term in (2 – 10) is called the data misfits and is going to be the forcing for the adjoint equation. The last term in (2 – 10) is added to the cost function because the main interest of this study is to estimate the parameters in the model. By adding this term, one makes sure that the new estimate of the parameters is not too far from the initial guess. In other words minimizing the cost function results in a solution which is close to the observations and new values of the parameters which are close to the estimate. The effect of the last term will be discussed further in section 4.

It is necessary to have several observations per degree of freedom in order to get reasonable results from the data assimilation. For numerical models of the ocean or the atmosphere, the number of degrees of freedom is related to the number of gridpoints used, or for a spectral model, the number of terms included in the spectral expansion. Most models already have more degrees of freedom than the available data bases. One way to supplement insufficient data is by adding a “penalty” or smoothing term to the cost function. Sasaki (1970) used this idea and he pointed out that adding a smoothing term to the cost function suppressed high frequencies

and wave numbers in the solution. Wahba and Wendelberger (1980) have shown that the degree of the highest derivative in the smoothing term must be at least two orders of magnitude greater than the highest derivative in the data constraint if a mathematically well behaved solution is going to be obtained. This idea has been used by e. g. Provost (1983), Provost and Salmon (1986), Harlan and O'Brien (1986), Thacker (1988) and Legler et al. (1989). There are many choices of these penalty terms. For instance the addition of a term of the form

$$K_p \nabla^2 \mathbf{x} \quad (2 - 11)$$

will tend to minimize the curvature of the assimilated field. In other words it tends to make the resulting field smoother.

Another way to add more information would be to use a solution of a previous model run as observations, or if no such run is available climatological values for the control parameters could be used as observations, Lorenc (1986). As long as the weights,  $K$ , used for the "observations" from previous forecasts or climatology are small compared to the  $K$ 's of the real observations, the effect of these data points will have a negligible effect on the solution except for the part that is not determined by the real data.

### 2.3. Outline of solution procedure

The procedure for using the variational method for data assimilation and parameter estimation can be formulated in the following way.

1. Choose a first guess for the control parameters.
2. Integrate the model forward for the period over which the observations are going to be assimilated.
3. Calculate the data misfits  $(\mathbf{x} - \mathbf{x}')$ .

4. Integrate the adjoint equations backward in time forced by the data misfits.
5. Calculate the gradient of  $L$  with respect to the control variables.
6. Use the gradient of  $L$  in a descent algorithm to find a new value of the control parameters which make the cost function move towards its minimum.
7. Check if the optimal solution has been found. This can be done by checking the norm of  $\nabla J$  or the value of  $J$  to see if it is less than a prescribed tolerance.
8. If this is not the case, the procedure described above can be repeated using the new values of the control parameters as a new guess and continue the iterative process until a satisfactory solution has been determined.

When trying to find the minimum of the cost function (the Lagrangian) it is important to realize that one cannot expect the minimum to be exactly zero. Experience from meteorology has shown that the decrease in the value of the cost function or its gradient is at most a couple of orders of magnitude. Hydrologist's experience in estimating parameters in groundwater flow using the variational technique, have shown that relying on the gradient norm as the only measure of convergence to the optimal solution may lead to an unnecessarily large number of iterations. A better way of determining if the optimal solution has been found would be to check and see if  $\frac{\|\nabla J\|}{\|\nabla J_0\|}$  is less than a prescribed tolerance, where  $\|\nabla J_0\|$  denotes the norm of the gradient during the first iteration. In view of the meteorological and hydrological experience the reduction of the tolerance can be expected to be of the order  $10^{-2}$ . Similarly, instead of checking the absolute value of  $J$ , the ratio  $\frac{J}{J_0}$  should be less than a specified tolerance. The decrease in the cost function itself may not be a good measure of how close the solution is to the minimum. Trying to fit a model to real observations is a difficult task. Depending on the complexity

of the model and its ability to predict the real world, it may not be possible to reduce the value of the cost function with several orders of magnitude. If the initial guess for the parameters is close to the optimal value, one cannot expect the cost function or its gradient to experience a large decrease. This does not mean that the minimization has not been successful. Other means of checking the behavior of the model should be used to see if the results are improved. In the experiments described later, a correlation coefficient is calculated to show that after the assimilation there is an increase in the correlation between the model and the observations, even if the decrease in the cost function is less than an order of magnitude.

During the first few iterations the reduction in the value of the cost function is fast. After this initial decrease, the reduction per iteration is going to be slower. The goal should be to keep the number of iterations below about 15-20. If it is necessary to continue with further iterations, the descent algorithm may have to be changed. Another choice may give a better convergence rate.

#### *2.4. Computational aspects*

The variational approach to data assimilation implies the solving of the system of equations (2 – 5) – (2 – 7). This may be computationally challenging. For an ocean model the number of variables can easily exceed  $10^6$ . The system (2 – 5) – (2 – 7) consists of the model equations and their adjoint. The adjoint equations are of similar complexity as the original model equations. Each iteration requires one forward integration of the model itself and one backward integration of the corresponding adjoint equations. Depending on how many iterations are necessary for the solution to converge, the method may require several times the computer time used by the integration of the ocean model. The necessity of using a descent algorithm which will find the minimum of the cost function in as few iterations as possible is therefore an important part of the variational formalism.

There are several important aspects which have to be considered when a descent algorithm is chosen. Some of the simpler algorithms, e. g., the steepest descent method, are easy to implement, but their slow convergence rate may make them impossible to use in practical situations. More sophisticated methods such as Newton or quasi-Newton methods have quadratic convergence rate, but these algorithms require the storage of Hessian matrices. For problems in oceanography with a large number of variables, the dimension of these matrices may be too large to fit on the available computer. Another aspect which is important in the descent algorithm, is the determination of the step length. The step length determines the distance which one is moving down the gradient of the cost function. There are different ways of calculating this parameter. Most of the methods require an additional integration of the forward model. One iteration of the data assimilation procedure may therefore result in 3 times the computer time required by integrating the model itself.

### 3. Descent methods

#### 3.1. Introduction

There are several different descent methods available for minimizing the function in (2-4). The oldest and also the simplest way to find the minimum of a function of several variables, is to use the method of steepest descent. One of the reasons why this method is still important is that it is one of the simplest for which a satisfactory analysis exists. Several of the more advanced algorithms are often a result of a modification of the basic steepest descent method in such a way that the new algorithm has improved convergence properties.

The models used in oceanographic studies often have a large number of variables  $N$ , which can easily exceed  $10^4$ , and for such cases the only choice for a descent method is the conjugate-gradient algorithm with its better convergence rate than the steepest descent. There are methods with better rate of convergence than the conjugate-gradient algorithm, e. g. Newton and quasi-Newton methods. These methods have quadratic rates of convergence and superlinear rates of convergence, respectively, but the disadvantage is that they require storage of Hessian matrices of second derivatives of size  $(N \times N)$ . Conjugate-gradient algorithms require storage of only a few vectors of length  $N$ .

The conjugate-gradient algorithm is an iterative method for unconstrained minimization of a general function of  $N$  variables. The method produces a better approximation of the minimum of the unconstrained nonlinear function with each iteration. During each iteration an estimate is made of the best way to change

each component of the vector  $\mathbf{x}$  in order to produce the maximum reduction in the function. The descent direction is found by combining information about the gradient of the function with information from earlier iterations to produce a new search direction. The algorithm also estimates an optimal step size which must be used to find the magnitude of the changes along the search direction.

The conjugate-gradient method has been successfully applied in meteorology to minimize the cost function used in variational analysis, e. g. Hoffman (1982, 1984), Navon and de Villiers (1983) and Derber (1985). Navon and Legler (1987) compared different conjugate-gradient algorithms by applying them to two different meteorological problems, and their conclusion was that the most consistent method was the Shanno and Phua (1980) quasi-Newton limited-memory (memoryless) conjugate-gradient algorithm. An introduction to the conjugate-gradient method is going to be presented in the next section. To implement the Shanno (1978a, b) algorithm, the subroutine CONMIN of Shanno and Phua (1980) is going to be used. This subroutine contains both a limited-memory quasi-Newton conjugate gradient algorithm and a quasi-Newton method with a BFGS update (Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno (1970)). Detailed description of different minimization algorithms can be found in e. g. Gill et al. (1981) or Luenberger (1984). The conjugate-gradient method was initiated by Hestenes and Stieffel (1952) and the historical development of the method can be found in Hestenes (1980).

### 3.2. *The conjugate-gradient method*

The conjugate-gradient method, or more generally the conjugate-direction method is analyzed for a purely quadratic function. The problem which will be considered is the problem of minimizing the cost function in (2 – 10). This is a



problem of unconstrained minimization and it can be written in the following form

$$\text{minimize } J(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{Q}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c \quad (3-1)$$

where  $J(\mathbf{x})$  represents the cost function and  $\mathbf{x}$  is a vector representing the  $N$  variables,  $x_0, \dots, x_{N-1}$ , of the function. It is assumed that  $J(\mathbf{x})$  can be written as a quadratic function, where  $\mathbf{Q}$  is a positive definite symmetric matrix,  $\mathbf{b}$  is a vector and  $c$  is a constant.  $\mathbf{x}^T$  represents the transpose of  $\mathbf{x}$ . In general an iterative algorithm for solving (3-1) takes the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \quad (3-2)$$

where  $\mathbf{d}_k$  is a descent direction and  $\alpha_k$  is a positive step size parameter. (3-2) is referred to as a generalized gradient method (or just a gradient method).

Let

$$\mathbf{g}_k = \nabla J(\mathbf{x}_k) \quad (3-3)$$

denote the gradient of  $J$  with respect to  $\mathbf{x}_k$ . If the method of steepest descent is used, the descent direction  $\mathbf{d}_k$  is simply given by  $-\mathbf{g}_k$ , the negative gradient of the function.

A couple of useful properties of conjugate-directions are going to be stated. The definition of conjugacy can be formulated as

**Definition:** Given a symmetric matrix  $\mathbf{Q}$ , two vectors  $\mathbf{d}_i$  and  $\mathbf{d}_j$  are said to be conjugate with respect to  $\mathbf{Q}$  if  $\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_j = 0$  for  $i \neq j$ .

If  $\mathbf{Q} = \mathbf{I}$ , conjugacy is equivalent to the usual notion of orthogonality.

**Theorem 1:** If  $\mathbf{Q}$  is positive definite and the vectors  $\mathbf{d}_i$  are mutually conjugate, then these vectors are linearly independent.

The next step will be to construct a set of mutually conjugate directions. For a given set of  $N$  linearly independent vectors  $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{N-1}$ , it is possible to construct a set of conjugate directions  $\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_{N-1}$  in the following way. Let

$$\mathbf{d}_0 = \mathbf{v}_0 \quad (3-4)$$

and then define

$$\mathbf{d}_i = \mathbf{v}_i + \sum_{j=0}^{i-1} a_{ij} \mathbf{d}_j \quad (3-5)$$

for  $i = 1, \dots, N-1$ . The coefficients  $a_{ij}$  are chosen so that  $\mathbf{d}_i$  is conjugate to the previous directions  $\mathbf{d}_{i-1}, \dots, \mathbf{d}_0$ . Multiplying (3-5) by  $\mathbf{Q}\mathbf{d}_l$  for  $l = 0, \dots, i-1$  gives

$$\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_l = \mathbf{v}_i^T \mathbf{Q} \mathbf{d}_l + \sum_{j=0}^{i-1} a_{ij} \mathbf{d}_j^T \mathbf{Q} \mathbf{d}_l = 0 \quad (3-6)$$

If all the previous  $a_{ij}$ 's have been chosen so that  $\mathbf{d}_0, \dots, \mathbf{d}_{i-1}$  are conjugate, then

$$\mathbf{d}_j^T \mathbf{Q} \mathbf{d}_l = 0 \quad , \quad j \neq l \quad (3-7)$$

and from (3-6) it follows

$$a_{ij} = -\frac{\mathbf{v}_i^T \mathbf{Q} \mathbf{d}_j}{\mathbf{d}_j^T \mathbf{Q} \mathbf{d}_j} \quad (3-8)$$

for all  $i = 0, \dots, N-1$  and  $j = 0, \dots, i-1$ .

The set of vectors  $\mathbf{d}_0, \dots, \mathbf{d}_{N-1}$  defined by (3-4) – (3-8) is conjugate with respect to  $\mathbf{Q}$ .

The above procedure can now be used to develop the conjugate-gradient method. The first step is to let

$$\mathbf{v}_0 = -\mathbf{g}_0 \quad (3-9)$$

i. e. the initial step is in the direction of the negative gradient of  $J$  and it is identical to a steepest descent direction. The rest of the  $\mathbf{v}$ 's are chosen as  $\mathbf{v}_1 = -\mathbf{g}_1, \dots, \mathbf{v}_{N-1} = -\mathbf{g}_{N-1}$ . Using (3-2) one gets

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 \quad (3-10)$$

The next conjugate direction  $\mathbf{d}_1$  can be found by using (3-5) and (3-8) with  $\mathbf{v}_1 = -\mathbf{g}_1$ .

$$\mathbf{d}_1 = -\mathbf{g}_1 + \frac{\mathbf{g}_1^T \mathbf{Q} \mathbf{d}_0}{\mathbf{d}_0^T \mathbf{Q} \mathbf{d}_0} \mathbf{d}_0 \quad (3-11)$$

From (3-1)

$$\mathbf{g}_{k+1} - \mathbf{g}_k = \mathbf{Q}(\mathbf{x}_{k+1} - \mathbf{x}_k) \quad (3-12)$$

and using (3-2)

$$\mathbf{g}_1 - \mathbf{g}_0 = \mathbf{Q}(\mathbf{x}_1 - \mathbf{x}_0) = \alpha_0 \mathbf{Q} \mathbf{d}_0 \quad (3-13)$$

(3-11) can now be written as

$$\mathbf{d}_1 = -\mathbf{g}_1 + \frac{\mathbf{g}_1^T (\mathbf{g}_1 - \mathbf{g}_0)}{\mathbf{d}_0^T (\mathbf{g}_1 - \mathbf{g}_0)} \mathbf{d}_0 \quad (3-14)$$

The process can be repeated and the result is

$$\mathbf{d}_k = -\mathbf{g}_k + \sum_{j=0}^{k-1} \frac{\mathbf{g}_k^T (\mathbf{g}_{j+1} - \mathbf{g}_j)}{\mathbf{d}_j^T (\mathbf{g}_{j+1} - \mathbf{g}_j)} \mathbf{d}_j \quad (3-15)$$

This expression can be simplified using the fact that

$$\mathbf{g}_k^T \mathbf{d}_j = 0 \quad (3-16)$$

for  $j = 0, \dots, k-1$ . The subspace spanned by  $\mathbf{d}_0, \dots, \mathbf{d}_{k-1}$  is also the subspace spanned by  $\mathbf{g}_0, \dots, \mathbf{g}_{k-1}$  and hence

$$\mathbf{g}_k^T \mathbf{g}_j = 0 \quad (3-17)$$

for  $j = 0, \dots, k-1$ . (3-15) now reduces to

$$\mathbf{d}_k = -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1} \quad (3-18)$$

where

$$\beta_k = \frac{\mathbf{g}_k^T (\mathbf{g}_k - \mathbf{g}_{k-1})}{\mathbf{d}_{k-1}^T (\mathbf{g}_k - \mathbf{g}_{k-1})} \quad (3-19)$$

Again using (3 – 16), (3 – 17) and (3 – 18) for  $\mathbf{d}_{k-1}$ , the expression for  $\beta$  can be rewritten as

$$\beta_k = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}} \quad (3 - 20)$$

It is important to note that in order to calculate the new direction  $\mathbf{d}_k$ , it is only necessary to know the current and the previous gradients of  $J$  and the previous direction  $\mathbf{d}_{k-1}$ .

There is one problem with using the conjugate-gradient algorithm described above. If the number of variables is large as in oceanographical problems, the conjugate-gradient method can produce a direction of search after a few iterations which is not very efficient due to loss of conjugacy. There have been several attempts to avoid this effect. One of the first applications used periodic restarts with the steepest descent direction. However, there is a disadvantage using this method. The reduction at the restart iteration is often poor compared with the reduction that would have occurred without restarting. Beale (1972) proposed another restart procedure. He showed that using the direction

$$\mathbf{d}_k = -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1} + \gamma_k \mathbf{d}_t \quad (3 - 21)$$

which is similar to the conjugate-gradient direction with the addition of a multiple of the restart direction  $\mathbf{d}_t$ . In this method  $\beta_k$  and  $\gamma_k$  are given by

$$\beta_k = \frac{(\mathbf{g}_{k-1} - \mathbf{g}_k)^T \mathbf{g}_{k+1}}{(\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{d}_k} \quad (3 - 22)$$

$$\gamma_k = \frac{(\mathbf{g}_{t-1} - \mathbf{g}_t)^T \mathbf{g}_{k+1}}{(\mathbf{g}_{t+1} - \mathbf{g}_t)^T \mathbf{d}_t} \quad (3 - 23)$$

$\mathbf{d}_t$  is the last direction of the previous conjugate-gradient cycle along which a linear search was made. Powell (1977) suggested a condition for when a restart should be performed. If the inequality

$$|\mathbf{g}_k^T \mathbf{g}_{k+1}| \geq 0.2 \|\mathbf{g}_{k+1}\| \quad (3 - 24)$$

holds or if there have been  $N$  linear searches, the conjugate gradient should be restarted. A restart should also take place if the direction of search is not sufficiently downhill, where an adequate downhill direction is one that satisfies the two inequalities

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} \leq -0.8 \|\mathbf{g}_{k+1}\|^2 \quad (3-25)$$

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} \geq -1.2 \|\mathbf{g}_{k+1}\|^2 \quad (3-26)$$

This means that frequent restarts will take place.

### 3.3. Newton, quasi-Newton and limited-memory quasi-Newton methods

Newton's method is based on the idea that the function  $J$  which is going to be minimized is approximated by a quadratic function. If  $\mathbf{x}^*$  is the minimum of the function, then a Taylor series expansion of  $J$  near a point  $\mathbf{x}_k$  is given by

$$J(\mathbf{x}_k + \mathbf{p}) \sim J(\mathbf{x}_k) + \mathbf{g}_k^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{Q}(\mathbf{x}_k) \mathbf{p} \quad (3-27)$$

where as before  $\mathbf{g}_k = \nabla J(\mathbf{x}_k)$  and  $\mathbf{p} = \mathbf{x}^* - \mathbf{x}_k$ , the distance to the minimum. The right hand side is minimized if  $\mathbf{p}_k$  satisfies the system

$$\mathbf{Q}_k \mathbf{p}_k = -\mathbf{g}_k \quad (3-28)$$

which gives the Newton method. A descent algorithm using Newton's method is therefore given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{Q}_k^{-1} \mathbf{g}_k \quad (3-29)$$

where  $\mathbf{Q}_k^{-1}$  is the inverse Hessian matrix. It is also assumed that  $\mathbf{Q}_k^{-1}$  exists and that the direction  $\mathbf{p}_k$  determined from (3-28) is a descent direction, i. e.  $\mathbf{p}_k^T \mathbf{g}_k < 0$ . If the Hessian matrix  $\mathbf{Q}_k$  is positive definite, the direction  $\mathbf{p}_k$  is clearly a descent direction, since  $\mathbf{p}_k^T \mathbf{g}_k = -\mathbf{g}_k^T \mathbf{Q}_k^{-1} \mathbf{g}_k < 0$ . On the other hand if  $\mathbf{Q}_k$  is not positive definite, the quadratic function need not have a minimum or even a stationary

point. To be able to use this method it is necessary to know the zero, first and second derivatives of  $J(\mathbf{x})$  at any point.

Quasi-Newton methods are based on the assumption that information about the curvature (the Hessian) of the quadratic function can be obtained as the iterations of a descent method proceed using the observed behavior of  $J$  and  $\mathbf{g}$ . The whole idea is therefore to compute an approximation to the curvature of the function without actually forming the Hessian matrix. A quasi-Newton iteration method can be described by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \quad (3-30)$$

$$\mathbf{B}_k \mathbf{d}_k = -\mathbf{g}_k \quad (3-31)$$

where  $\mathbf{B}_k$  is an approximation to the Hessian matrix. At the first iteration it is usual to take  $\mathbf{B}_0$  as the identity matrix. After each iteration  $\mathbf{B}_k$  is updated to obtain a new approximation of the Hessian,  $\mathbf{B}_{k+1}$

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \mathbf{U}_k \quad (3-32)$$

where  $\mathbf{U}_k$  is the update matrix. The new approximation  $\mathbf{B}_{k+1}$  is required to satisfy the quasi-Newton condition

$$\mathbf{B}_{k+1} \mathbf{s}_k = \mathbf{y}_k \quad (3-33)$$

where

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k \quad (3-34)$$

the change in  $\mathbf{x}$  during the  $k$ -th iteration, and

$$\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k \quad (3-35)$$

the change in the gradient.

There are several different ways of updating the Hessian matrix. The update formula used in the subroutine CONMIN is the BFGS formula

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k} + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \quad (3-36)$$

where  $\mathbf{s}_k$  and  $\mathbf{y}_k$  are given by (3-34) and (3-35), respectively. The derivation of (3-36) can be found in e. g. Luenberger (1984).

The idea behind the limited-memory (memoryless) quasi-Newton methods is that the Hessian matrix is not computed, but a limited number of quasi-Newton corrections are added to the identity matrix. The additional storage required is only the vectors defining the updates. Different methods exist depending on the number of vectors used in the updating. The method used by Shanno (1978a, b) is defined by

$$\begin{aligned} \mathbf{d}_{k+1} = & -\mathbf{g}_{k+1} - \left[ \left( 1 + \frac{\mathbf{y}_k^T \mathbf{y}_k}{\mathbf{p}_k^T \mathbf{y}_k} \right) \frac{\mathbf{p}_k^T \mathbf{g}_{k+1}}{\mathbf{p}_k^T \mathbf{y}_k} - \frac{\mathbf{y}_k^T \mathbf{g}_{k+1}}{\mathbf{p}_k^T \mathbf{y}_k} \right] \mathbf{p}_k \\ & + \frac{\mathbf{p}_k^T \mathbf{g}_{k+1}}{\mathbf{p}_k^T \mathbf{y}_k} \mathbf{y}_k \end{aligned} \quad (3-37)$$

where

$$\mathbf{p}_k = \mathbf{x}_{k+1} - \mathbf{x}_k \quad (3-38)$$

#### 3.4. Computing the step length

In order for the descent method to converge, it is important that the step length produces a “sufficient decrease” in  $J$  at each iteration. The “sufficient decrease” may be satisfied by several different conditions on  $\alpha_k$ . An exact line search would have given

$$\min_{\alpha} J(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \quad (3-39)$$

but in general the solution of the nonlinear function  $\frac{\partial J}{\partial \alpha_k} = 0$  cannot be implemented in a finite number of operations.

The Goldstein-Armijo principle is one method for choosing the step length  $\alpha_k$  so that a sufficient decrease in  $J$  is achieved, Goldstein and Price (1967). The principle states that  $\alpha_k$  should be chosen so that the following is true

$$0 < -\mu_1 \alpha_k \mathbf{g}_k^T \mathbf{d}_k \leq J(\mathbf{x}_k) - J(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq -\mu_2 \alpha_k \mathbf{g}_k^T \mathbf{d}_k \quad (3-40)$$

where  $\mu_1$  and  $\mu_2$  are scalars satisfying

$$0 < \mu_1 \leq \mu_2 \leq 1 \quad (3-41)$$

The upper and lower bounds of (3-40) ensure that  $\alpha_k$  is neither “too large” nor “too small”. It is important to realize that (3-40) alone does not guarantee a good value of  $\alpha_k$ . For instance, choosing  $\alpha_0$  as  $10^{-5}$  would satisfy (3-40) for almost all functions encountered in practice, if  $\mu_1$  and  $\mu_2$  are appropriate small values. This would be an “efficient” way to find the  $\alpha$ 's in that a suitable  $\alpha_k$  would be found with only a single function evaluation per iteration, but the descent method would be extremely inefficient if such a step length was chosen. In choosing the step length it is therefore essential to consider the performance of the algorithm not only in terms of function evaluations per iteration, but also in terms of the decrease in  $J$  at each iteration.

Another way of interpreting  $\alpha_k$  is in terms of univariate minimization which requires that the magnitude of the directional derivative at  $\mathbf{x}_k + \alpha_k \mathbf{d}_k$  be sufficiently reduced from that at  $\mathbf{x}_k$

$$| \mathbf{g}(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k | \leq -\eta \mathbf{g}_k^T \mathbf{d}_k \quad (3-42)$$

where

$$0 \leq \eta \leq 1 \quad (3-43)$$

Since (3-42) does not involve the change in  $J$ , it is not adequate to ensure a sufficient decrease. Many step length algorithms therefore include the following



condition (Gill et al., 1982)

$$J(\mathbf{x}_k) - J(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \geq -\mu \alpha_k \mathbf{g}_k^T \mathbf{d}_k \quad (3-44)$$

where

$$0 < \mu \leq \frac{1}{2} \quad (3-45)$$

The method for finding the optimal step length which will be followed here is a method used by Derber (1985). The first step is to use an initial guess for the step length and then update the parameters in interest according to

$$\mathbf{c}_{k+1} = \mathbf{c}_k + \alpha_i \mathbf{d}_k \quad (3-46)$$

where  $\alpha_i$  is an initial guess for the step length. The model can then be integrated forward with the new estimate of the parameters. The goal will be to maximize the reduction in the cost function between the two model runs.

$$\Delta J = J(\mathbf{x}_k, \mathbf{c}_k) - J(\mathbf{x}_k, \mathbf{c}_k + \alpha_i \mathbf{d}_k) \quad (3-47)$$

Previous experience will give some idea of what the initial value of the step length should be.

The new value of the parameters produce a perturbation in the model representation of the observations. Let  $\delta \mathbf{x}_k$  denote the perturbation. Using the definition of the cost function in (2-10), the perturbation of the cost function as a result of the perturbation to the model representation, leads to

$$\delta J = \int_{\Sigma} \frac{1}{2} (\delta \mathbf{x}_k^T K_x \delta \mathbf{x}_k + 2(\mathbf{x}_k - \mathbf{x}')^T K_x \delta \mathbf{x}_k) d\sigma \quad (3-48)$$

Assume that another choice of the step length,  $\alpha_k$ , would result in a perturbation  $\delta \mathbf{x}'_k$  of the model field

$$\delta \mathbf{x}'_k = \frac{\alpha_k}{\alpha_i} \delta \mathbf{x}_k \quad (3-49)$$

For linear dynamics (as for the case studied here) this would be exactly true, but for nonlinear dynamics it may not be a good approximation. Derber (1985) used the approximation for a nonlinear atmospheric model, and for the cases he studied, (3 – 49) worked satisfactory. (3 – 48) now becomes

$$\delta J = \int_{\Sigma} \frac{1}{2} \left( \frac{\alpha_k^2}{\alpha_i^2} \delta \mathbf{x}_k^T K_x \delta \mathbf{x}_k + 2 \frac{\alpha_k}{\alpha_i} (\mathbf{x}_k - \mathbf{x}')^T K_x \delta \mathbf{x}_k \right) d\sigma \quad (3 - 50)$$

where the prime on  $\delta \mathbf{x}_k'$  has been dropped. The goal is to maximize  $\delta J$ , i. e. the derivative of  $\delta J$  with respect to  $\alpha_k$  is set equal to zero.

$$\frac{\partial \delta J}{\partial \alpha_k} = 0 \quad (3 - 51)$$

which gives

$$\alpha_k = -\alpha_i \frac{\int_{\Sigma} (\mathbf{x}_k - \mathbf{x}')^T K_x \delta \mathbf{x}_k d\sigma}{\int_{\Sigma} \delta \mathbf{x}_k^T K_x \delta \mathbf{x}_k d\sigma} \quad (3 - 52)$$

for the optimal step length. In later iterations  $\alpha_i$  is replaced by  $\alpha_{k-1}$ .

The limited-memory quasi-Newton conjugate gradient algorithm in the subroutine CONMIN is implemented in the experiments described later. A few modifications of the routine have been done. The step length is no longer computed as in the original code, but instead (3 – 52) is used to calculate the new step length. In CONMIN the cost function and its gradient are given in a function call. In the modified version the values of the cost function and its gradient are calculated and written to a file, which is read in the subroutine.

#### 4. Parameter estimation

If the parameter estimation is going to be successful, an important question must be addressed; Under what condition can one expect the proposed estimation method to give unique and stable results? Similar problems have been studied in several papers concerning groundwater flow, e. g. Chavent et al. (1975), Neuman (1980), Yakowitz and Duckstein (1980), Carrera and Neuman (1986a, b, c) and Yeh (1986). Panchang and O'Brien (1988) studied the estimation of the friction factor in tidal rivers. In these papers which consider both time-dependent and steady state flows, important aspects about data assimilation and parameter estimation are discussed. There are four important terms which must have a clear definition: well-posedness, uniqueness, identifiability and stability. These definitions will be given in the following paragraphs.

Strictly speaking, if the method is going to work, the problem should be well-posed. The inverse problem is often ill-posed, but there are situations under which a meaningful solution can be found, although in a limited sense. It will therefore be important to be able to recognize the circumstances which allow a solution of the problem to be found. The inverse problem can be defined as follows: Let a functional relationship  $p(x, y, t) = F(c^2(x, y, t))$  be given between a parameter which in our case is the phase speed  $c^2(x, y, t)$  and  $p(x, y, t)$  where  $p$  represents the variables  $U(x, y, t)$ ,  $V(x, y, t)$  and  $h(x, y, t)$  of our problem. The inverse problem will then be to determine  $c^2$  on the basis of  $p(x, y, t)$  and the inverse relationship  $c^2(x, y, t) = R(p(x, y, t))$ . This problem is said to be well-posed if and only if (1)

to every  $p$  there corresponds a solution  $c^2(x, y, t)$ ; i. e. a solution exists; (2) the solution is unique for any given  $p(x, y, t)$ ; and (3) the solution depends continuously on  $p(x, y, t)$ ; i. e. the solution is stable. If the inverse problem fails to satisfy one or more of these three requirements, it is said to be ill-posed.

Uniqueness can be defined in the following way. If  $c_1^2(x, y, t) = R(p_1(x, y, t))$  and  $c_2^2(x, y, t) = R(p_2(x, y, t))$  are two solutions of the inverse problem, then

$$\| c_1^2(x, y, t) - c_2^2(x, y, t) \| = 0 \implies \| p_1(x, y, t) - p_2(x, y, t) \| = 0 \quad (4-1)$$

where  $\| \cdot \|$  represents a norm over the appropriate space. In practical problems  $p$  is only given at discrete points in space and time, and  $R$  represents a minimization of a functional  $J$  as given in section 2.

Parameter identifiability can be defined as

$$\| p_1(x, y, t) - p_2(x, y, t) \| = 0 \implies \| c_1^2(x, y, t) - c_2^2(x, y, t) \| = 0 \quad (4-2)$$

where  $p_1(x, y, t) = F(c_1^2(x, y, t))$  and  $p_2(x, y, t) = F(c_2^2(x, y, t))$ . While uniqueness refers to the inverse problem,  $R$ , identifiability refers to the forward problem,  $F$ . If two sets of parameters lead to the same function  $p(x, y, t)$ , the parameters are said to be unidentifiable. Uniqueness on the other hand is concerned with the problem whether different parameters may be found from a given  $p(x, y, t)$ , if so the parameters are nonunique.

Stability can be defined in the following way. For every  $\varepsilon > 0$  there exists a  $\delta$  such that for  $c_1^2(x, y, t) = R(p_1(x, y, t))$  and  $c_2^2(x, y, t) = R(p_2(x, y, t))$  one has

$$\| p_1(x, y, t) - p_2(x, y, t) \| < \delta \implies \| c_1^2(x, y, t) - c_2^2(x, y, t) \| < \varepsilon \quad (4-3)$$

(4-3) states that small errors in the variables must not lead to large changes in the computed parameter.

The work done by hydrologists concerning parameter estimation in ground-water flow considers parabolic partial differential equations, since this is the type of equations which hydrologists like to believe governs groundwater flow. Kitamura and Nikagiri (1977) studied the identifiability of continuously varying and constant parameters in a linear one-dimensional parabolic partial differential equation

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left[ a(x) \frac{\partial u}{\partial x} \right] + b(x)u + f(x, t) \quad (4-4)$$

where  $a(x)$  and  $b(x)$  represents the parameters. They assume that the parameters in the system equation are unknown but continuous, together with their first (and for  $a(x)$ ) second derivatives, over the whole flow domain. They also assume that the boundary conditions are known. Their results show that the parameters are only identifiable under certain fairly restrictive conditions.

The problem of identifiability and also stability can be solved by reducing the number of parameters which is going to be estimated. In the hydrological literature the most common way to do this is to approximate the parameters by a known class of functions depending on a finite number of parameters. If the parameter dimension is not reduced, it may be difficult or even impossible to determine the spatial structure of the parameter field. Instability is often characterized by the fact that during the solution process the parameter values are bouncing back and forth between high and low values.

As mentioned in section 2.2 the last term in the cost function

$$J(\mathbf{x}, \mathbf{c}) = \int_{\Sigma} \frac{1}{2} (\mathbf{x} - \mathbf{x}')^T K_{\mathbf{x}} (\mathbf{x} - \mathbf{x}') d\sigma + \int_{\Sigma} \frac{1}{2} (\mathbf{c} - \mathbf{c}')^T K_{\mathbf{c}} (\mathbf{c} - \mathbf{c}') d\sigma \quad (4-5)$$

is added because the interest of this research is parameter estimation, and the term makes sure that the new estimate of the parameters is not too far from the initial guess. In the hydrologist's nomenclature the term represents prior information about the parameters. The definition of uniqueness (4-1), implied that the solution

would be unique if the cost function  $J$  was convex. The requirements for this condition is that the Hessian matrix is positive definite. Writing the cost function as

$$J = J_{\mathbf{x}} + J_{\mathbf{c}} \quad (4 - 6)$$

where  $J_{\mathbf{x}}$  represents the first term in (4 - 5), while  $J_{\mathbf{c}}$  represents the last term or the term representing prior information about the parameters. The Hessian is represented by

$$\frac{\partial^2 J}{\partial \mathbf{c}^2} = \frac{\partial^2 J_{\mathbf{x}}}{\partial \mathbf{c}^2} + \frac{\partial^2 J_{\mathbf{c}}}{\partial \mathbf{c}^2} \quad (4 - 7)$$

The first term gives

$$\frac{\partial^2 J_{\mathbf{x}}}{\partial \mathbf{c}^2} = \int_{\Sigma} K_{\mathbf{x}} \left[ \left( \frac{\partial \mathbf{x}}{\partial \mathbf{c}} \right)^T \left( \frac{\partial \mathbf{x}}{\partial \mathbf{c}} \right) + (\mathbf{x} - \mathbf{x}')^T \frac{\partial^2 \mathbf{x}}{\partial \mathbf{c}^2} \right] d\sigma \quad (4 - 8)$$

which can be positive or negative. So the first term in the minimization criterion does not guarantee that the cost function is convex. The last term in (4 - 7) gives

$$\frac{\partial^2 J_{\mathbf{c}}}{\partial \mathbf{c}^2} = K_{\mathbf{c}} \quad (4 - 9)$$

which is clearly a positive term. Adding prior information about the parameters therefore increases the chance that the cost function will be convex. Of course there is no guarantee that the term in (4 - 9) will make the Hessian positive definite. Carrera and Neuman (1986b) discuss the effect of prior information in a few simple examples of estimation of aquifer parameters. Their examples clearly show that the addition of prior information lead to unique solutions.

## 5. Specification of the problem

In this section the variational formalism developed in section 2 is going to be applied to a model of the equatorial Pacific Ocean. The governing equations are presented and the adjoint equations are derived. The set of equations will be able to assimilate sea level observations into the Pacific Ocean model, and the main interest will be focused on the ability of sea level observations to adjust (estimate) parameters in the model. The parameter which will be estimated is the phase speed. In fact it is the square of the phase speed which will be estimated, but the term phase speed is going to be used. The method is not restricted to the use of only sea level observations, but could easily be modified so it could assimilate other types of observations e. g. velocity data. However, this is not going to be considered in the present study. It is also possible to adjust other parameters than the phase speed, e. g. the eddy viscosity or even adjusting the windstress, i. e., the forcing function, so the model solution fits the observations in an optimal way. In the initial tests of the data assimilation algorithm, it is important to keep the problem under consideration as simple as possible, so the focus of this research will be on estimating only one parameter, the phase speed. As described in section 4, hydrologists have had some experience with parameter adjustment, and one of their results is that it is sometimes necessary to reduce the number of degrees of freedom in order to be able to find unique and stable parameters. The important physics in the model under consideration consist of equatorial waves propagating across the

Pacific Ocean, and to reduce the number of degrees of freedom it is assumed that the parameter is a function of longitude only.

### 5.1. The numerical model

The model to be used is a reduced-gravity, linear transport model for the equatorial Pacific Ocean. Due to the large latitudinal extent, spherical coordinates are used, with  $\phi$  (longitude) increasing eastward and  $\theta$  (latitude) increasing northward. In the reduced gravity model the ocean is assumed to consist of two layers of slightly different density ( $\rho, \rho_2$ ), with the interface being an approximation of the pycnocline. The lower layer of the model is dynamically inactive, while the motion of the upper layer represents the first baroclinic mode. Similar models (but in an equatorial  $\beta$ -plane) have been successfully used to simulate the ocean circulation in the equatorial regions (e. g. Adamec and O'Brien, 1978; Cane, 1979; Busalacchi and O'Brien, 1980).

Let  $U = uh$  and  $V = vh$  represent the eastward and northward component of upper-layer transport, respectively, where  $(u, v)$  are the depth-independent  $(\phi, \theta)$  velocity components in the upper layer and  $h$  the thickness of the upper layer. Figure 1 shows a schematic picture of the reduced gravity model. The model equations are:

$$\frac{\partial U}{\partial t} - fV = -\frac{c^2}{a \cos \theta} \frac{\partial h}{\partial \phi} + \frac{\tau^\phi}{\rho} + A \nabla^2 U \quad (5-1)$$

$$\frac{\partial V}{\partial t} + fU = -\frac{c^2}{a} \frac{\partial h}{\partial \theta} + \frac{\tau^\theta}{\rho} + A \nabla^2 V \quad (5-2)$$

$$\frac{\partial h}{\partial t} + \frac{1}{a \cos \theta} \left[ \frac{\partial U}{\partial \phi} + \frac{\partial}{\partial \theta} (V \cos \theta) \right] = 0 \quad (5-3)$$

where  $\nabla^2$  is given by the simplified form

$$\nabla^2 = \frac{1}{a^2 \cos^2 \theta} \left[ \frac{\partial^2}{\partial \phi^2} + \cos \theta \frac{\partial}{\partial \theta} \left( \cos \theta \frac{\partial}{\partial \theta} \right) \right] \quad (5-4)$$



and  $c^2 = gH \left( \frac{\rho_2 - \rho}{\rho} \right)$  is the barotropic gravity wave speed, which is the parameter that is going to be estimated.  $c^2$  is assumed to be a function of longitude only,  $c^2 = c^2(\phi)$ .  $\rho_2$  and  $\rho$  are the densities of the lower and upper layer, respectively.  $a$  is the radius of the earth,  $f = 2\omega \sin\theta$  is the Coriolis parameter and  $\omega$  is the rotation rate of the earth.  $A$  is a horizontal kinematic eddy viscosity. The wind stress  $\tau = (\tau^\phi, \tau^\theta)$ , is applied as a body force over the upper layer. The values used for some of the different parameters are :  $A = 2000\text{m}^2\text{s}^{-1}$ ,  $\omega = 0.729 \cdot 10^{-4}\text{s}^{-1}$ ,  $a = 6378\text{km}$ . The initial thickness of the upper layer is 300m.

The model geometry extends from 120°E to 75°W and from 20°S to 25°N as in Kubota and O'Brien (1988), see Figure 2. The effect of the coast at the eastern and the western end of the basin is included in the model by simulating the coastline as a closed boundary, i. e.  $U = V = 0$ . The northern and the southern boundaries are open and the open boundary condition described by Camerlengo and O'Brien (1980) is applied along these boundaries.

In solving the equations (5 – 1) – (5 – 3) the staggered C-grid of Arakawa (1966) is used, see Figure 3. The gridspacing is  $\frac{1}{4}^\circ$  in longitude and  $\frac{1}{2}^\circ$  in latitude. The equations are integrated in time using a leapfrog scheme, with a forward scheme used every 99<sup>th</sup> time step to eliminate the computational mode. The timestep is 40 minutes. A Dufort-Frankel scheme is used for the diffusive term.

## 5.2. The adjoint equations

The formalism developed in section 2 can now be followed using the equations for the Pacific Ocean model described above as constraints. It is necessary to define a cost function  $J$ , which measures the distance between the observations and the model solution.  $J$  will be chosen as the quadratic function

$$J(h, c^2) = \int_{\Sigma} \left[ \frac{K_h}{2} (h - h')^2 + \frac{K_c}{2} (c^2 - c^{2'})^2 \right] d\sigma \quad (5 - 5)$$

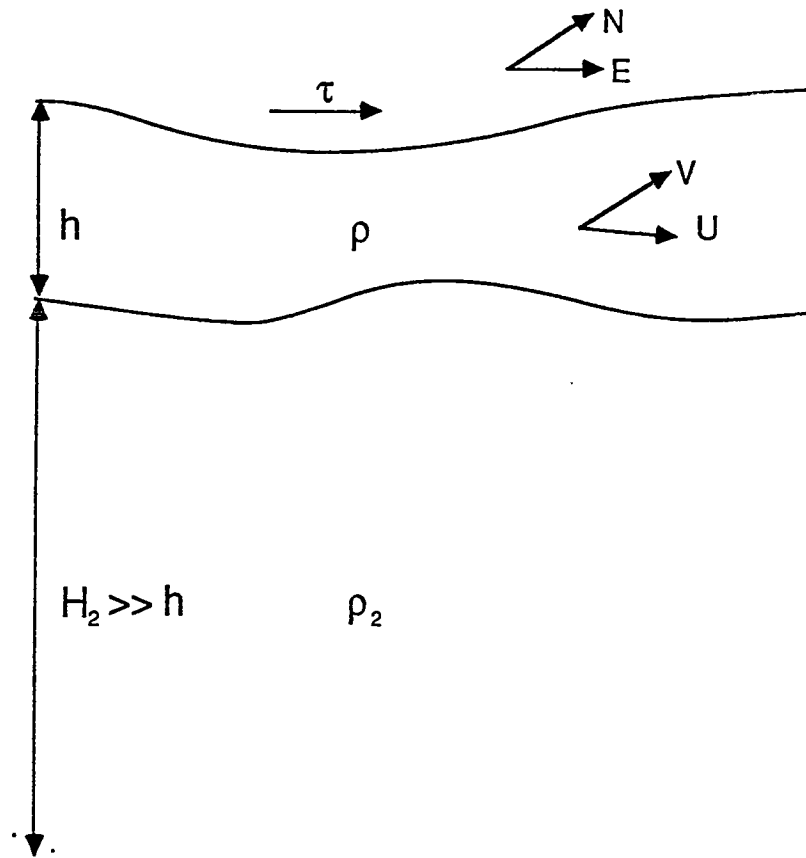


Figure 1. The geometry of the reduced gravity model

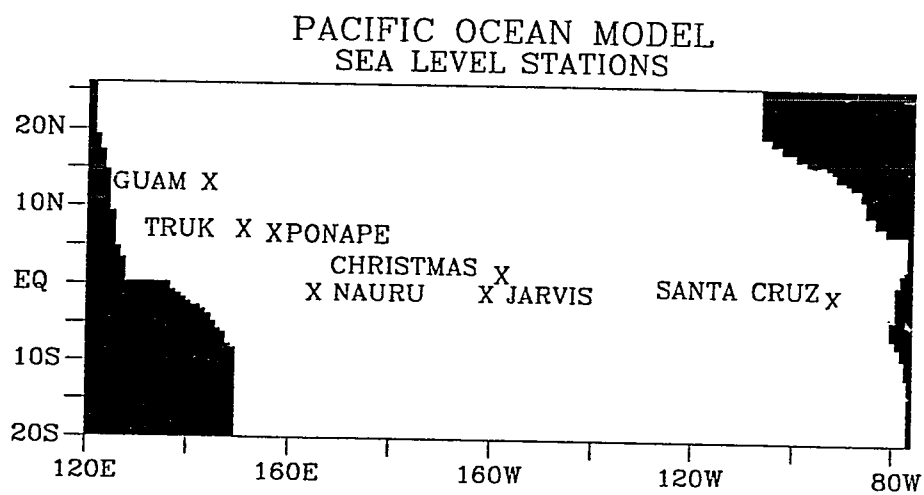


Figure 2. The Pacific Ocean model domain. The southern and northern boundaries are open. The position of the sea level stations used in the assimilation is also shown.

where  $h'$  represents an observation of the upper layer thickness and  $c^{2'}$  is an estimate of the phase speed. It has been assumed that the observational errors are uncorrelated.  $K_h$  and  $K_c$  are validity coefficients.  $\Sigma$  represents the spatial and temporal domain over which the Pacific model is integrated. The last term in (5-5) makes sure that the final value of the phase speed is not too far from the initial estimate. This is important since there are certain physical limits for  $c^2$ . The initial estimate should be a reasonable one so that the minimization process just has to make a "small" correction to the initial guess.

The next step is to form the associated Lagrange function

$$\begin{aligned}
L(U, V, h, \lambda_u, \lambda_v, \lambda_h, c^2) = & \\
& \int_{\Sigma} \lambda_u \left[ \frac{\partial U}{\partial t} - fV + \frac{c^2}{a \cos \theta} \frac{\partial h}{\partial \phi} - \frac{\tau^\phi}{\rho} - A \nabla^2 U \right] d\sigma \\
& + \int_{\Sigma} \lambda_v \left[ \frac{\partial V}{\partial t} + fU + \frac{c^2}{a} \frac{\partial h}{\partial \theta} - \frac{\tau^\theta}{\rho} - A \nabla^2 V \right] d\sigma \quad (5-6) \\
& + \int_{\Sigma} \lambda_h \left[ \frac{\partial h}{\partial t} + \frac{1}{a \cos \theta} \left( \frac{\partial U}{\partial \phi} + \frac{\partial}{\partial \theta} (V \cos \theta) \right) \right] d\sigma \\
& + \int_{\Sigma} \frac{K_h}{2} (h - h')^2 d\sigma + \int_{\Sigma} \frac{K_c}{2} (c^2 - c^{2'})^2 d\sigma
\end{aligned}$$

where  $\lambda_u$ ,  $\lambda_v$  and  $\lambda_h$  are the Lagrangian multipliers for  $U$ ,  $V$ , and  $h$ , respectively. The details of the derivation of the appropriate continuous equations are given in appendix A. The stationary points of the Lagrange function (which correspond to the minimum of the cost function) can be found by letting the first variation of  $L$  with respect to the variables  $U$ ,  $V$ ,  $h$ ,  $\lambda_u$ ,  $\lambda_v$ ,  $\lambda_h$  and  $c^2$  vanish. The first variation of  $L$  with respect to  $\lambda_u$ ,  $\lambda_v$  and  $\lambda_h$  gives the following equations:

$$\frac{\partial U}{\partial t} - fV = -\frac{c^2}{a \cos \theta} \frac{\partial h}{\partial \phi} + \frac{\tau^\phi}{\rho} + A \nabla^2 U \quad (5-7)$$

$$\frac{\partial V}{\partial t} + fU = -\frac{c^2}{a} \frac{\partial h}{\partial \theta} + \frac{\tau^\theta}{\rho} + A \nabla^2 V \quad (5-8)$$

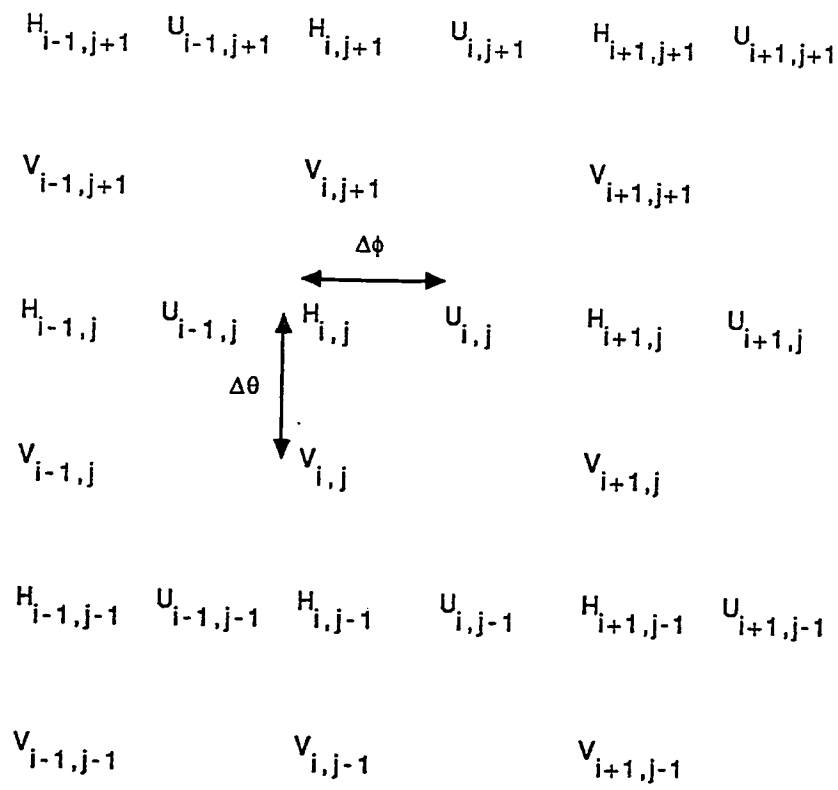


Figure 3. The Arakawa C-grid

$$\frac{\partial h}{\partial t} + \frac{1}{a \cos \theta} \left[ \frac{\partial U}{\partial \phi} + \frac{\partial}{\partial \theta} (V \cos \theta) \right] = 0 \quad (5-9)$$

which are identical to the original model equations (5-1) – (5-3). Letting the first variation of  $L$  with respect to  $U$ ,  $V$  and  $h$  vanish, gives the adjoint equations:

$$-\frac{\partial \lambda_u}{\partial t} + f \lambda_v = \frac{c^2}{a \cos \theta} \frac{\partial \lambda_h}{\partial \phi} + A \nabla^2 \lambda_u \quad (5-10)$$

$$-\frac{\partial \lambda_v}{\partial t} - f \lambda_u = \frac{c^2}{a} \frac{\partial \lambda_h}{\partial \theta} + A \nabla^2 \lambda_v \quad (5-11)$$

$$-\frac{\partial \lambda_h}{\partial t} - \frac{1}{a \cos \theta} \left[ \frac{\partial \lambda_u}{\partial \phi} + \frac{\partial}{\partial \theta} (\lambda_v \cos \theta) \right] + K_h (h - h') = 0 \quad (5-12)$$

It is worth noting that the adjoint equations are forced by the data misfits represented by the last term in equation (5-12). The windstress does not enter into these equations. Comparing equations (5-10) – (5-12) to the original model equations (5-7) – (5-9) one can see they have a similar form, except that the adjoint equations correspond to an evolution backward in time. The Lagrangian multipliers will therefore propagate information about the data misfits back to the initial time of the data assimilation period. Since the system of equations governing the Lagrangian multipliers are similar to the model equations, the information will propagate in the form of equatorial Kelvin and Rossby waves. However, in the adjoint equations the equatorial Kelvin wave is going to have a westward phase propagation, while the Rossby waves will have an eastward phase propagation.

In the derivation of the adjoint equations, the natural boundary conditions for the adjoint variables follow. The “initial” conditions at  $t=T$  for the adjoint variables are homogeneous ones, namely  $\lambda_u = \lambda_v = \lambda_h = 0$ . The adjoint equations are forced by the data misfits, and after the last computational time level for the physical model there are no data.  $\lambda_u = \lambda_v = \lambda_h = 0$  is therefore the natural initial condition for the adjoint variables. At the solid boundaries in the western and eastern part of the basin, the boundary conditions are given by  $\lambda_u = \lambda_v = 0$ . These

conditions are identical to the no-slip conditions used in the model equations,  $U = V = 0$  at the solid boundaries. Using the open boundary conditions of Camerlengo and O'Brien (1980) for  $U$  and  $V$ , the conditions for the adjoint variables can be derived. It turns out that identical open boundary conditions must be used for  $\lambda_u$  and  $\lambda_v$ .

Letting the first variation of  $L$  with respect to  $c^2$  vanish

$$\int_t \int_{\theta} \left[ \frac{\lambda_u}{c^2 a \cos \theta} \frac{\partial h}{\partial \phi} + \frac{\lambda_v}{c^2 a} \frac{\partial h}{\partial \theta} \right] d\theta dt + T L_{\theta} K_c (c^2 - c^{2'}) = 0 \quad (5 - 13)$$

which gives the gradient of  $L$  with respect to  $c^2$ .  $T$  and  $L_{\theta}$  are the period over which the data are assimilated and a length scale in the latitudinal direction, respectively.

It is important to determine the correct adjoint equations. If there is an error in the equations calculating the Lagrange variables, the minimization process may converge to a solution which is not the one which is sought. The model equations are going to be solved using finite differences. One has to be careful when going from the continuous equations to the finite difference representation. Deriving the adjoint of the continuous equations and then taking the finite difference of the result may not give the desired adjoint that is the adjoint of the finite difference version of the model. The adjoint of the finite difference equations are therefore derived in Appendix B.

The temporal domain in which the cost function (5 - 5) is going to be minimized must be determined. The spatial domain is given by the spatial domain of the model under consideration. It is not so clear what to choose for the time interval. Numerical models are not perfect representations of the real world, and it would be meaningless to try to assimilate data over too long time intervals. Different models will require different assimilation periods. The time scales of the model is an important factor. For instance in a meteorological forecast model it would be

natural to choose an assimilation period of a few days, since the forecast is not very accurate after this period. On the other hand, for an oceanographical model such as the reduced gravity model used in this study, it would not be correct to choose a period of a few days. The major time scales of the motion in this type of model are of the order of months. A natural assimilation period would therefore be of the order of months too. In the initial experiments described in the next section a period of six months was chosen. When the model is driven by real wind forcing, it is important that the information have time to propagate throughout the model. A Kelvin wave takes about 3 months to cross the Pacific Ocean, while the first symmetric mode Rossby waves need 9 months to cross. A data assimilation period of 12 months was therefore chosen in this case.

### *5.3. The wind forcing*

The wind stress data used as forcing of the Pacific Ocean model come from ship observations. The individual wind observations have been grouped into monthly values on a  $2 \times 2^\circ$  grid (see Legler and O'Brien (1986) for more details). From this coarse grid the data are interpolated to the model grid using a cubic spline interpolation. Figure 4 shows an example of the interpolated wind stress for June 1979. The quantity plotted is the pseudostress, i. e. the magnitude of the wind velocity times the component.

### *5.4. The observations*

Sea surface elevation (SSE) observations from island stations are going to be assimilated into the model. The model does not include sea level as one of the dependent variables, and it will therefore be necessary to transform SSE to upper layer thickness (ULT). The SSE data will be transformed to ULT by dividing by



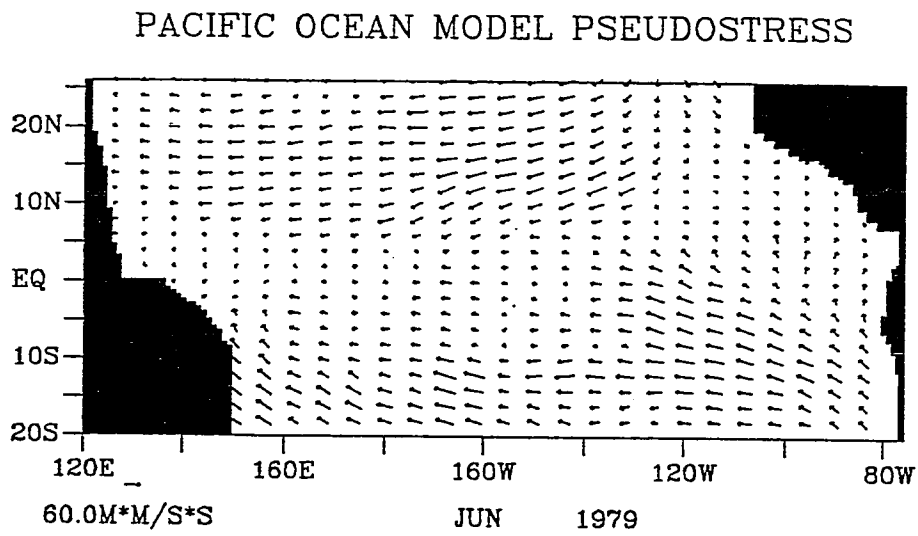


Figure 4. Pseudostress for the Pacific Ocean model for June 1979.

the density difference between the two layers of the model, i. e.

$$\text{ULT} = \frac{\text{SSE}}{\frac{\Delta\rho}{\rho}} \quad (5 - 14)$$

Time series of daily observed SSE from the island stations are going to be used. These observations will contain both high and low frequencies. The main interest of this study is to be able to determine the basic stratification of the Pacific Ocean, and the high frequencies will not give important information for this purpose. The time series are therefore filtered using a low pass filter before the data are assimilated into the model. As mentioned above the wind stress used to force the Pacific Ocean model, is a monthly average. The shortest time scale resolved by the model is therefore 2 months. The sea level data are as a consequence of this low pass filtered with a pass period of 60 days. As an example of the time series of sea level observation converted into ULT, Figure 5a shows ULT as a function of time for the station at Santa Cruz. The figure shows the daily observation for the period 1978-1983. Figure 5b is a plot of the filtered time series for the same period, using a pass period of 60 days. Missing data have been filled in by linear interpolation or if data are missing at the beginning or end of the time series, persistence is used. This is not going to influence the assimilation in this research, since the periods considered here did not have any missing data points.

The Pacific Ocean model saves time series of the height field once a day for the island stations in Figure 2. The  $U$ ,  $V$  and  $h$  fields are saved every 6 days for the whole domain. Similarly for the adjoint equations,  $\lambda_u$ ,  $\lambda_v$  and  $\lambda_h$  fields are saved every 6 days. The  $h$ ,  $\lambda_u$  and  $\lambda_v$  fields are used when the gradient of  $L$  with respect to  $c^2$  is calculated by equation (5 - 13). Saving the variables every 6 days may seem not to be often enough. But as mentioned earlier, the main interest of this study is to be able to determine the large scale variation in the Pacific Ocean

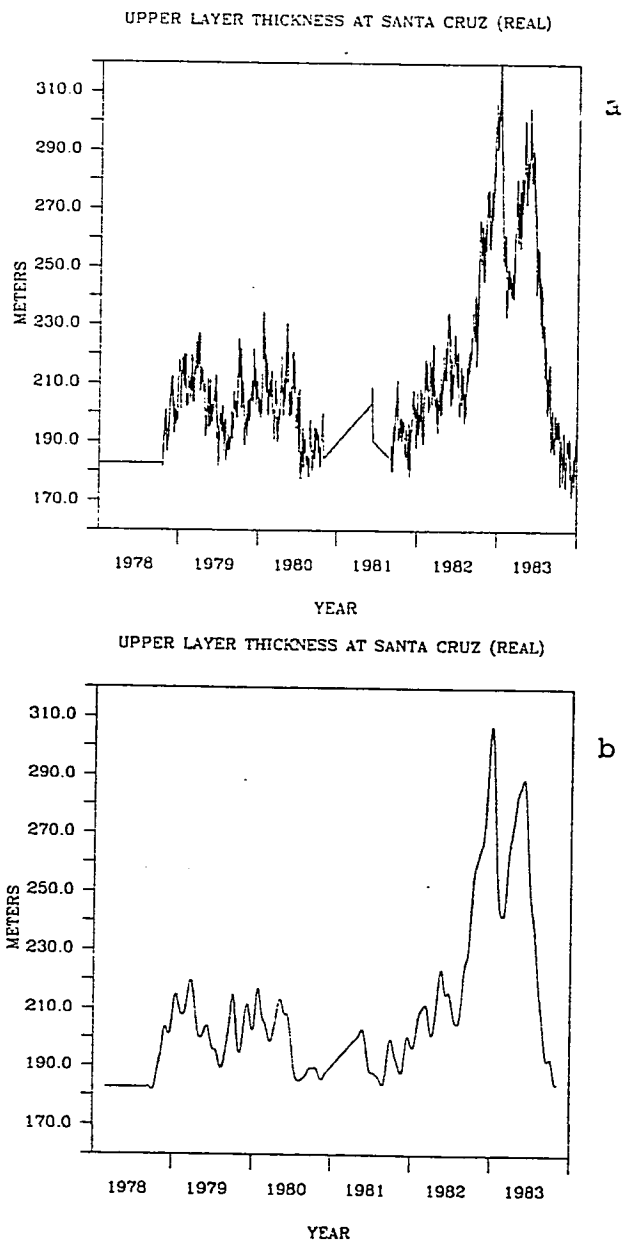


Figure 5. Time series of upper layer thickness (ULT) for the station at Santa Cruz. ULT in meters is plotted as a function of time. a) Daily observations of ULT. b) The filtered time series using a pass period of 60 days.

and the time integral in (5 – 13) will be adequately approximated using data every 6 days.

The adjoint equations are forced by the difference between the model solution and the observations, the data misfits. The timestep in the adjoint model is the same as in the original model, 40 minutes. In order to obtain a smooth forcing of the adjoint equations, the data misfits are linearly interpolated so that the forcing is a smooth function of time.

## 6. Data assimilation in the Pacific Ocean model

As initial experiments of data assimilation in the Pacific Ocean model, a few simple examples are going to be discussed. The primary objective of this research is to be able to determine the phase speed in the model. Adjustment of the external forcing is not considered here, and in the first experiments there is no wind stress applied. The Pacific Ocean model is instead initialized by the height and velocity field of a Kelvin wave. It is important to be able to check the performance of the data assimilation algorithm. Identical twin data are therefore going to be used in the preliminary experiments, that is the “observations” are results from the model. The simple set up of an initial Kelvin wave gives a nice picture of how the adjoint equations propagate information throughout the model domain. The number of waves present is limited, and it is relatively easy to explain what is happening.

After the initial test of the algorithm, the estimation of the phase speed in “real” model runs is considered. The Pacific Ocean model was spun up from rest using the wind stress described in section 5.3. The integration started in 1972, and the model was run through 1983. As an initial guess for the parameter  $c^2$ , a value of  $6.0\text{m}^2\text{s}^{-2}$  was used. This value corresponds to what Kubota and O’Brien (1988) used in their calculations. A few experiments using identical twin data are considered initially. Since  $c^2$  gives information about the basic stratification of the Pacific Ocean, it would be interesting to be able to determine  $c^2$  for two different periods. First, a “normal” year is going to be used as observations. The year 1979

was chosen to represent what is going to be called the normal situation. Second, the situation in the Pacific Ocean changes dramatically during an El Nino event and to investigate the effect of a major climatic event on the estimated value of the parameter, 1982/1983 was chosen as an example of an El Nino year. The El Nino of 1982/83 is one of the strongest on record in this century.

### 6.1. *The Pacific Ocean model initialized by a Kelvin wave*

In the first experiment the Pacific Ocean model was initialized by the height and velocity field of a Kelvin wave. The model was integrated forward for 180 days using a constant phase speed  $c^2 = 6.0\text{m}^2\text{s}^{-2}$ . The results from this run are going to represent the “observations”. The value of the phase speed was then changed to  $c^2 = 4.0\text{m}^2\text{s}^{-2}$  and the model was again integrated forward for 180 days to create the model results. In Figure 6a-d the results of the model using  $c^2 = 6.0\text{m}^2\text{s}^{-2}$  are shown. The figures show the upper layer thickness (ULT) and the velocity field. Upper layer thickness is represented by contour lines, while the velocity components  $U$  and  $V$  are represented by arrows. The initial Kelvin wave at day 0 can be seen in Figure 6a. Figure 6b shows the wave at day 30. The wave has now hit the eastern boundary of the model, and coastal Kelvin waves are propagating northward and southward along the coastline. The open boundaries at the south and north boundary let the coastal Kelvin waves leave the model domain. The Rossby waves created by the reflection can also be observed. These waves can be more clearly seen in Figure 6c, which shows the ULT and velocity field at day 90. The model results after 180 days are shown in Figure 6d. The Rossby waves have now propagated into the middle of the basin. Similar results are obtained when a phase speed of  $c^2 = 4.0\text{m}^2\text{s}^{-2}$  is used. These results are not shown. Experiments with a different assimilation period have shown that the obtained results are not dependent on the data assimilation period in these initial experiments.

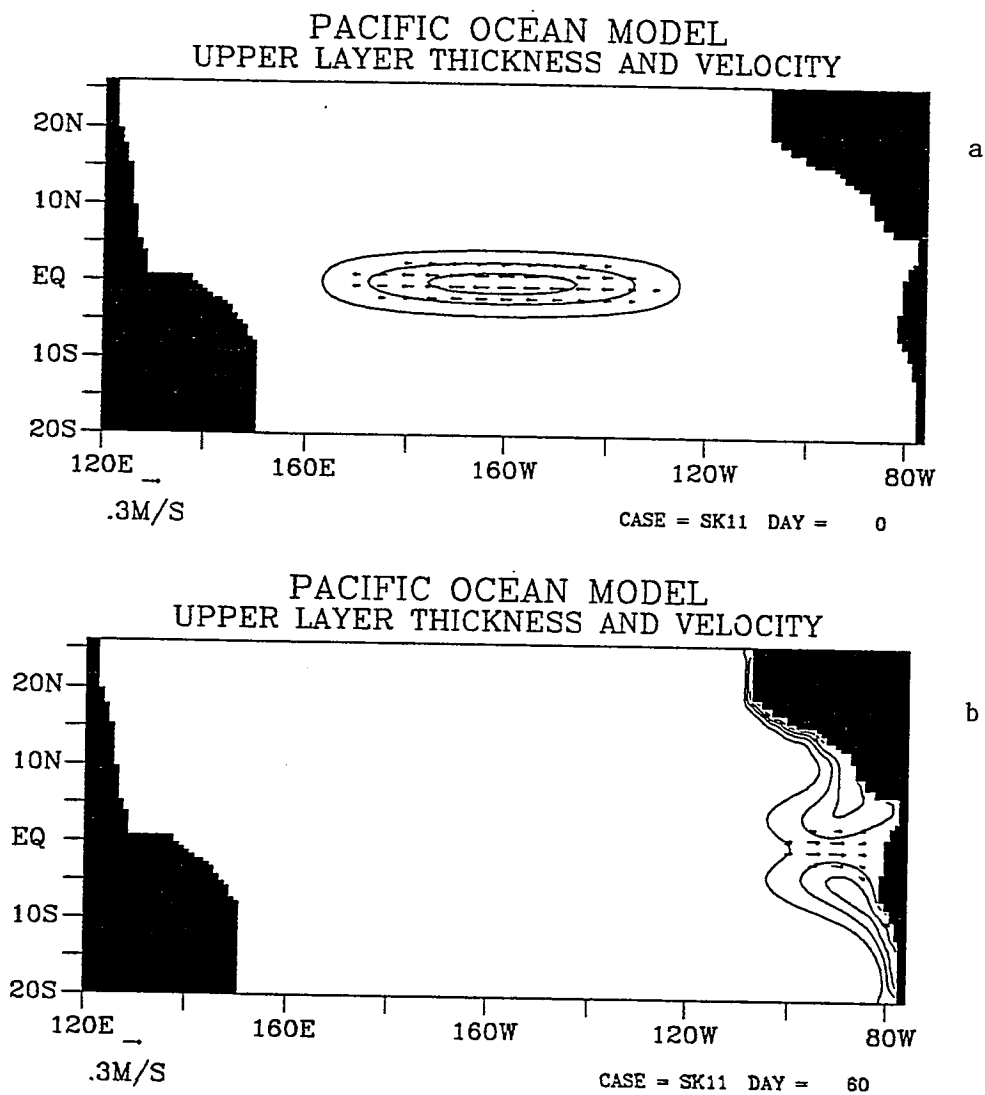


Figure 6. Plot of the upper layer thickness (ULT) and velocity field from the Pacific Ocean model. A phase speed  $c^2 = 6.0\text{m}^2\text{s}^{-2}$  was used. The initial ULT was 300 meter. Contour interval is 10 meter. The Figure shows the results at a) day 0, b) day 60,

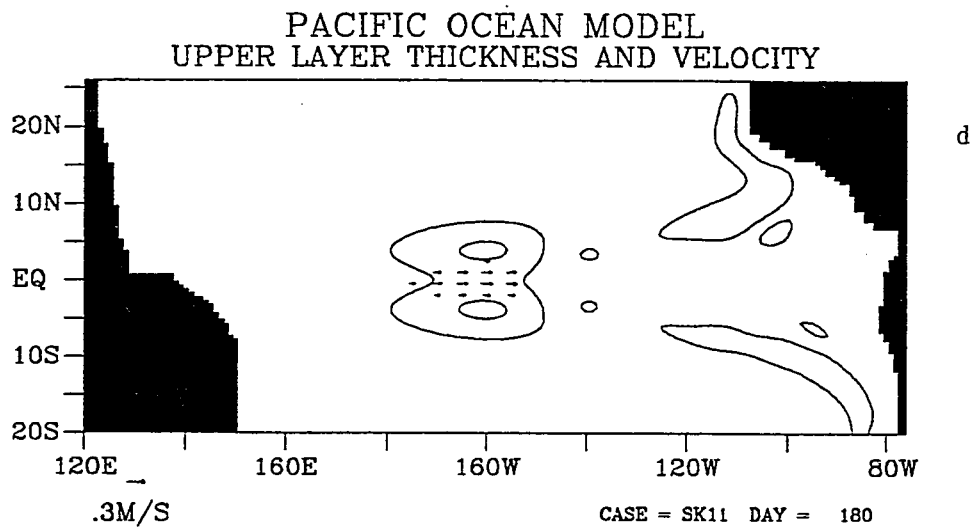
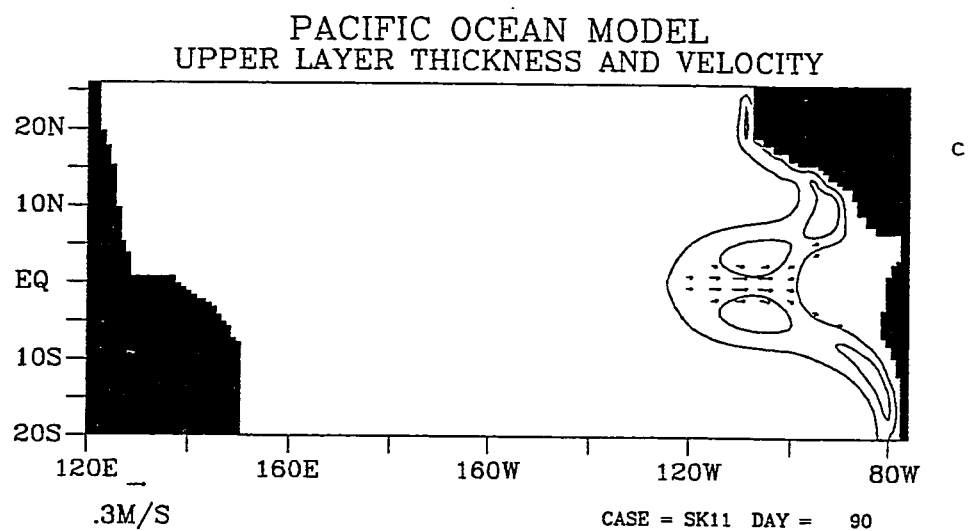


Figure 6. c) day 90, d) day 180.



*6.1.1. Experiment 1: Constant phase speed and observations everywhere*

The experiment considered assumes the best possible situation which can occur, that is observations of ULT are available everywhere in space. The observations are assumed to be perfect, i. e. they are not contaminated by observational errors. The question of what effect such errors will have on the estimation process is going to be addressed later.

The next step in the variational data assimilation algorithm is to calculate the data misfits between the two model runs described above. The linearly time interpolated data misfits (in meters) for the model gridpoint corresponding to the island station at Santa Cruz are shown as a function of time (days), in Figure 7. The figure shows the misfits during the first iteration. Note that day 180 corresponds to the end of the integration period of the model, which is the “initial” time for the integration of the adjoint equations.

The adjoint equations are integrated “backward” for 180 days using the data misfits as forcing. In this case there is a forcing term at every gridpoint of the model. Figure 8a-d shows the time evolution of the fields of the Lagrangian variables  $\lambda_u$ ,  $\lambda_v$  and  $\lambda_h$ . In Figure 8a the three fields are shown after the adjoint equations have been integrated for 30 days. This corresponds to day 150 in the forward model. As described earlier, in the adjoint model Kelvin and Rossby waves move in opposite directions compared to the model itself. From the figures it is apparent that the Rossby waves indeed move eastward (Figure 8a and b), while the Kelvin wave caused by the reflection of the Rossby waves is seen to move westward, see Figure 8c and d. It is interesting to note that the higher mode Rossby waves propagate in front of the gravest mode Rossby wave in Figure 8a.

The next step in the assimilation procedure is to calculate the gradient of the cost function with respect to the phase speed using (5 – 13). In Figure 9

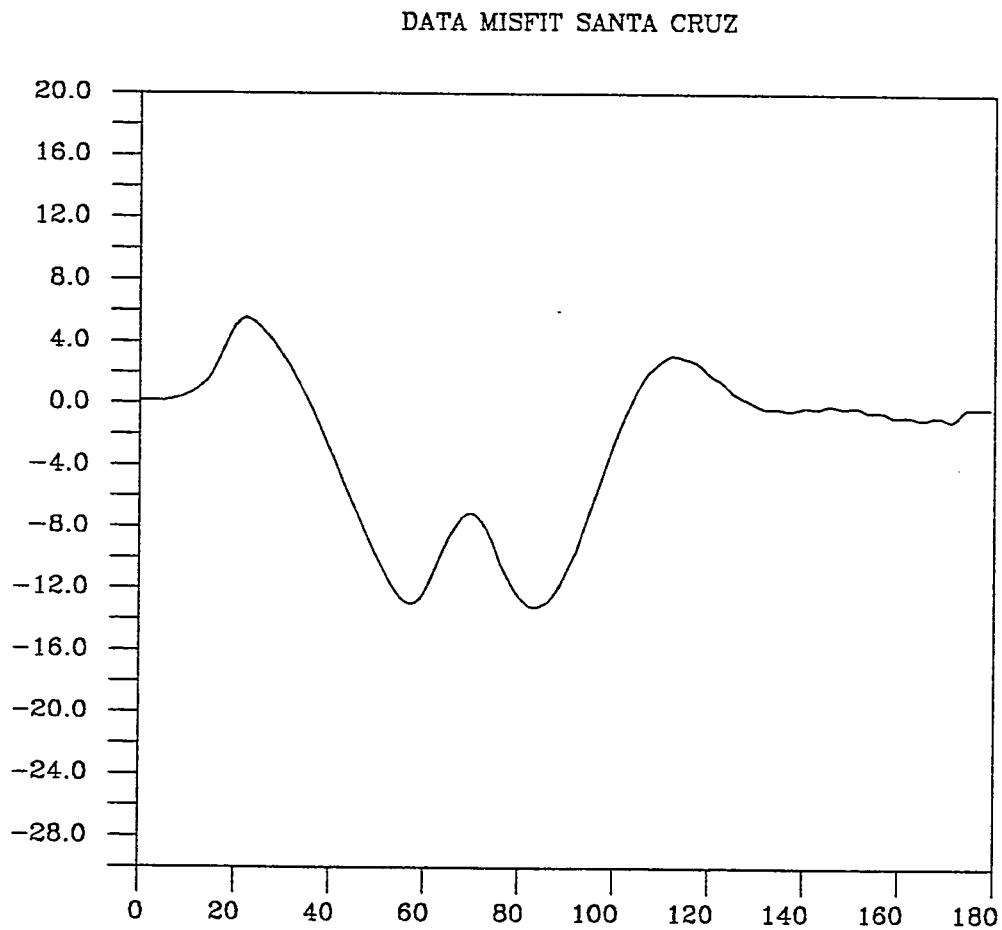


Figure 7. Data misfits during the first iteration for the model gridpoint corresponding to the island station at Santa Cruz. Misfits in meters are shown along the vertical axis, while time in days is plotted along the horizontal axis.

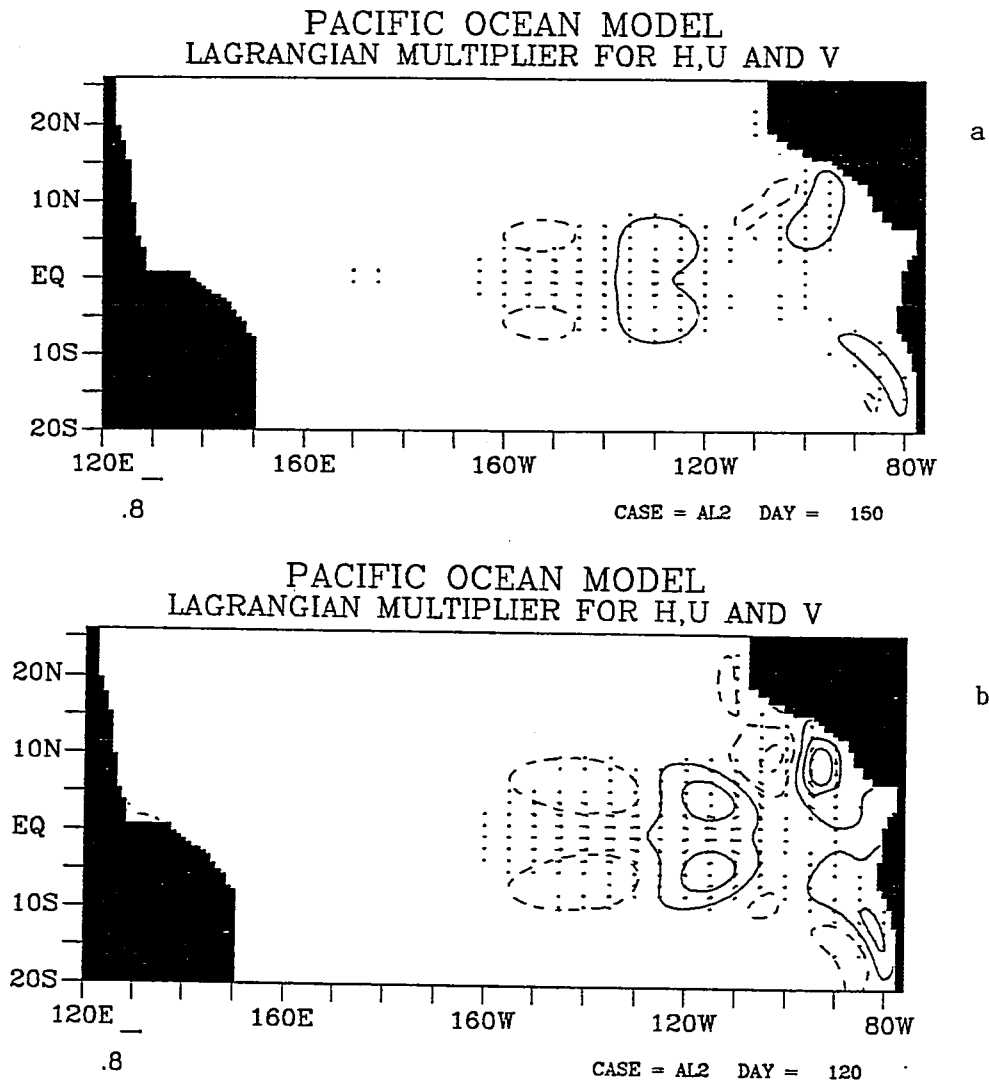


Figure 8. Plot of the Lagrangian multipliers for the case where observations are available everywhere. The Lagrangian multiplier for the  $h$  equation is shown using contours, while the Lagrangian multipliers for the  $U$  and  $V$  equations are represented by vectors. The scale of the vectors are shown in the lower left corner of each figure. Contour interval is 10.0. The Figure shows the results at a) at day 150, b) day 120,

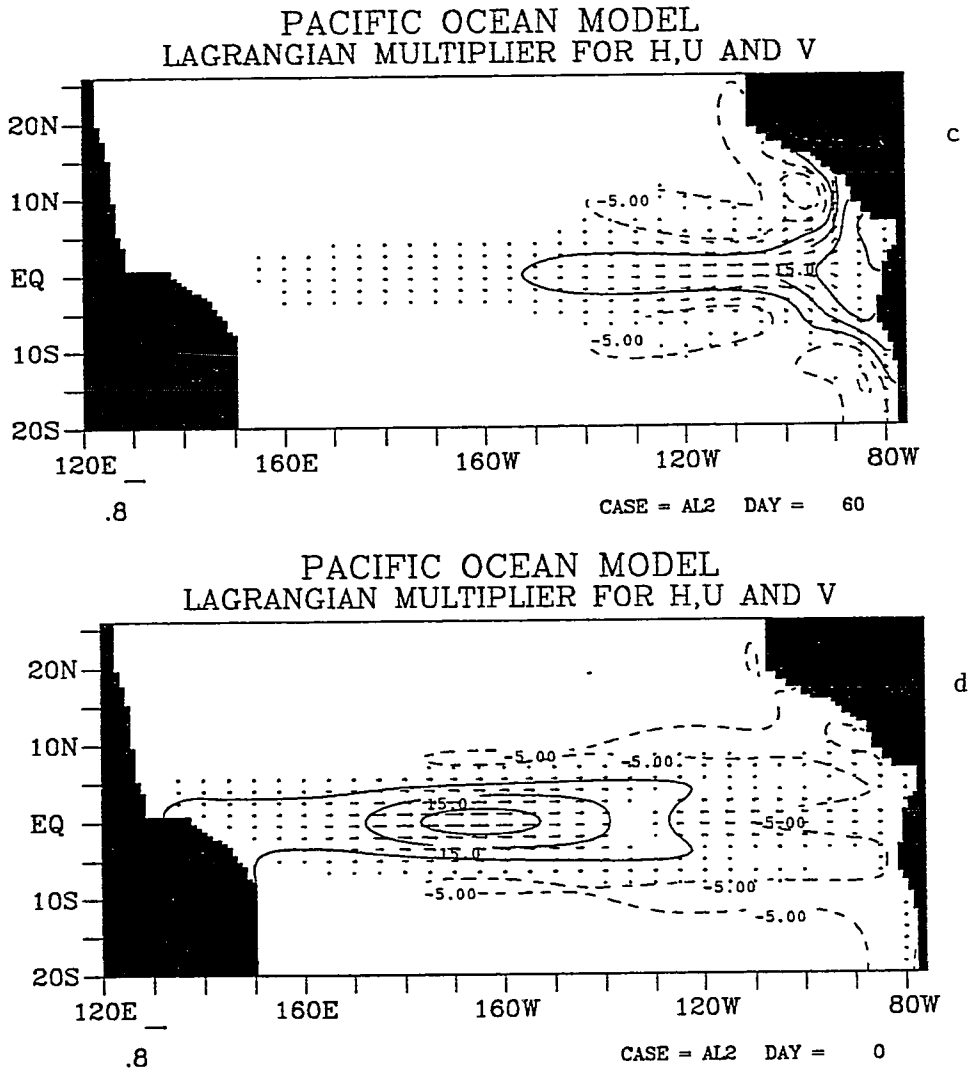


Figure 8. c) day 60, d) day 0.

the time and longitudinal variation are shown. The latitudinal dependence has been integrated out. The figure shows the results after the first iteration. It is worth noting how the information about the gradient propagates in time and space. Initially, that is at the end of the assimilation period, the information propagates from west to east with the Rossby wave speed. Calculating the phase speed of the propagation gives a value of  $c = 0.64\text{ms}^{-1}$ , which is very close to the mode 1 Rossby wave phase speed. The Rossby waves connected with the eastward propagation can be seen in Figure 8a and b. After the information hits the eastern boundary, the information propagates westward as a Kelvin wave. Again calculating the phase speed from Figure 9, the westward propagation can be found in Figure 8c and d. The Kelvin wave is about to leave the eastern end of the basin in Figure 8c, which shows the results at day 60. Looking at Figure 9 again, day 60 is exactly when the information about the gradient starts to propagate westward. At day 0 the information has come all the way to  $160^\circ\text{E}$ . The initial conditions for the model, see Figure 6a, show that the gradient of the Kelvin wave height field is close to zero west of  $160^\circ\text{E}$ . This gradient is a part of the integral determining the gradient of the cost function with respect to  $c^2$ . The information does not have a chance to propagate further west because of the chosen initial conditions. As will be discussed later, this has an important effect on the ability to determine the spatial structure of the phase speed. Of course, in a real situation this limitation will not occur.

The simplest test is to determine a constant phase speed. The space and time dependence of the gradient of  $L$  is removed by integrating over the remaining space direction (longitudinal) and over the time interval of the data assimilation. Since  $c^2$  and the gradient of the cost function is a constant, a steepest descent minimization algorithm was used. The step length was calculated according to (3 – 52). After a new value of the phase speed was determined, the iterative

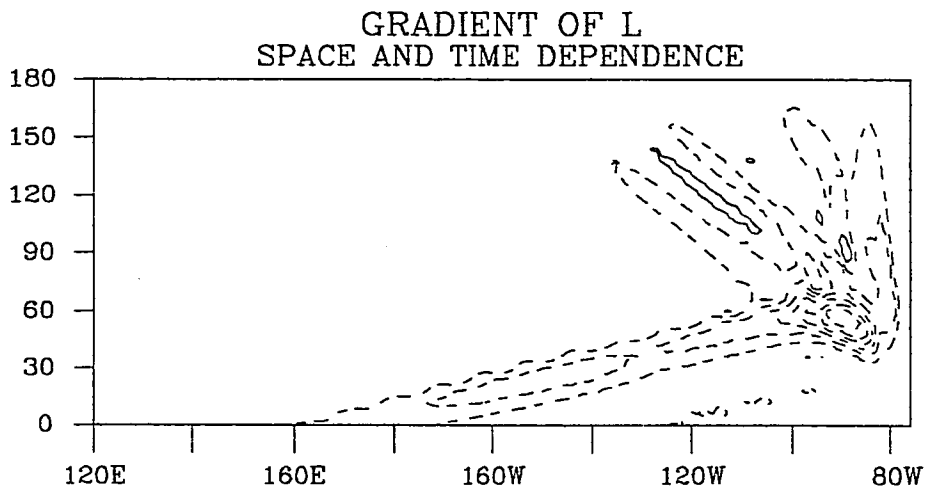


Figure 9. Plot of the longitudinal and time variation of the integral in equation (5-13) after the first iteration. Dashed lines represents negative values of the gradient, while solid lines represents positive values. The gradient has been scaled by its maximum value. Contour interval is 0.15.

procedure continued. The model was integrated forward using the new value of the phase speed and the adjoint model was integrated backward so that a new value of the gradient of the cost function could be calculated. A new minimization was performed and the convergence of the algorithm could be checked. As a convergence criterion  $\frac{\|g\|}{\|g_0\|} < 1.0 \cdot 10^{-2}$  was used in this and the following experiments. The results of the iterative process can be seen in Figure 10a-c. Figure 10a shows the value of the cost function, normalized by its initial value, as a function of the number of iterations. As can be observed during the first few iterations, there is a rapid decrease in the cost function. After 3 iterations  $\frac{J}{J_0}$  has dropped to about  $10^{-3}$ , i. e. the cost function has been reduced by 3 orders of magnitude. In Figure 10b the value of the gradient of the cost function, normalized by its initial value, is shown. The gradient also has a rapid decrease in magnitude during the first couple of iterations. After 4 iterations the value of the gradient has decreased by 3 orders of magnitude, just as the cost function itself. The change in the value of the parameter  $c^2$  during the iterations is shown in Figure 10c. After 4 iterations the value of  $c^2$  is  $6.0\text{m}^2\text{s}^{-2}$ , which is equal to the exact value.

Setting  $c^2$  and thus the gradient of the cost function to be a constant, is the simplest scenario which can occur. To determine  $c^2$  as a function of longitude is a much more difficult problem. Instead of treating the gradient of the cost function as a constant, the gradient is allowed to vary with longitude, but not with time. The time dependence of the gradient was therefore removed by integrating over the time interval of the assimilation. To find the optimal  $c^2$  the modified version of the limited-memory conjugate gradient subroutine CONMIN of Shanno and Phua (1980) described in section 3, was used. This subroutine gave very good convergence rates for the experiments performed in this research. The algorithm converged in ten iterations or less for all the cases considered. Compared to the simple method of

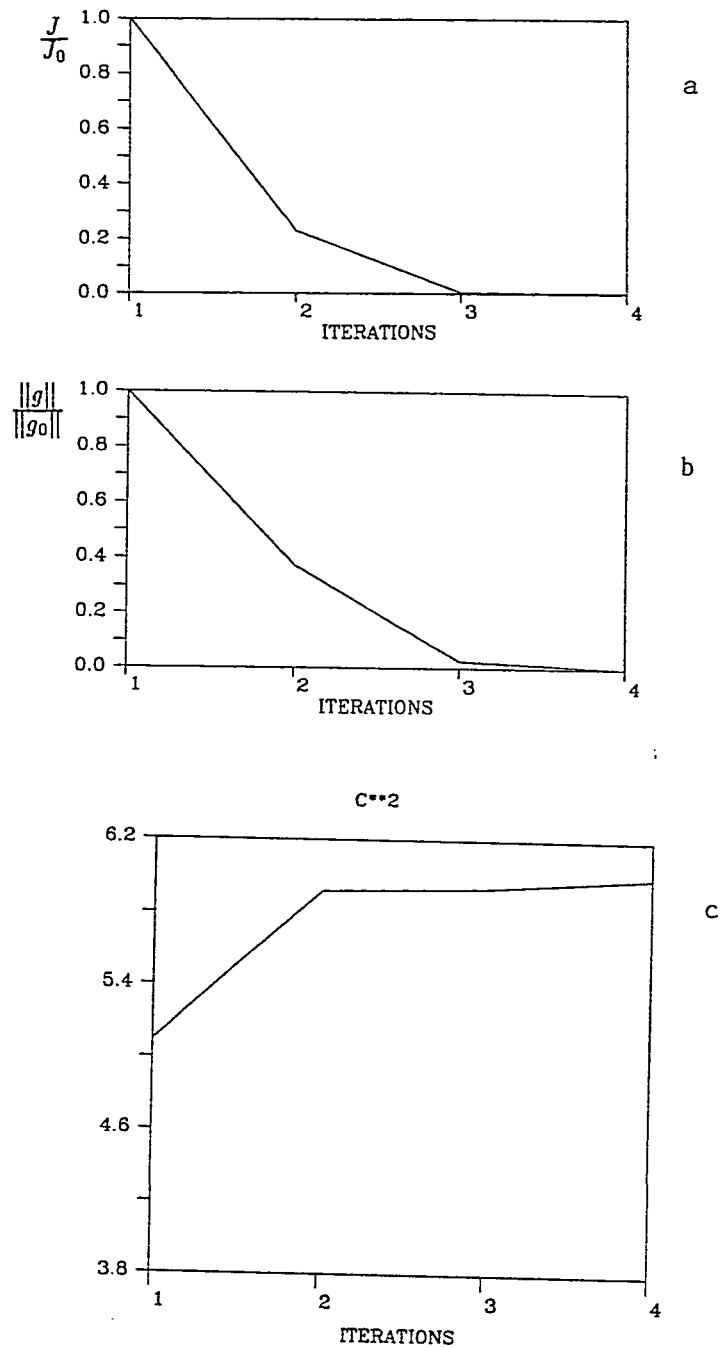


Figure 10. The case when the gradient of the cost function is not a function of longitude. a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\|g\|}{\|g_0\|}$  is plotted as a function of the number of iterations. c) The value of  $c^2$  is plotted as a function of the number of iterations.



steepest descent, the number of iterations dropped by a factor of about three. Using the quasi-Newton method with BFGS-updates did not change the convergence rate appreciably. For most cases the limited-memory version of the subroutine converged faster. The difference between the two methods was never more than two iterations. The step length was calculated using the expression in (3 – 52).

The main interest of this research is to determine the large scale variation of the phase speed which will be a representation of the basic stratification of the Pacific Ocean. The short length scale variation in the estimated values of  $c^2$  are therefore removed by applying a filter to the solution. If these short scale variations were not removed, the estimated values of  $c^2$  varied rapidly with longitude. As the hydrologists have experienced, this kind of instability can be overcome by reducing the parameter dimension. Restricting the attention to the large scale variation of  $c^2$  is of course one way of reducing the dimension of the parameter.

The results of the iterative process are shown in Figure 11a and b and Figure 12a-d. The algorithm converged after 7 iterations with a convergence criterium of  $\frac{\|g\|}{\|g_0\|} < 1.0 \cdot 10^{-2}$ . The cost function normalized by its initial value as a function of iterations, is shown in Figure 11a. There is a rapid decrease during the first three iterations. After four iterations the value of the cost function has decreased an order of magnitude. The value does not decrease much during the next couple of iterations. In Figure 11b the normalized value of the gradient of the cost function is shown. The gradient also experiences a rapid decrease in the beginning. After seven iterations the value of the gradient has decreased below the convergence criterium.

The evolution of the estimated value of the phase speed is shown in Figure 12a-d. Figure 12a is a plot of the longitudinal variation of the phase speed after the first iteration. As can be seen there is a relatively strong correction in the eastern

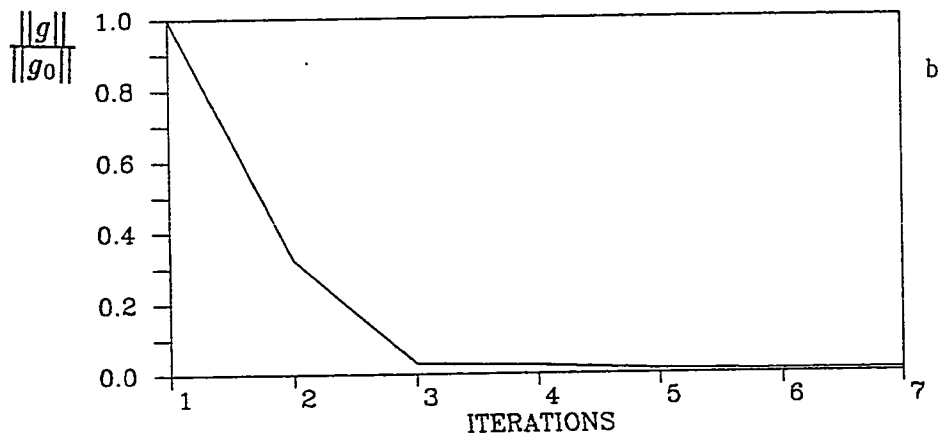
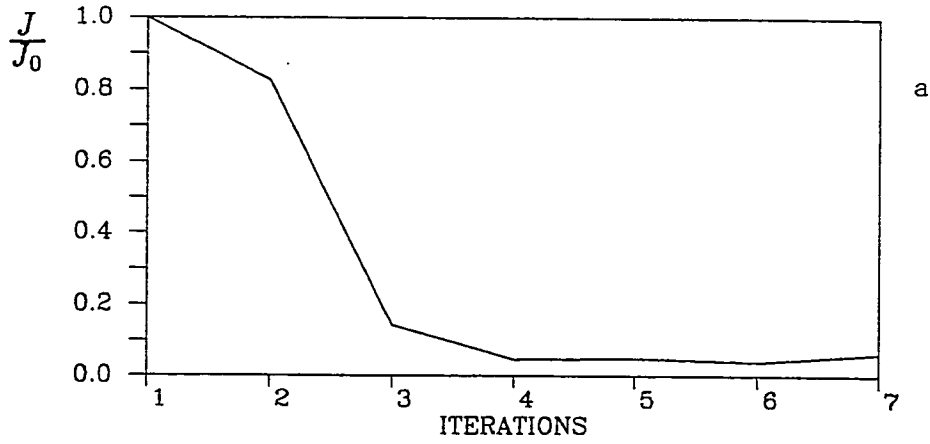


Figure 11. The results for the case when the gradient of the cost function is a function of longitude. a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\|g\|}{\|g_0\|}$  is plotted as a function of the number of iterations.

part of the basin, while the correction is getting smaller and smaller as the western end of the basin is approached. West of about  $160^{\circ}\text{E}$  there is hardly any correction at all. The reason is as mentioned above, the initial conditions of this experiment. In the area west of  $160^{\circ}\text{E}$  the data misfits are basically equal to zero during the period of the data assimilation, which means that there is no direct forcing of the adjoint equations in the area. The dynamics of the model can propagate information given by the data misfits elsewhere, into the region. The Lagrangian multipliers in Figure 8d are not equal to zero west of  $160^{\circ}\text{E}$ . This does not on the other hand make the gradient of the cost function much different from zero, since the expression for the gradient (5 – 13) also contains the terms  $\frac{\partial h}{\partial \phi}$  and  $\frac{\partial h}{\partial \theta}$  which both are small in the area. It is therefore to be expected that the correction in the western part of the basin is going to be smaller than in the eastern part.

Figure 12b shows  $c^2$  after the second iteration. The correction in the eastern part of the basin has now given a maximum value of  $c^2$  slightly higher than the “correct” value. In the western part the correction is still very small. After the fifth iteration, Figure 12c, a flat area has developed in the eastern part with  $c^2$  very close to the value used to create the observations ( $c^2 = 6.0\text{m}^2\text{s}^{-2}$ ). Because of the reason explained above, the correction in the western end is much smaller than in the eastern part. Figure 12d shows the results after seven iterations. The area of “correct” values for  $c^2$  has now spread to nearly half the basin, but there is still a discrepancy between the real and estimated  $c^2$  in the eastern part of the basin. After seven iterations the gradient of the cost function has decreased about 2 orders of magnitude and the algorithm converged.

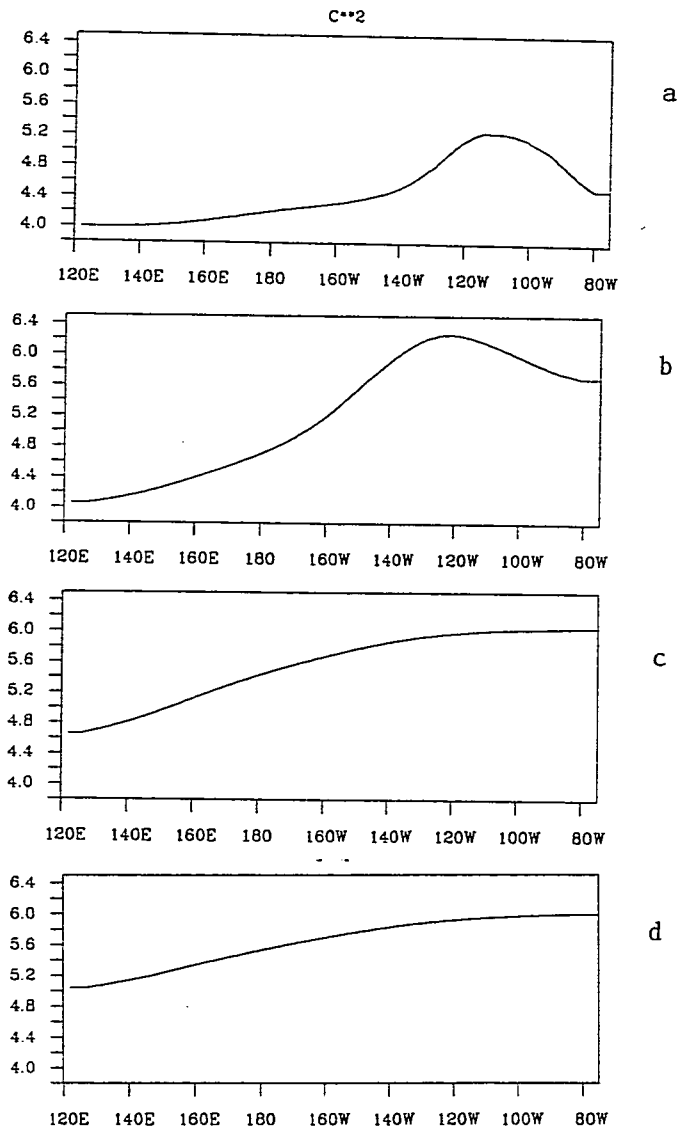


Figure 12. The results for the case when the gradient of the cost function is a function of longitude. The longitudinal variation of the phase speed during the iterative process, a) after one iteration, b) after two iterations, c) after five iterations and d) after seven iterations.

### 6.1.2. Verification of the gradient of the cost function

As mentioned earlier, it is important to get the right gradient of the cost function. One way to check if the correct gradient is found, is described below. Let

$$J(h, c_0^2 + \alpha h) = J(h, c_0^2) + \alpha h \nabla_{c_0^2} J(h, c_0^2) + \text{h.o.t.} \quad (6-1)$$

be a Taylor expansion of the cost function in (5-5).  $\alpha$  is a small scalar and  $h$  is a random vector of unit length. Rewriting (6-1) one can define a function of  $\alpha$  by

$$f(\alpha) = \frac{J(h, c_0^2 + \alpha h) - J(h, c_0^2)}{\alpha h \nabla_{c_0^2} J(h, c_0^2)} \approx 1 + \text{h.o.t.} \quad (6-2)$$

The result of the calculation in (6-2) is only valid for values of  $\alpha$  sufficiently large so that the numerator in (6-2) can be calculated accurately. If  $\alpha$  is chosen close to the machine zero one cannot expect to be able to verify that the correct gradient has been found. For values of  $\alpha$  which are not too close to the machine zero one should expect to obtain a value for  $f(\alpha)$  which is close to 1. For the CYBER 205 the machine zero is about  $10^{-15}$  and if  $\alpha$  is chosen between say  $10^{-3} - 10^{-10}$  one would expect to find  $f(\alpha) \sim 1$ .

To check if the calculation of the cost function gives consistent answers one can calculate  $J(h, c_0^2 - \alpha h)$  and check if

$$J(h, c_0^2 + \alpha h) - J(h, c_0^2) \approx \frac{1}{2} \left( J(h, c_0^2 + \alpha h) - J(h, c_0^2 - \alpha h) \right) \quad (6-3)$$

The experiment performed in section 6.1.1 was used to check the gradient. The gradient of the cost function with respect to the phase speed was calculated using (5-13). Different values of  $\alpha$  were chosen and equation (6-3) was verified. In figure 13 a plot of  $f(\alpha)$  is shown. It is clearly seen that for  $\alpha$  between  $10^{-3}$  and  $10^{-10}$ , (6-2) is verified. The correct gradient is therefore found.

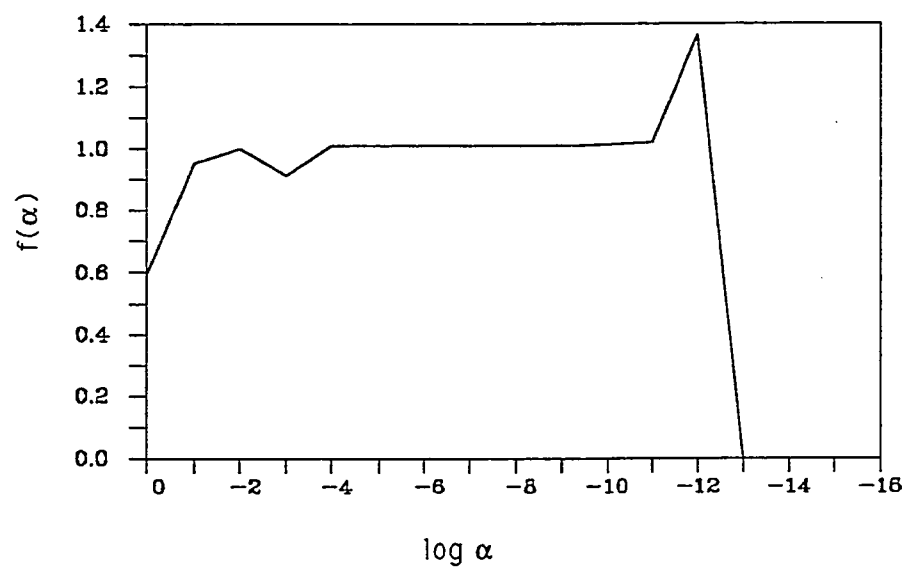


Figure 13. Plot of the function  $f(\alpha)$  in (6 - 2).

## 6.2. Assimilation of identical twin data in the Pacific Ocean model forced by real winds

As mentioned above two different scenarios are going to be investigated, but first a few test cases are studied. In these initial experiments the results from the year 1979 are used, with identical twin data as observations.

### 6.2.1. Experiment 2: Constant phase speed and observations everywhere

In this experiment the results from the original model integration are used as observations. A new integration of the model for the year 1979 was performed, changing the constant  $c^2$  to  $4.0\text{m}^2\text{s}^{-2}$ . A data assimilation and parameter estimation procedure as described in experiment 1 was performed. Perfect observations were available at every gridpoint of the model. In the experiments with the model forced by real winds, the model is restarted using the end of the previous year as initial conditions. The phase speed determined by the algorithm is a time averaged phase speed over the period of the assimilation. After each iteration the model was integrated for two years, the year prior to the year of the assimilation plus the year of the actual assimilation. This was done in order for the model to adjust to the new phase speed even in the beginning of the assimilation period.

In figure 14a and b, the normalized cost function and its gradient is shown, respectively. As before there is a rapid decrease in both the cost function and the gradient during the first few iterations. After seven iterations the values have dropped about two orders of magnitude and the algorithm converged. Figure 15a-c shows the corresponding evolution of the spatial structure of  $c^2$ . In figure 15a the results after the first iteration is shown. There is a correction all over the basin, with a slightly larger correction in the central region. After the third iteration, figure 15b, the “correct” value of  $6.0\text{m}^2\text{s}^{-2}$  has been found in the western part of

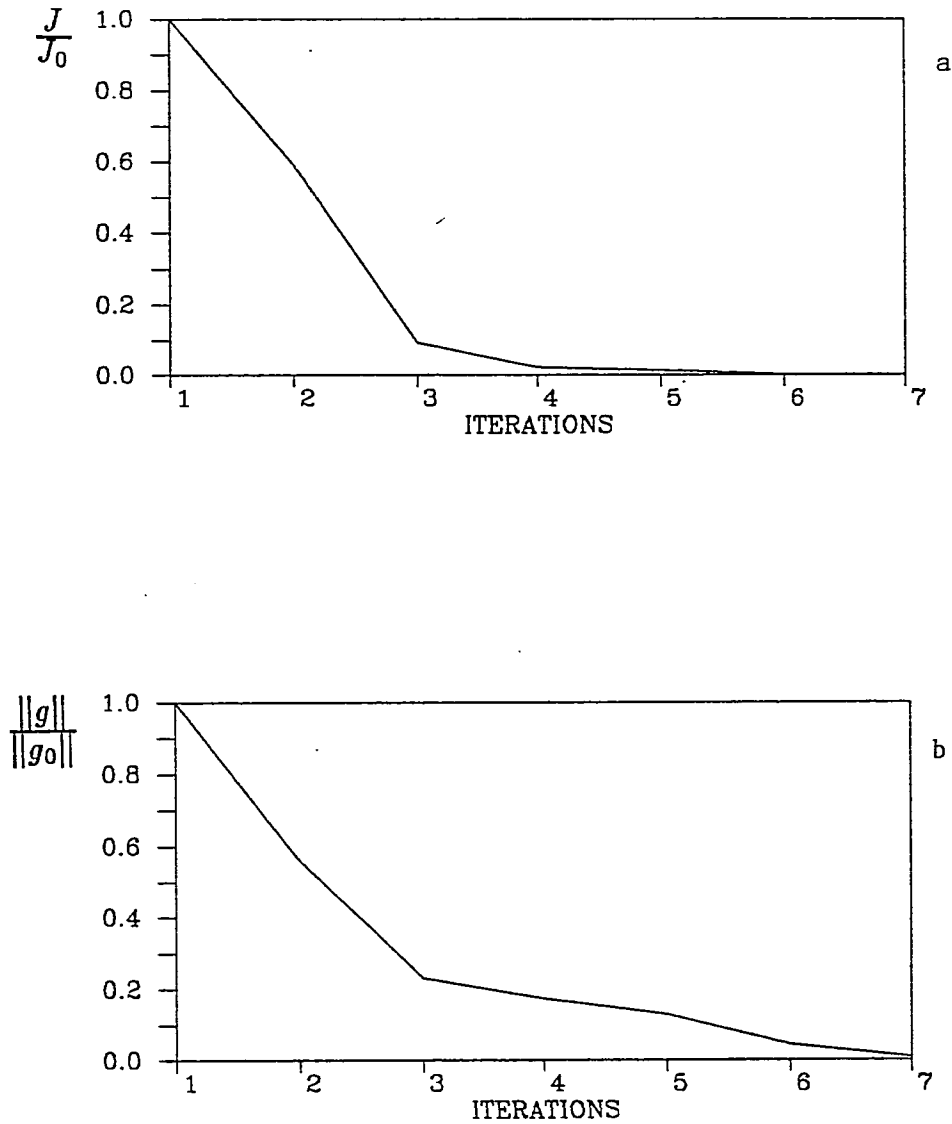


Figure 14. The results from experiment 2 are shown. a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\|g\|}{\|g_0\|}$  is plotted as a function of the number of iterations.



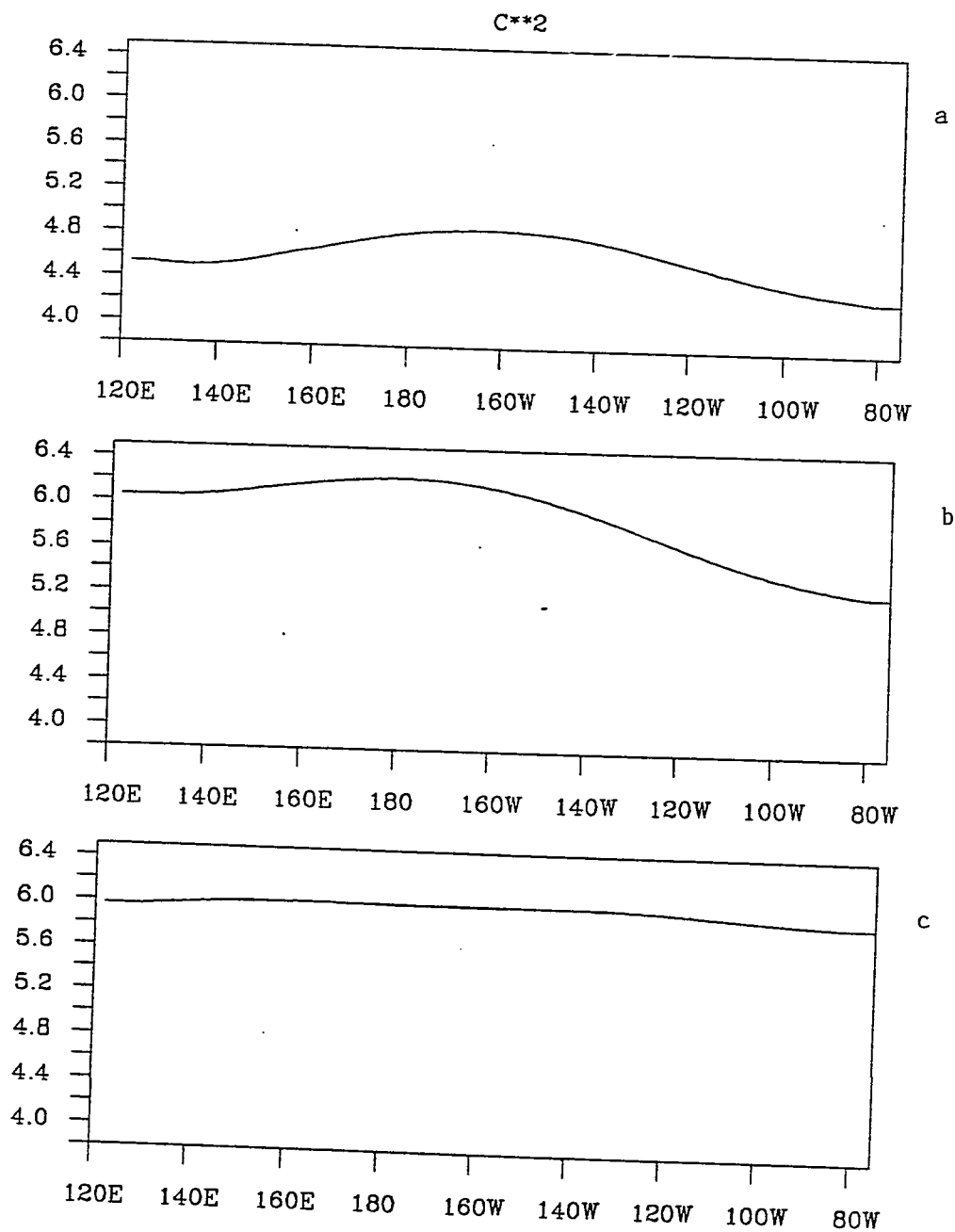


Figure 15. The longitudinal variation of the phase speed during the iterative process in experiment 2, a) after one iteration, b) after three iterations, c) after seven iterations.

the basin, while there is still a discrepancy between the estimated and the correct value in the eastern part. Figure 15c which is a plot of  $c^2$  after the seventh iteration, shows that the correct value of  $c^2$  has been determined.  $c^2$  is now very close to a constant over the whole domain, and there are only small departures from the value of  $c^2 = 6.0\text{m}^2\text{s}^{-2}$  used to create the “observations”.

### 6.2.2. Experiment 3: Variable phase speed and observations everywhere

The experiments discussed so far have tried to estimate a constant  $c^2$ . Even if  $c^2$  was allowed to vary with longitude during the iterative process, the end result should give a constant phase speed. In a real ocean this is not very realistic, and an important question arises if the method is able to determine a spatially varying phase speed. In this experiment the algorithm is therefore tested on a more realistic case, where the phase speed is a function of longitude. Figure 16 is a plot of the phase speed used in the model integration creating the “observations”. The variation of  $c^2$  is supposed to model what is a realistic variation, since one would expect to have higher values in the west compared to the east. This is at least correct under “normal” conditions, i. e. when an El Nino event is not dominating the Pacific Ocean. The upper layer thickness is shallow in the east and deep in the west, and consequently  $c^2$  would have a longitudinal distribution as in figure 16. The model was restarted in 1977 and integrated through 1979 using the new phase speed in order to let the model adjust.

As initial guess,  $c^2$  was assumed to be constant with a value of  $4.0\text{m}^2\text{s}^{-2}$ . Again perfect observations were available at every gridpoint of the model. The results of the assimilation are shown in figure 17a-b and in figure 18a-d. The normalized cost function is plotted as a function of the number of iterations in figure 17a, while the normalized gradient is plotted in figure 17b. The rapid decrease

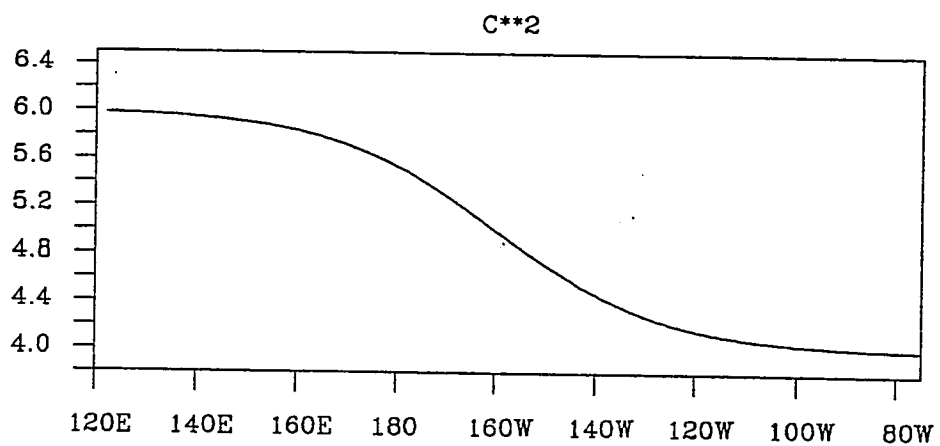


Figure 16. The longitudinal variation of the phase speed used to create the observations.

during the first few iterations can be observed. After ten iterations the decrease in the gradient is more than two orders of magnitude and the algorithm converged. figure 18a-d shows the evolution of the spatial variation of  $c^2$ . After the first iteration there is a strong adjustment in the western part of the basin and a smaller correction further to the east. Comparing the results after the first iteration to the optimal  $c^2$  in figure 16, it is difficult to see if the estimation is going to work. Figure 18b which is a plot of  $c^2$  after three iterations, shows a promising development of the spatial structure. A slope has become evident with higher values in the western part compared to the values further east, but there is still a rather large discrepancy if the results from the third iteration is compared to the real  $c^2$ . After six iterations, Figure 19c, the spatial variation of  $c^2$  is beginning to look very much like the spatial variation of  $c^2$  in figure 16. The maximum value in the west is slightly lower than the real values, while in the eastern part  $c^2$  has a slightly higher value than it is supposed to. Finally, in figure 18d the results after ten iterations is shown. The spatial variation is now very close to the variation of the  $c^2$  used to create the "observations", see figure 16. As a conclusion one can say that the algorithm has been successful in determining the spatially varying phase speed.

#### *6.2.3. Experiment 4: Variable phase speed and observations at 3 stations*

In all the experiments described above, observations have been assumed to be available everywhere. This is not going to be the case in a real data assimilation. Even with the new satellites, the coverage will not be good enough to give observation at every gridpoint of the model. The time resolution will also be quite different from what has been used in the above experiments. The TOPEX/POSEIDON project is still trying to decide what the repeat period of the satellite should be. Periods of 10 or 20 days have been suggested, which mean that for a certain position a new observation will be available every 10(20) days. These repeat periods will

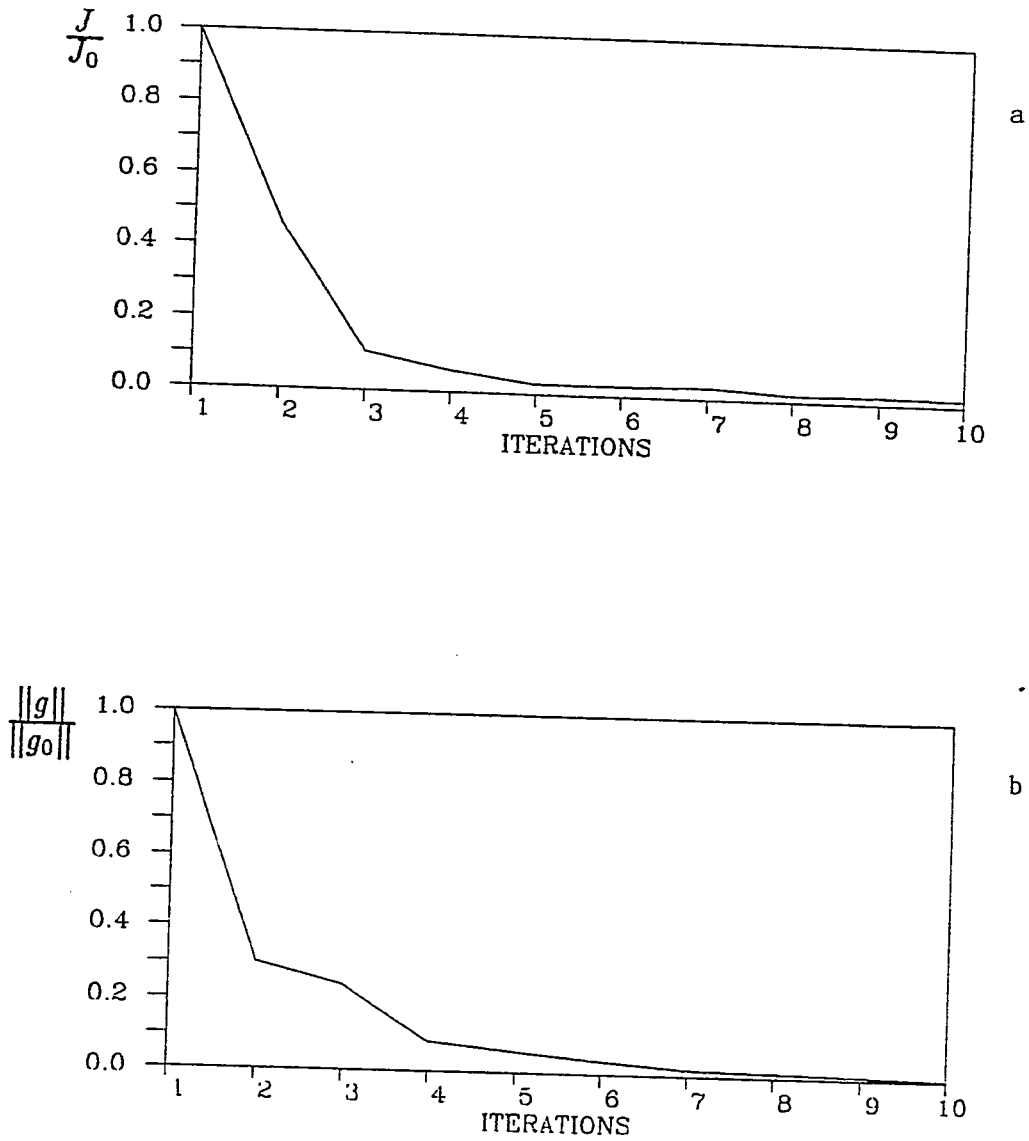


Figure 17. The results from experiment 3 are shown. a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\|g\|}{\|g_0\|}$  is plotted as a function of the number of iterations.

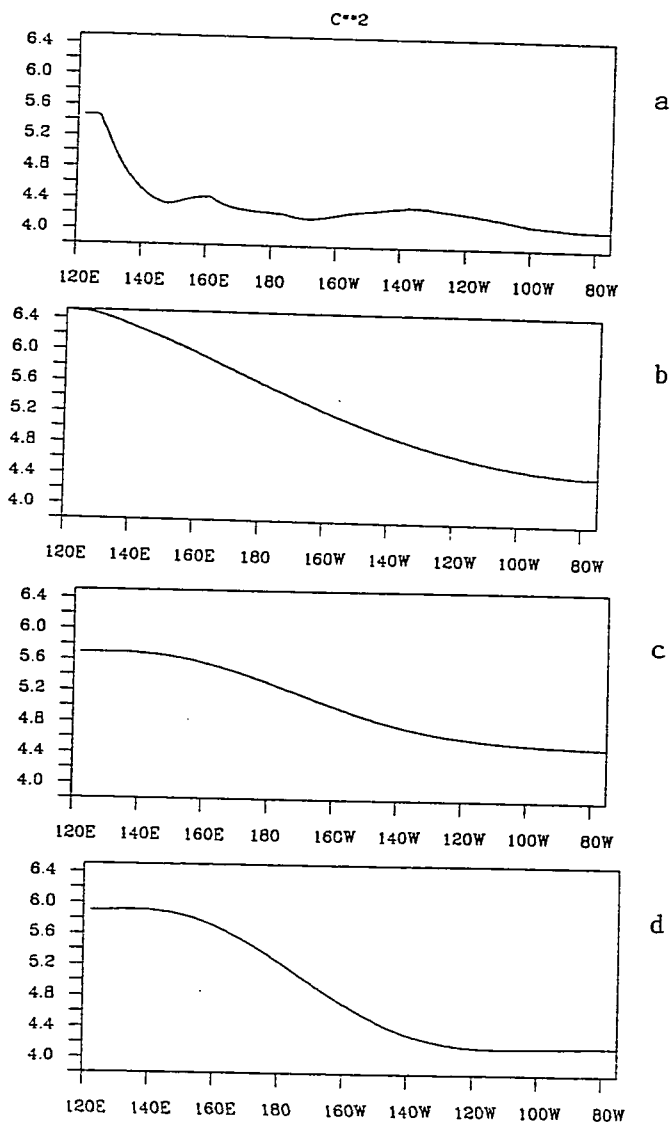


Figure 18. The longitudinal variation of the phase speed during the iterative process in experiment 3, a) after one iteration, b) after three iterations, c) after six iterations, d) after ten iterations.

give a spatial resolution of about  $320 \times 320 \text{ km}^2$  (10 days) and  $160 \times 160 \text{ km}^2$  (20 days) in midlatitudes.

The goal of this research is to assimilate sea level observations from island stations. The number of stations are limited and the number of observations are therefore drastically reduced compared to the cases discussed above. In this experiment it is assumed that data are only given at 3 stations (the same stations will be used when real sea level observations are assimilated). The stations are, Santa Cruz, Jarvis, and Truk, see Figure 2. These three stations are distributed over the Pacific Ocean, with Santa Cruz representing the eastern part of the ocean, Jarvis the central region and Truk the western region. The experiment is identical to experiment 3, with the phase speed in figure 16 creating the observations and the initial guess for  $c^2$  is  $4.0 \text{ m}^2 \text{ s}^{-2}$ . The observations are again assumed to be perfect.

Figure 19a-b and figure 20a-d show the results of the iterative process. The normalized cost function is shown in figure 19a, while the normalized gradient is shown in figure 19b. The algorithm converged after ten iterations. Figure 20a-d show the evolution of the spatial variation of  $c^2$ . After the first iteration, figure 20a, there is a small adjustment over most of the basin, with slightly higher values in the west. Figure 20b which is the result after the third iteration starts to show the right spatial structure. A slope from west to east has developed. The variation of  $c^2$  after six iterations is shown in figure 20c. Now a definite resemblance to the "real"  $c^2$  has developed. The values in the western and eastern part are too high, but overall there is a quite good agreement. After ten iterations the algorithm converged and the results are shown in figure 20d. Using observations from only three stations, it is still possible to estimate the correct spatial structure of  $c^2$ .

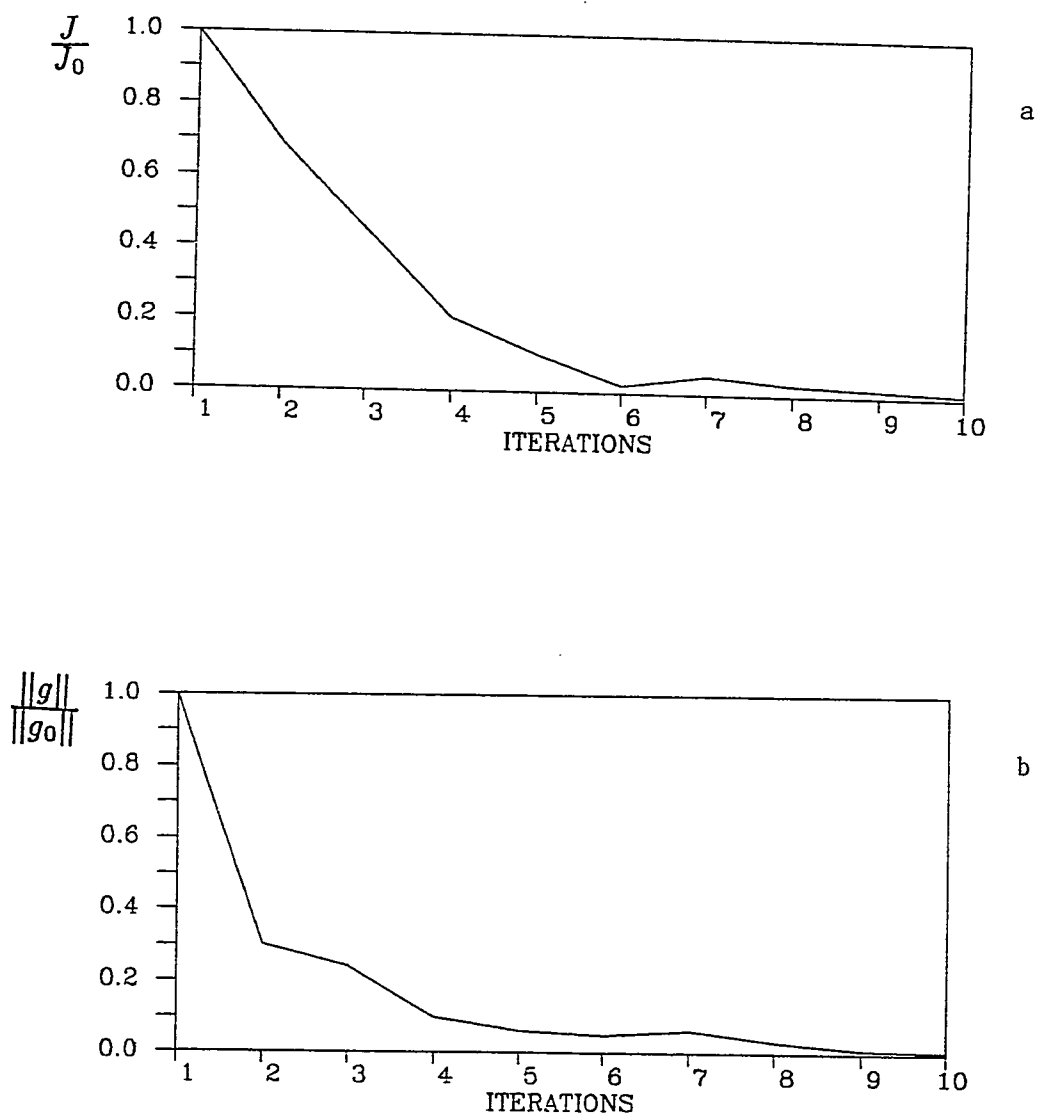


Figure 19. The results from experiment 4 are shown. a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\|g\|}{\|g_0\|}$  is plotted as a function of the number of iterations.



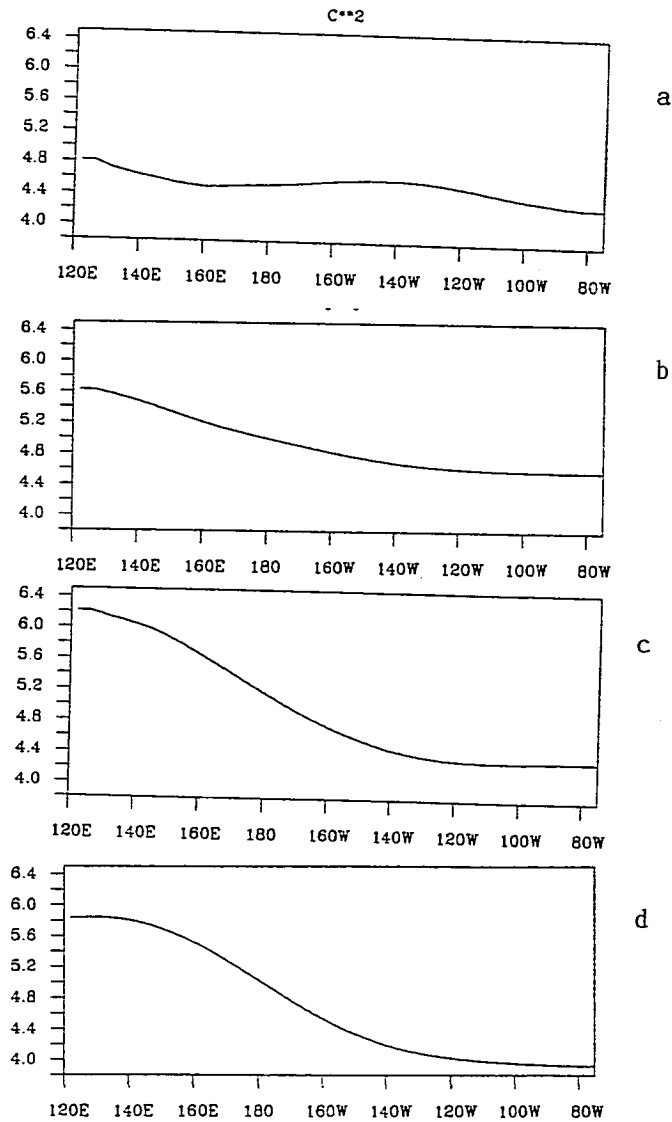


Figure 20. The longitudinal variation of the phase speed during the iterative process in experiment 4, a) after one iteration, b) after three iterations, c) after six iterations, d) after ten iterations.

#### 6.2.4. Uniqueness and stability of the solution

An important question is whether the solution is unique or not. If the solution is unique, the cost function  $J$  should be convex, i. e. only one global minimum of the function should exist. An experiment with observations in only one region of the Pacific Ocean, showed that the estimated structure of the phase speed could look completely different from the case when observations were available in more than one region. However, this does not mean that the parameter is nonunique. If observations are available at only one station, the algorithm is going to adjust the parameter so that the cost function has a minimum at this station. Adding observations in other regions will change the spatial structure of the estimated  $c^2$ , in order for the total cost function to decrease. For a given number of stations, the estimated structure of  $c^2$  is unique. Given observations at only one station it is difficult and in most cases impossible to determine a spatially varying  $c^2$ . To be able to determine a spatially varying phase speed it turned out that it was necessary to have observations in the different dynamical regions of the model. Observations should be available in the western region where the model is dominated by the westward propagating Rossby waves, in the central region where there is a mixture between Kelvin and Rossby waves influencing the model and in the eastern region where for the most part Kelvin waves dominate the solution.

The final solution is not sensitive to the initial guess of  $c^2$ . Several experiments starting the assimilation with different guesses for the initial phase speed, showed that the algorithm always converged to the correct solution. The number of iterations needed for the algorithm to converge, did not depend strongly on the initial guess either. Changing the initial guess from  $4.0\text{m}^2\text{s}^{-2}$  to  $3.0\text{m}^2\text{s}^{-2}$  in experiment 3, the algorithm converged in 11 iterations instead of 10.

Another question is whether the estimated parameter is stable, i. e. will small errors in the variables lead to large changes in the computed parameter. To investigate this the same scenario as in the experiment 4 was used, but the “observations” were contaminated by what would represent errors in the observations. The errors consisted of normally distributed random numbers added to the time series of the observations. The maximum error would represent an accuracy of measuring the sea surface elevation of  $\pm 0.05$  meter. Even with this large error in the observations, the algorithm was able to find the correct spatial structure of the phase speed. The algorithm converged in ten iterations and the final result of the assimilation was nearly identical to the result obtained for the case with perfect observations shown in figure 20d.

### *6.3. Assimilation of real sea level observations in the Pacific Ocean model*

#### *6.3.1. Experiment 5: Real sea level observations from 3 station in 1979*

Real sea level observations from three stations are used in this experiment. The stations are the same as in the experiment described above, Santa Cruz, Jarvis and Truk. The initial guess for  $c^2$  was a constant value of  $6.0\text{m}^2\text{s}^{-2}$ . The results of the assimilation procedure are shown in figure 21a-b and figure 22a-c. Figure 21a is a plot of the normalized cost function. The cost function for the three stations dropped to about 35% of its initial value after six iterations. Looking at the cost function at the three different stations separately, showed that for the station at Santa Cruz the decrease was much larger. After six iterations the cost function had dropped to about 10% of its initial value. For the station in the central region, Jarvis, there was basically no change at all, while for Truk in the western region, there was a decrease of about 60%. The cost function in the central region had a low value during the whole iterative process. The initial value of the cost function at Jarvis was less than 20% of the value at the two other stations, and at the end

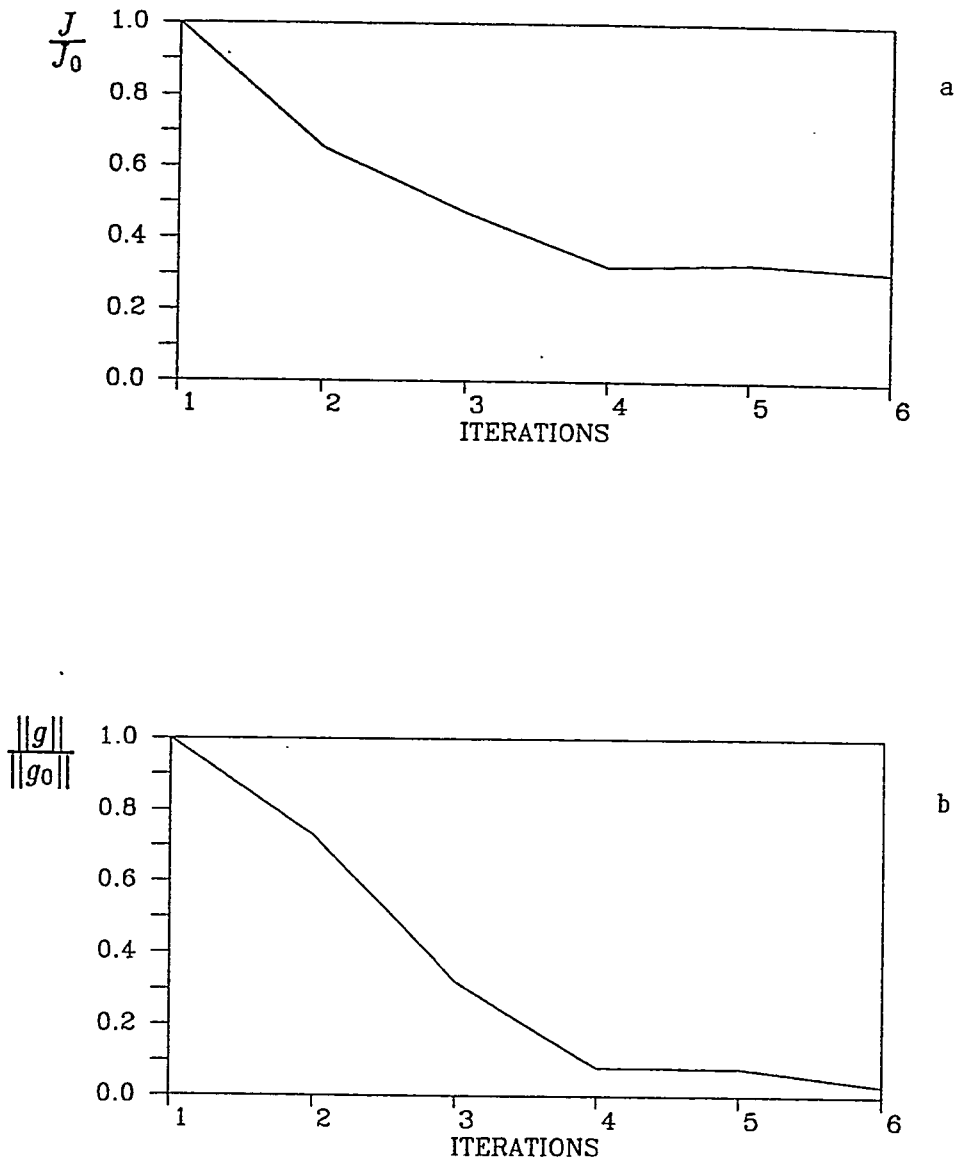


Figure 21. The results from experiment 5 are shown. a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\|g\|}{\|g_0\|}$  is plotted as a function of the number of iterations.

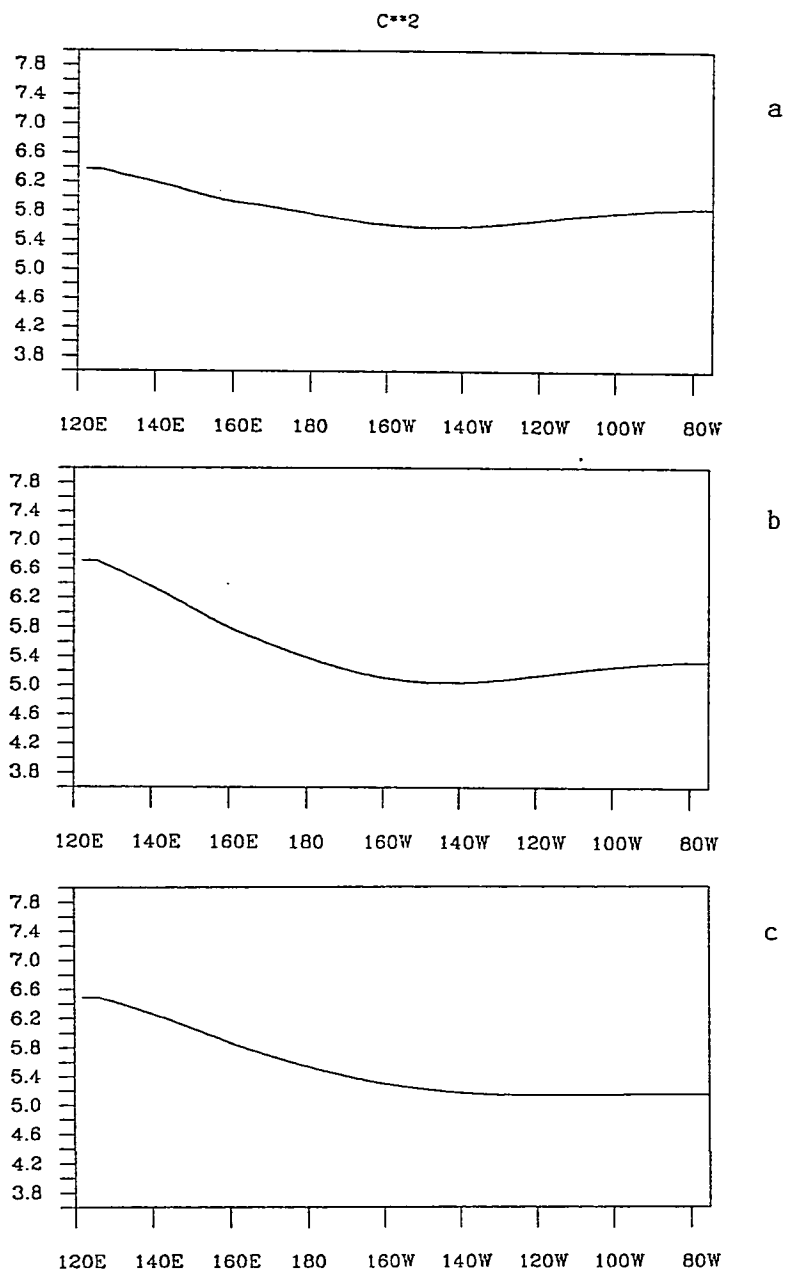


Figure 22. The longitudinal variation of the phase speed during the iterative process in experiment 5, a) after one iteration, b) after three iterations, c) after six iterations.

of the assimilation the cost function at Santa Cruz and Jarvis had about the same value. Since the data misfits (the cost function is a measure of the misfits) are the forcing for the adjoint equations, a station where the misfits are small will not have as much influence on the final solution. This could be changed by letting the validity coefficients be functions of space and not as here constants. On the other hand, there is no reason to assume that the observations at Jarvis should have a larger influence on the solution than the other stations. Figure 21b shows the corresponding normalized gradient of the cost function. The gradient dropped to about 5% of its initial value. Increasing the number of iterations did not result in a further decrease in the gradient or the cost function. Even if the decrease is not two orders of magnitude as in the previous identical twin experiments, the norm of the gradient after six iterations has about the same value as in the identical twin examples. The conclusion is therefore that the minimum of the cost function has been found.

Figure 22a is a plot of the spatial structure of  $c^2$  after the first iteration. There is an adjustment in the central and western region, with higher values in the west and lower values in the middle of the basin. After three iterations, figure 22b, there is a strong adjustment. A slope has developed with high values in the west and low values in the east. Finally, the result after six iterations is shown in figure 22c. There is a slope from west to east with most of the slope confined to the western end of the Pacific Ocean. From about  $160^\circ\text{W}$ ,  $c^2$  is more or less constant with a value of  $5.2\text{m}^2\text{s}^{-2}$ . The maximum value in the west is about  $6.5\text{m}^2\text{s}^{-2}$ . The spatial structure of  $c^2$  with high values in the west and lower values in the east, is in good agreement with observations from the Pacific. During normal conditions the basic stratification consists of a deep thermocline in the west and a shallow

thermocline in the east, which corresponds exactly to the results obtained, ( $c^2$  is proportional to the depth of the thermocline (ULT)).

A question now arises, did the new  $c^2$  improve the model results? In terms of the cost function the answer is yes. To investigate this issue further, the correlation coefficient between the model results and the observations is calculated for the different stations. The coefficient is calculated using the formula

$$r = \frac{\sum_{i=1}^{360} hh'}{\left( \sum_{i=1}^{360} h^2 \sum_{i=1}^{360} h'^2 \right)^{\frac{1}{2}}} \quad (6-4)$$

where  $h$  and  $h'$  represents the time series of the model results and the observations, respectively. The mean value is removed from both  $h$  and  $h'$ . figure 23 is a plot of the correlation coefficient as a function of iterations for the stations at Santa Cruz, Jarvis, Truk, Nauru and Guam. The stations at Santa Cruz, Jarvis and Truk were used in the data assimilation, while Nauru and Guam are stations which were not included in the assimilation. The correlation coefficient increases for all the stations, with a dramatic increase for the station at Santa Cruz. The model did not do a good job in predicting the ULT at Guam prior to the assimilation, but at the end of the assimilation there is a significant increase in the correlation coefficient. It is interesting to note that the new estimate of  $c^2$  was able to improve the model results at this location, even if Guam was not one of the stations used in the assimilation. An increase in the correlation coefficient at Nauru can also be observed. The change is not as dramatic at this station, but there is still a significant increase. As discussed above the cost function at Jarvis had a low value during the whole assimilation period. Figure 23 shows that the correlation coefficient for this station has the highest value of the three stations used in the assimilation. At the end of the iterations the correlation coefficient at Santa Cruz and Jarvis has about

the same value, which agrees well with the fact that the values of the cost functions at the two stations were almost equal.

### 6.3.2. Experiment 6: Real sea level observations from 3 stations in 1982/83

As in experiment 5, real sea level observations from the 3 stations at Santa Cruz, Jarvis and Truk were used. In this case the assimilation started in June 1982 and continued for one year. The El Nino which occurred in 82/83 is therefore a part of this assimilation. Again the initial guess of  $c^2$  was  $6.0\text{m}^2\text{s}^{-2}$ . Figure 24a is a plot of the normalized cost function. As in the previous experiment the cost function decreased to almost 35% of its initial value after five iterations. The station at Santa Cruz experienced the largest decrease in the cost function. After five iterations the value had dropped to about 10% of the initial value. As in the previous experiment the cost function at Jarvis was more or less constant during the iterative process, with a value of about the same magnitude as the final result at Santa Cruz. The reduction of the cost function at Truk was about 50%. Figure 24b shows the normalized gradient of the cost function, and as in the previous experiment the gradient dropped to about 5% of its initial value. Further iterations did not improve the results in this case either. The norm of the gradient after five iterations had about the same value as when the identical twin experiments converged, and the minimum of the cost function has therefore been found.

The corresponding evolution of the spatial structure of  $c^2$  is shown in figure 25a-c. After the first iteration, figure 25a, there is an adjustment in the western and central region of the basin. Close to the western boundary the values of  $c^2$  have dropped, while higher values can be found in the central area. Figure 25b shows the results after the third iteration. Now there is a large decrease in the western area, while higher values than the initial can be found east of about



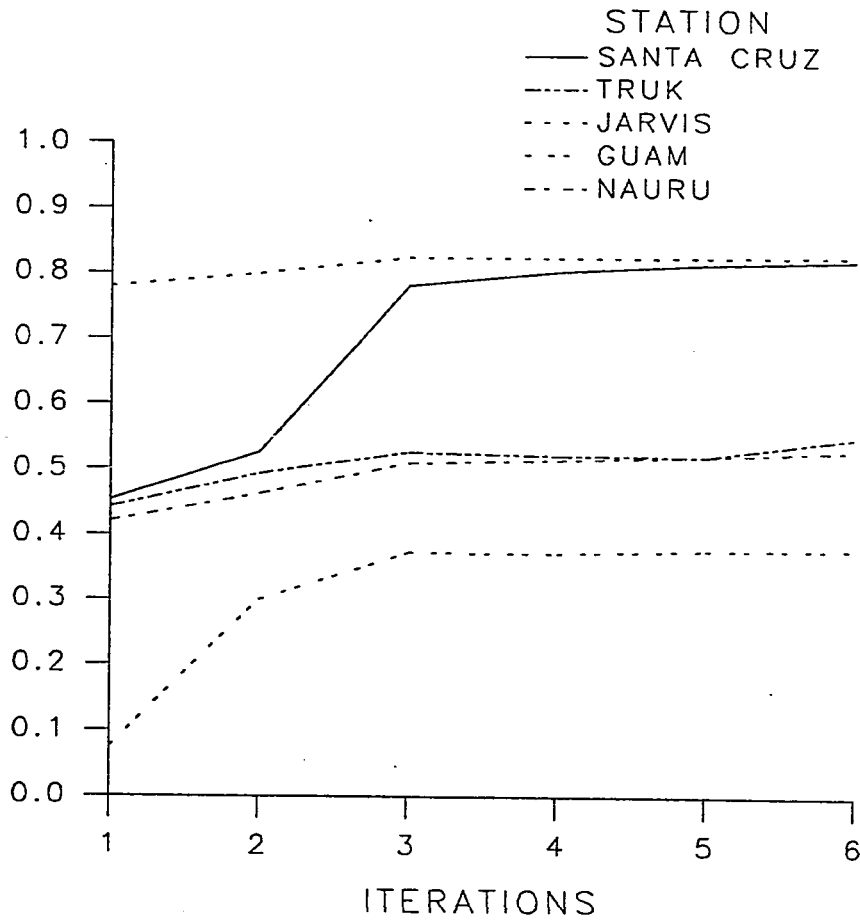


Figure 23. The correlation coefficient for different stations as a function of the number of iterations in experiment 5.

160°E. A maximum occurs around 160°W. After five iterations, figure 25c, a steep slope has developed with the lowest values close to the western boundary where  $c^2 = 3.3\text{m}^2\text{s}^{-2}$ . The maximum value of  $c^2 = 7.0\text{m}^2\text{s}^{-2}$  occurs around 160°W with slowly decreasing values as the eastern boundary is approached. As in the previous experiment, the spatial structure of the estimated phase speed is in good agreement with observations. An El Nino event is recognized by the fact that the eastern Pacific Ocean becomes warmer, which correspond to an increase in the upper layer thickness. The western Pacific is cooler than normal corresponding to a decrease in the upper layer thickness.

The correlation coefficients for the stations are shown in figure 26. The general picture is that the model gives better correlation for all the stations during an El Nino year. During the iterative process the correlation coefficient increases for all the stations. The stations at Guam and Nauru are not shown, since continuous observations were not available for the assimilation period. The observations at Ponape were not used in the assimilation, and as in the previous experiment, the new estimate of  $c^2$  is able to improve the model results even at locations which are not part of the assimilation.

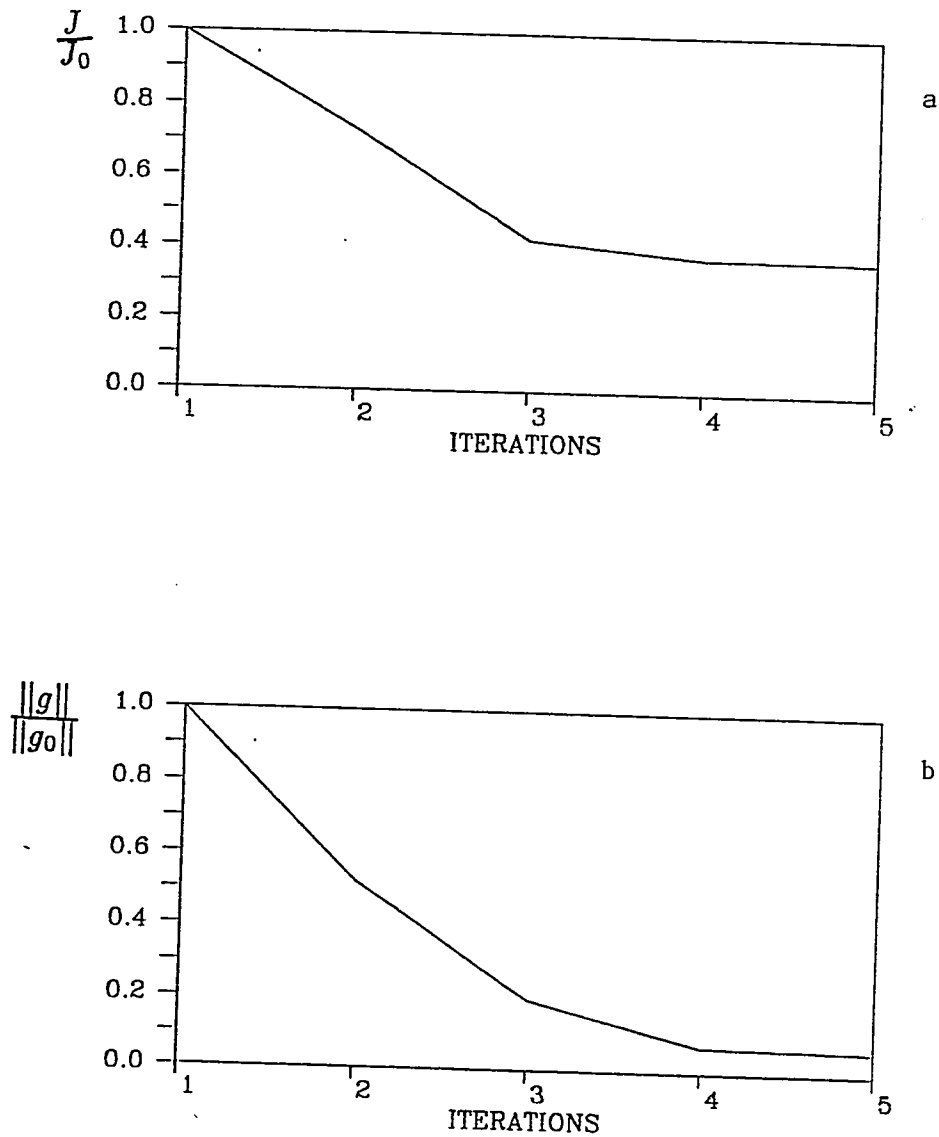


Figure 24. The results from experiment 6 are shown. a) The normalized cost function  $\frac{J}{J_0}$  is plotted as a function of the number of iterations. b) The normalized gradient of the cost function  $\frac{\|g\|}{\|g_0\|}$  is plotted as a function of the number of iterations.

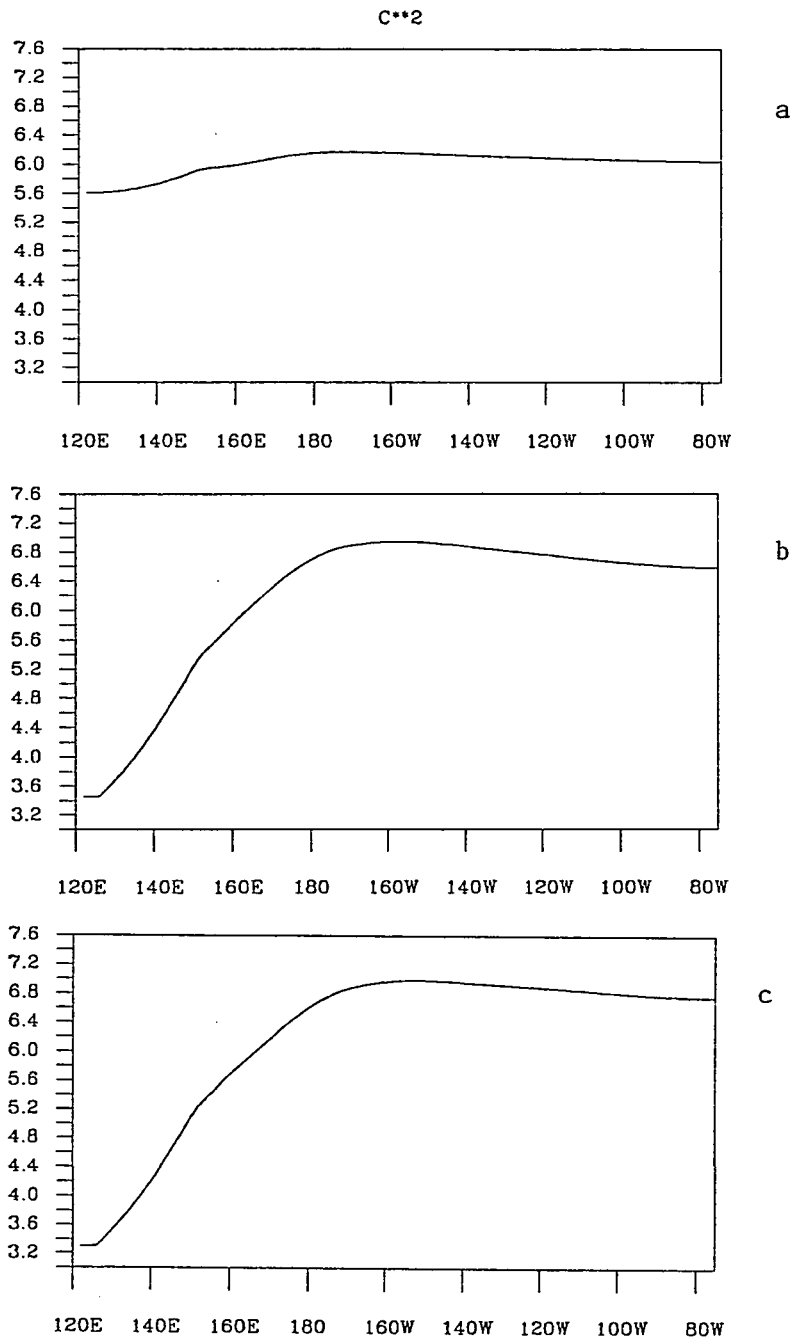


Figure 25. The longitudinal variation of the phase speed during the iterative process in experiment 6, a) after one iteration, b) after three iterations, c) after five iterations.

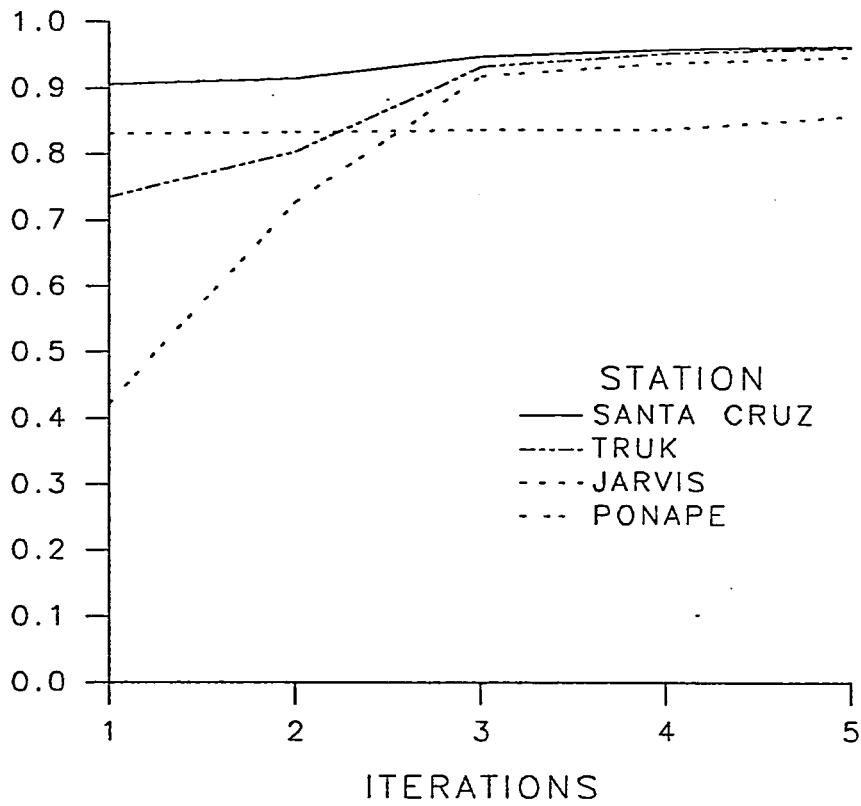


Figure 26. The correlation coefficient for different stations as a function of the number of iterations in experiment 6.

## 7. Summary and conclusions

A variational data assimilation method for a reduced gravity model has been developed. The method was applied to the equatorial Pacific Ocean. The phase speed of the model was chosen to be the control parameter of the problem, and the optimal spatial structure of the parameter was determined. Knowledge of the spatial structure of the phase speed will give information about the basic stratification of the Pacific Ocean. In the variational formalism a cost function measuring the “distance” between the observations and the model results is minimized. The method consists of integrating the model equations forward in time over the period which data are going to be assimilated. Data misfits between the model and the observations are then calculated and the adjoint equations of the model are integrated backward using the data misfits as forcing. It is necessary to determine the gradient of the cost function with respect to the control parameter. The gradient can be found using the model and the adjoint variables and it is used in a minimization algorithm to determine a new value of the phase speed. The minimization procedure utilized a conjugate gradient method to determine the search direction. The quasi-Newton limited-memory conjugate gradient subroutine of Shanno and Phua (1980) was implemented with a few modifications. A step length was calculated using a similar method as in Derber (1985). The convergence criterion for the algorithm required the value of the cost function and/or its gradient to have decreased by a specified tolerance. If this was not the case the iterative procedure must be continued. The new value of the parameter can be used in the procedure

described above, and it can be repeated until a satisfactory decrease in the cost function and its gradient has been achieved.

Several different experiments have been performed to test the assimilation algorithm. The first example used a simple set up of the Pacific Ocean model. The model was initialized by the height and velocity field of a Kelvin wave and the model was integrated forward in time using a value of  $c^2 = 6.0\text{m}^2\text{s}^{-2}$  to create what was called the observations. The assimilation algorithm was started using an initial guess for  $c^2 = 4.0\text{m}^2\text{s}^{-2}$  and the optimal value of  $c^2$  was estimated. This simple example clearly demonstrated how information propagated backward in time through the adjoint variables. Perfect observations were assumed to be available everywhere, and it was shown that it was possible to determine the original phase speed. The algorithm converged in seven iterations. Due to the initial conditions of this experiment, the gradient of the cost function was small in certain areas of the domain and the final result was not as close to the “real” solution as was the case in the next experiments.

The Pacific Ocean model was forced with real winds in the experiments which followed. The model was integrated from 1972 through 1983, and in the initial experiments, identical twin observations from the year 1979 were used. A similar experiment as the one described for the Kelvin wave case, was performed. The algorithm converged in seven iterations and the final result of the assimilation gave a  $c^2$  which was constant with longitude. Since a constant  $c^2$  is not very realistic as the optimal solution when real data are assimilated, another experiment was performed. In this case the  $c^2$  used to create the “observations” was a function of longitude, with high values in the western part of the basin and then decreasing values as the eastern part of the Pacific Ocean was approached. Again perfect observations were used, and the assimilation was started with an initial guess of a

constant  $c^2 = 4.0\text{m}^2\text{s}^{-2}$ . The algorithm converged after ten iterations and the final result of the estimation was a spatially varying  $c^2$  which had the correct structure with high values in the west, and low values in the east.

In a real data assimilation, observations will not be available at every grid-point of the numerical model. Since the goal of this research was to assimilate sea level observations from island stations, the next experiment used observations at only three stations as forcing for the adjoint equations. The stations which was chosen were Santa Cruz, Jarvis and Truk. These three stations represent the three different regions of the Pacific Ocean. Santa Cruz representing the eastern part of the ocean, Jarvis the central region and Truk the western region. The experiment was identical to the previous experiment except for the number of available observations. Even in this case the algorithm converged after ten iterations and the spatial variation of the estimated  $c^2$  was again very close to the correct structure.

Real sea level observations were assimilated during two different periods. In the first experiment, observations from 1979 were used. This is a normal year, i. e. there was no El Nino event during 1979. In the second experiment observations from 1982/83 were assimilated. A strong El Nino occurred during this period. In both experiments sea level observations from the three stations used in the previous experiment were assimilated. As an initial guess for the phase speed a constant value of  $6.0\text{m}^2\text{s}^{-2}$  was chosen. The results from the assimilation of observations from 1979 gave a spatial structure of the phase speed with higher values in the west and then a gradual slope towards the east where the values had decreased compared to the initial guess. This is in good agreement with observations. The general picture of the upper layer thickness in the Pacific Ocean is a thick ULT in the west and then a slope towards a thinner ULT in the east. Assimilating the observations during 1982/83 gave a completely reversed picture. Now the phase speed had lower values



in the west and higher values in the east, which corresponds to a decrease in the ULT in the west and an increase in the east. This confirms what is happening during an El Nino. A warming of the eastern Pacific Ocean occurs, which corresponds to an increase in the ULT and in the west there is a cooling corresponding to a decrease in the upper layer thickness.

A correlation coefficient between the model results and the observations was also calculated. The correlation increased for all the stations considered during the assimilation period. An important result is that the new estimate of the phase speed was able to improve the correlation not only at the stations used in the assimilation, but also at stations which were not a part of the assimilation.

The results have shown that the estimated phase speed was a unique solution to the problem. It has also been shown that to be able to determine the spatial structure of the phase speed, it was necessary to have observations in the three different dynamical regions of the model. Observations should be available in the eastern region where Kelvin waves dominate the dynamics, in the central region where there is a mixture between Kelvin and Rossby waves influencing the solution and in the western region where Rossby waves play the most important role.

The estimated parameter is stable, i. e. errors in the observations do not result in large variations in the parameter. Normally distributed random numbers were added to the observation in an identical twin experiment. The errors corresponded to an uncertainty in the sea level measurement of  $\pm 0.05$  meter. The final result of the assimilation gave a solution which was very close to the "real" spatial variation.

The variational data assimilation algorithm can be computationally expensive to use. It is necessary to be able to find the optimal solution in as few iterations as possible. The results from the different experiments have shown that

the algorithm is computationally effective. Convergence was achieved in ten iterations or less for all the cases considered. One of the crucial parts of the algorithm is the choice of the minimization method. Several different conjugate gradient algorithms exist. Navon and Legler (1987) tested some of the available subroutines and their conclusion was that the subroutine CONMIN of Shanno and Phua (1980) (which was used here) was the one that gave the best convergence rates. There is a constant development in the area of optimization, and one way of improving the convergence rate could be to implement another subroutine. Navon (personal communication) has tested a new subroutine based on an algorithm of Liu and Nocedal (1988) which gave even better convergence rates than CONMIN. A saving of up to 25% in the number of iterations were found. For problems which require a large number of iteration this is an important improvement. The two assimilations of real sea level observations studied here, required only five or six iterations to converge, and a saving of one or maybe two iterations would not mean much in terms of saved computer time.

Another crucial part of the method is the way the step length is calculated. It is of extreme importance to have a routine which gives the optimal step length. The method used here (Derber (1985)), worked satisfactorily for the experiments considered but may not be adequate for nonlinear models. The initial guess had some influence on the convergence rate. It turned out that the initial guess should not be too small.

The estimated parameter has been assumed to be a function of longitude only. Letting the phase speed be an arbitrary function of both latitude and longitude would not be a feasible problem to solve given the limited number of observations available from sea level stations. Satellite observation may give enough information to make it possible to determine the spatial structure of a phase speed which is a

function of both latitude and longitude. At least it should be possible to determine the large scale structure of the phase speed. Short scale variations will probably have to be removed from the solution. As mentioned earlier experience has shown that it is sometimes necessary to reduce the dimension of the parameter in order to obtain stable results which do not bounce back and forth between high and low values during the iterative process. Restricting the attention to the large scale variation is one way of reducing the dimension of the parameter.

## References

- Adamec, D. and J. J. O'Brien, 1978: The seasonal upwelling in the Gulf of Guinea due to remote forcing. *J. Phys. Oceanogr.*, 8, 1050–1060.
- Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of atmospheric motion. *J. Comput. Phys.*, 1, 119–143.
- Beale, E. M., 1972: A derivation of conjugate-gradients. *Numerical methods for non-linear optimization*. F. A. Lootsma, Ed., Academic Press, 39–43.
- Bengtsson, L., M. Ghil and E. Källén, editors, 1981: *Dynamic meteorology: Data assimilation methods*. Springer Verlag, New York, 330pp.
- Bennett, A. F. and P. C. McIntosh, 1982: Open ocean modelling as an inverse problem: Tidal theory. *J. Phys. Oceanogr.*, 12, 1004–1018.
- Bennett, A. F. and W. P. Budgell, 1987: Ocean data assimilation and the Kalman filter: Spatial regularity. *J. Phys. Oceanogr.*, 17, 1583–1601.
- Bergthorsson, P. and B. R. Döös, 1955: Numerical weather map analysis. *Tellus*, 7, 329–340.
- Bertsekas, D. P., 1982: *Constrained optimization and Lagrange multiplier methods*. Academic Press, London, 395pp.
- Bloom, S. C., 1983: The use of dynamical constraints in the analysis of mesoscale rawinsonde data. *Tellus*, 35A, 363–378.
- Bratseth, A. M., 1986: Statistical interpolation by means of successive corrections. *Tellus*, 38A, 439–447.

- Broyden, C. G., 1970: The convergence of a class of double-rank minimization algorithms. *J. Inst. Math. Applic.*, 6, 76–90.
- Bryan, K., 1969: A numerical method for the study of the circulation of the World Ocean. *J. Comput. Phys.*, 3, 347–376.
- Budgell, W. P., 1986: Nonlinear data assimilation for shallow water equations in branched channels. *J. Geophys. Res.*, 91, 10633–10644.
- Budgell, W. P., 1987: Stochastic filtering of linear shallow water wave processes. *SIAM J. Sci. Stat. Comput.*, 8, 152–170.
- Busalacchi, A. J. and J. J. O'Brien, 1980: The seasonal variability in a model of the Tropical Pacific. *J. Phys. Oceanogr.*, 10, 1929–1951.
- Camerlengo, A. L. and J. J. O'Brien, 1980: Open boundary conditions in rotating fluids. *J. Comp. Physics*, 35, 12–35.
- Cane, M. A., 1979: The response of an equatorial ocean to simple wind stress patterns: I Model formulation and analytic results. *J. Marine Res.*, 37, 233–252.
- Carrera, J. and S. P. Neuman, 1986a: Estimation of aquifer parameters under transient and steady state conditions, 1: Maximum likelihood method incorporating prior information. *Water Resour. Res.*, 22(2), 199–210.
- Carrera, J. and S. P. Neuman, 1986b: Estimation of aquifer parameters under transient and steady state conditions, 2: Uniqueness, stability and solution algorithms. *Water Resour. Res.*, 22(2), 211–227.
- Carrera, J. and S. P. Neuman, 1986c: Estimation of aquifer parameters under transient and steady state conditions, 3: Application to synthetic and field data. *Water Resour. Res.*, 22(2), 228–242.
- Chavent, G., M. Dupuy and P. Lemonnier, 1975: History matching by use of optimal control theory. *Soc. Petrol. Eng. J.*, 15(1) 74–86.

- Courtier, P. and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation - Part II, Numerical results. *Quart. J. Roy. Meteor. Soc.*, 113, 1329–1348.
- Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, 87, 367–374.
- Derber, J. C., 1985: The variational four-dimensional assimilation of analyses using filtered models as constraints. Ph. D. dissertation, University of Wisconsin, Madison, Wisconsin, USA.
- Derber, J. C., 1987: Variational four-dimensional analysis using quasi-geostrophic constraints. *Mon. Wea. Rev.*, 115, 998–1008.
- Eliassen, A., 1954: Provisional report on calculation of spatial covariance and autocovariance of the pressure field. Videnskapsakademiets Institutt for Vær og Klimaforskning, report no. 5. (Reprinted in *Dynamic meteorology: Data assimilation methods*. L. Bengtsson, M. Ghil and E. Källén, Eds. Springer Verlag, New York, 1981, pp 319–330.)
- Fletcher, R., 1970: A new approach to variable metric algorithms. *Computer J.*, 13, 317–322.
- Gandin, L. S., 1965: Objective analysis of meteorological fields (translated from Russian). Israel program for scientific translations, Jerusalem, 242pp.
- Ghil, M. S., E. Cohn, J. Tavantis, K. Bube and E. Isaacson, 1981: Application of estimation theory to numerical weather prediction. In *Dynamic meteorology: Data assimilation methods*. L. Bengtsson, M. Ghil and E. Källén, Eds. Springer-Verlag, 139–224.
- Gilchrist, B. and G. P. Cressman, 1954: An experiment in objective analysis. *Tellus*, 6, 97–101.

- Gill, P. E., W. Murray and M. H. Wright, 1981: Practical optimization, Academic Press, 401 pp.
- Gill, P. E., M. A. Saunders and M. S. Wright, 1982: A note on a sufficient decrease criterion for a non-derivative step length procedure. *Math. Program.*, 23, 349–352.
- Goldfarb, D., 1970: A family of variable metric methods derived by variational means. *Math. Comput.*, 24, 23–26.
- Goldstein, A. A. and J. F. Price, 1967: An effective algorithm for minimization. *Numer. Math.*, 10, 184–189.
- Harlan Jr., J. and J. J. O'Brien, 1986: Assimilation of scatterometer winds into surface pressure fields using a variational method. *J. Geophys. Res.*, 91, 7816–7836.
- Hestenes, M. R., 1980: Conjugate direction methods in optimization. *Applications of Mathematics*, 12, Springer Verlag, 325pp.
- Hestenes, M. R. and E. Stiefel, 1952: Methods of conjugate-gradients for solving linear systems. *J. Res. Natl. Bur. Stand.*, 48, 409–436.
- Hoffman, R. N., 1982: SASS wind ambiguity removal by direct minimization. *Mon. Wea. Rev.*, 110, 434–445.
- Hoffman, R. N., 1984: SASS wind ambiguity removal by direct minimization. Part II: Use of smoothness and dynamic constraints. *Mon. Wea. Rev.*, 112, 1829–1852.
- Hoffman, R. N., 1986: A four-dimensional analysis exactly satisfying equations of motion. *Mon. Wea. Rev.*, 114, 388–397.
- Holland, W. R. and P. Malanotte-Rizzoli, 1989: Along-track assimilation of altimeter data into an ocean circulation model; space versus time resolution studies. Submitted to *J. Phys. Oceanogr.*

- Hurlburt, H. E., 1986: Dynamic transfer of simulated altimeter data into subsurface information by a numerical ocean model. *J. Geophys. Res.*, 91, 2372–2400.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Engr. (Trans. ASME)*, 82D, 35–45.
- Kalman, R. E. and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *J. Basic Engr. (Trans. ASME)*, 83D, 95–108.
- Kindle J. G., 1986: Sampling strategies and model assimilation of altimetric data for ocean monitoring and prediction. *J. Geophys. Res.*, 91, 2418–2432.
- Kimeldorf, G. and G. Wahba, 1970: A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.*, 41, 495–502.
- Kitamura, S. and S. Nakagiri, 1977: Identifiability of spatially varying and constant parameters in distributed systems of parabolic type. *SIAM, J. Contr. and Optimiz.*, 15, 785–802.
- Kubota, M. and J. J. O'Brien, 1988: Variability of the upper tropical Pacific Ocean model. *J. Geophys. Res.*, 93, 13930–13940.
- Le Dimet, F. X. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, 38A, 97–110.
- Le Dimet, F. X. and I. M. Navon, 1989: Variational and optimization methods in meteorology - a review. Submitted to *Mon. Wea. Rev.*.
- Legler, D. M. and J. J. O'Brien, 1986: Analysis of near real-time wind stress data from the tropical Pacific Ocean. *Atmosphere Newsletter*, February 1986.
- Legler, D. M., I. M. Navon and J. J. O'Brien, 1989: Objective analysis of pseudo-stress over the Indian Ocean using a direct-minimization approach. *Mon. Wea. Rev.*, 117, 709–720.



- Lewis, J. M. and S. C. Bloom, 1978: Incorporation of time continuity into subsynoptic analysis by using dynamical constraints. *Tellus*, 30, 496–516.
- Lewis, J. M. and J. C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, 37A, 309–322.
- Lions, J. L., 1971: Optimal control of systems governed by partial differential equations. Springer-Verlag, Berlin, 396pp.
- Liu D. C. and J. Nocedal, 1988: On the limited BFGS method for large scale optimization. Technical Report NAM 03, Northwestern University, Dept. of Electrical Engineering and Computer Science, 26pp.
- Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, 109, 701–721.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 112, 1177–1194.
- Lorenc, A. C., 1988: Optimal nonlinear objective analysis. *Quart. J. Roy. Meteor. Soc.*, 114, 205–240.
- Luenberger, D. C., 1984: Linear and nonlinear programming, Addison-Wesley, 491 pp.
- Malanotte-Rizzoli, P. and W. R. Holland, 1986: Data constraints applied to models of the ocean general circulation. Part I: The steady case. *J. Phys. Oceanogr.*, 16 1665–1682.
- Malanotte-Rizzoli, P. and W. R. Holland, 1988: Data constraints applied to models of the ocean general circulation. Part II: The transient, eddy-resolving case. *J. Phys. Oceanogr.*, 18, 1093–1107.

- Malanotte-Rizzoli, P., R. E. Young and D. B. Haidvogel, 1989: Initialization and data assimilation experiments with a primitive equation model. To appear in *Dynamics of Atmospheres and Oceans*.
- Marshall, J. C., 1985: Determining the ocean circulation and improving the geoid from satellite altimetry. *J. Phys. Oceanogr.*, 15, 330-349.
- Miller, R. N., 1986: Toward the application of the Kalman filter to regional open ocean modelling. *J. Phys. Oceanogr.*, 16, 72-86.
- Moore, A. M., N. S. Cooper and D. L. T. Anderson, 1987: Initialization and data assimilation in models of the Indian Ocean. *J. Phys. Oceanogr.*, 17, 1965-1977.
- Morel, P., G. Lefevre and G. Rabreau, 1971: On initialization and non-synoptic data assimilation. *Tellus*, 23, 197-206.
- Navon, I. M., 1986: A review of variational and optimization methods in meteorology, in *Variational methods in geosciences*, Y. Sasaki ed., Elsevier, New York, 29-34.
- Navon, I. M. and R. de Villiers, 1983: Combined penalty-multiplier optimization methods to enforce integral invariants conservation. *Mon. Wea. Rev.*, 111, 1228-1243.
- Navon, I. M. and D. Legler, 1987: Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Wea. Rev.*, 115, 1479-1502.
- Neuman, S. P., 1980: A statistical approach to the inverse problem of aquifer hydrology, 3. Improved solution method and added perspective. *Water Resour. Res.*, 16(2), 331-346.
- Panchang, V. G. and J. J. O'Brien, 1988: On the determination of hydraulic model parameters using the strong constraint formulation. To appear in *Ocean Modelling*-CRC Press.

- Panofsky, H. A., 1949: Objective weather map analysis. *J. Met.*, vol 6, no. 6, 386-392.
- Powell, M. J. D., 1977: Restart procedures for the conjugate gradient method. *Math. Program.* 12, 241-254.
- Provost, C., 1983: A variational method for estimating the general circulation of the ocean. Ph. D. thesis U. of Cal. San Diego, 95pp.
- Provost, C. and R. Salmon, 1986: A variational method for inverting hydrographic data. *J. Mar. Res.*, 44, 1-34.
- Richtmyer, R. D. and K. W. Morton, 1967: *Difference methods for initial-value problems*. Second edition, Interscience Publishers.
- Sandstrøm, J. W. and B. Helland-Hansen, 1903: Über die berechnung von meerestromungen. Report on Norwegian fishery and marine investigations, 2 (1902): 4, 43pp.
- Sasaki, Y., 1955: A variational study of the numerical prediction based on the variational principle. *J. Meteor. Soc. Japan*, 33, 262-275.
- Sasaki, Y., 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, 36, 738-742.
- Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, 98, 875-883.
- Schröter, J. and C. Wunsch, 1986: Solution of nonlinear finite difference ocean models by optimization methods with sensitivity and observational strategy analysis. *J. Phys. Oceanogr.*, 16, 1855-1874.
- Semtner A. J., 1974: An oceanic general circulation model with bottom topography. University of California, L. A. Tech. Rep. No. 9.
- Shanno, D. F., 1970: Conditioning of quasi-Newton methods for function minimization. *Math. Comput.*, 24, 647-657.

- Shanno, D. F., 1978a: Conjugate-gradient methods with inexact searches. *Math. Operations Res.*, 3, 244–256.
- Shanno, D. F., 1978b: On the convergence of a new conjugate-gradient method. *SIAM J. Numer. Anal.*, 15, 1247–1257.
- Shanno, D. F. and K. H. Phua, 1980: Remark on algorithm 500 - a variable method for unconstrained nonlinear minimization. *ACM Trans. on Mathematical Software*, 6, 618–622.
- Stommel, H. and F. Schott, 1977: The beta spiral and the determination of the absolute velocity field from hydrographic data. *Deep Sea Res.*, 24, 325–329.
- Talagrand, O., 1981: A study of the dynamics of four-dimensional data assimilation. *Tellus*, 33, 43–60.
- Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation - Part I, Theory. *Quart. J. Roy. Meteor. Soc.*, 113, 1311–1328.
- Thacker, W. C., 1986: Relationships between statistical and deterministic methods of data assimilation, in *Variational methods in the geosciences*, Y. Sasaki ed., Elsevier, New York, 173–179.
- Thacker, W. C., 1988: Fitting models to inadequate data by enforcing spatial and temporal smoothness. *J. Geophys. Res.*, 93, 10655–10665.
- Thacker, W. C. and R. B. Long, 1988: Fitting dynamics to data. *J. Geophys. Res.*, 93, 1227–1240.
- Thompson, P. D., 1969: Reduction of analysis error through constraints of dynamical consistency. *J. Applied Meteor.* 8, 738–742.
- Wahba, G. and J. Wendelberger, 1980: Some new methods for variational objective analysis using splines and cross validation. *Mon. Wea. Rev.*, 108, 1122–1143.

- Webb, D. J. and A. Moore, 1986: Assimilation of altimeter data into ocean models. *J. Phys. Oceanogr.*, 16, 1901–1913.
- Wunsch, C., 1977: Determining the general circulations of the oceans: a preliminary discussion. *Science*, 196, 871–875.
- Wunsch, C., 1978: The North Atlantic general circulation west of 50°W determined by inverse methods. *Rev. Geophys. Space Phys.*, 16, 583–620.
- Yakowitz, S. and L. Duckstein, 1980: Instability in aquifer identification: Theory and case studies. *Water Resour. Res.*, 16(6), 1045–1064.
- Yeh, W. W-G., 1986: Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resour. Res.* 22(2), 95–108.

## Appendix A: Derivation of the continuous adjoint equations

From section 6 the Lagrangian is given by: (Note that in spherical coordinates  $d\sigma = a^2 \cos\theta d\theta d\phi dt$ .)

$$\begin{aligned}
 L(U, V, h, \lambda_u, \lambda_v, \lambda_h, c^2) = & \\
 & \int_{\Sigma} \lambda_u \left[ \frac{\partial U}{\partial t} - fV + \frac{c^2}{a \cos\theta} \frac{\partial h}{\partial \phi} - \frac{\tau^\phi}{\rho} - A\nabla^2 U \right] d\sigma \\
 & + \int_{\Sigma} \lambda_v \left[ \frac{\partial V}{\partial t} + fU + \frac{c^2}{a} \frac{\partial h}{\partial \theta} - \frac{\tau^\theta}{\rho} - A\nabla^2 V \right] d\sigma \quad (A-1) \\
 & + \int_{\Sigma} \lambda_h \left[ \frac{\partial h}{\partial t} + \frac{1}{a \cos\theta} \left( \frac{\partial U}{\partial \phi} + \frac{\partial}{\partial \theta} (V \cos\theta) \right) \right] d\sigma \\
 & + \int_{\Sigma} \frac{K_h}{2} (h - h')^2 d\sigma + \int_{\Sigma} \frac{K_c}{2} (c^2 - c'^2)^2 d\sigma
 \end{aligned}$$

where  $\lambda_u, \lambda_v, \lambda_h$  are the Lagrangian multipliers for the U, V and h equation, respectively.  $K_h$  and  $K_c$  are specified (constant) validity coefficients,  $c^2$  is a function of longitude only.

Letting the first order variation of  $L$  with respect to  $\lambda_u, \lambda_v$  and  $\lambda_h$  vanish, give the following equations:

$$\frac{\partial U}{\partial t} - fV + \frac{c^2}{a \cos\theta} \frac{\partial h}{\partial \phi} - \frac{\tau^\phi}{\rho} - A\nabla^2 U = 0 \quad (A-2)$$

$$\frac{\partial V}{\partial t} + fU + \frac{c^2}{a} \frac{\partial h}{\partial \theta} - \frac{\tau^\theta}{\rho} - A\nabla^2 V = 0 \quad (A-3)$$

$$\frac{\partial h}{\partial t} + \frac{1}{a \cos\theta} \left( \frac{\partial U}{\partial \phi} + \frac{\partial}{\partial \theta} (V \cos\theta) \right) = 0 \quad (A-4)$$

These three equations are identical to the original model equations (5-1) – (5-3).

In order to find the equations for the Lagrangian multipliers, the adjoint equations, one has to let the first variation of  $L$  with respect to  $U, V$  and  $h$  vanish. It is necessary to perform the following integrations by parts.

$$\frac{\partial}{\partial U} \int_{\Sigma} (\lambda_u \frac{\partial U}{\partial t}) d\sigma = \frac{\partial}{\partial U} \left[ \int_{\Sigma} \frac{\partial}{\partial t} (\lambda_u U) d\sigma - \int_{\Sigma} \frac{\partial \lambda_u}{\partial t} U d\sigma \right] = - \int_{\Sigma} \frac{\partial \lambda_u}{\partial t} d\sigma \quad (A-5)$$

where the initial condition for the adjoint variable  $\lambda_u = 0$  at  $t = T$  is used.  $T$  represents the time at the end of the integration period. At  $t = 0$ ,  $U$  is assumed to be zero. Similar integrations of the other two terms involving time differentiation give

$$\frac{\partial}{\partial V} \int_{\Sigma} (\lambda_v \frac{\partial V}{\partial t}) d\sigma = - \int_{\Sigma} \frac{\partial \lambda_v}{\partial t} d\sigma \quad (A-6)$$

$$\frac{\partial}{\partial h} \int_{\Sigma} (\lambda_h \frac{\partial h}{\partial t}) d\sigma = - \int_{\Sigma} \frac{\partial \lambda_h}{\partial t} d\sigma \quad (A-7)$$

where the initial conditions  $\lambda_v = \lambda_h = 0$  at  $t = T$  are used and  $V$  and  $h$  are assumed to be zero at  $t = 0$ .

From the continuity equation one has

$$\frac{\partial}{\partial U} \int_{\Sigma} \frac{1}{a \cos \theta} (\lambda_h \frac{\partial U}{\partial \phi}) d\sigma = \frac{\partial}{\partial U} \left[ \int_{\Sigma} \frac{1}{a \cos \theta} \left( \frac{\partial}{\partial \phi} (\lambda_h U) - \frac{\partial \lambda_h}{\partial \phi} U \right) d\sigma \right] = - \int_{\Sigma} \frac{1}{a \cos \theta} \frac{\partial \lambda_h}{\partial \phi} d\sigma \quad (A-8)$$

where the boundary condition  $U=0$  at the eastern and western end of the basin has been used. Similarly one finds

$$\begin{aligned}
& \frac{\partial}{\partial V} \int_{\Sigma} \frac{1}{a \cos \theta} (\lambda_h \frac{\partial}{\partial \theta} (V \cos \theta)) d\sigma = \\
& \frac{\partial}{\partial V} \left[ \int_t \int_{\phi} \int_{\theta} a \frac{\partial}{\partial \theta} (\lambda_h V \cos \theta) d\theta d\phi dt - \int_t \int_{\phi} \int_{\theta} a V \cos \theta \frac{\partial \lambda_h}{\partial \theta} d\theta d\phi dt \right] = \quad (A-9) \\
& - \int_{\Sigma} \frac{1}{a} \frac{\partial \lambda_h}{\partial \theta} d\sigma
\end{aligned}$$

where it has been used that  $V = 0$  at a solid northern and southern boundary. The question of the condition at the open boundaries will be delayed until the finite difference equations are derived, but for the time being the extra terms are neglected.

The pressure terms give the following two integrations

$$\begin{aligned}
& \frac{\partial}{\partial h} \int_{\Sigma} \frac{c^2}{a \cos \theta} (\lambda_u \frac{\partial h}{\partial \phi}) d\sigma = \\
& \frac{\partial}{\partial h} \left[ \int_t \int_{\phi} \int_{\theta} a \frac{\partial}{\partial \phi} (c^2 \lambda_u h) d\theta d\phi dt - \int_t \int_{\phi} \int_{\theta} c^2 a h \frac{\partial \lambda_u}{\partial \phi} d\theta d\phi dt \right. \\
& \left. - \int_t \int_{\phi} \int_{\theta} a h \lambda_u \frac{\partial c^2}{\partial \phi} d\theta d\phi dt \right] = - \int_{\Sigma} \frac{c^2}{a \cos \theta} \frac{\partial \lambda_u}{\partial \phi} d\sigma - \int_{\Sigma} \frac{\lambda_u}{a \cos \theta} \frac{\partial c^2}{\partial \phi} d\sigma \quad (A-10)
\end{aligned}$$

where the boundary condition  $\lambda_u = 0$  at the eastern and the western end of the basin has been used, and

$$\begin{aligned}
& \frac{\partial}{\partial h} \int_{\Sigma} \frac{c^2}{a} (\lambda_v \frac{\partial h}{\partial \theta}) d\sigma = \frac{\partial}{\partial h} \left( \int_t \int_{\phi} \int_{\theta} c^2 a \frac{\partial}{\partial \theta} (\lambda_v h \cos \theta) d\theta d\phi dt \right. \\
& \left. - \int_t \int_{\phi} \int_{\theta} c^2 a \cos \theta h \frac{\partial \lambda_v}{\partial \theta} d\theta d\phi dt + \int_t \int_{\phi} \int_{\theta} c^2 a h \lambda_v \sin \theta d\theta d\phi dt \right) = \quad (A-11) \\
& - \int_{\Sigma} \frac{c^2}{a \cos \theta} \frac{\partial}{\partial \theta} (\lambda_v \cos \theta) d\sigma
\end{aligned}$$



where it is assumed that  $\lambda_v = 0$  at a solid northern and southern boundary. A discussion of the term at the open boundary will be delayed until the finite difference equations are derived. For the time being the extra terms are neglected.

The Laplacian friction term can be written as:

$$\begin{aligned}
 & A \frac{\partial}{\partial U} \int_{\Sigma} (\lambda_u \nabla^2 U) d\sigma = \\
 & A \frac{\partial}{\partial U} \int_{\Sigma} \left( \lambda_u \frac{1}{a^2 \cos^2 \theta} \left[ \frac{\partial^2}{\partial \phi^2} + \cos \theta \frac{\partial}{\partial \theta} (\cos \theta \frac{\partial}{\partial \theta}) \right] U \right) d\sigma
 \end{aligned} \tag{A-12}$$

where the simplified  $\nabla^2$ -operator given in (5-4) has been used. The first term in (A-12) gives:

$$\begin{aligned}
 & A \frac{\partial}{\partial U} \int_t \int_{\phi} \int_{\theta} \frac{1}{\cos \theta} \lambda_u \frac{\partial^2 U}{\partial \phi^2} d\theta d\phi dt = \\
 & A \frac{\partial}{\partial U} \left[ \int_t \int_{\phi} \int_{\theta} \frac{1}{\cos \theta} \frac{\partial}{\partial \phi} (\lambda_u \frac{\partial U}{\partial \phi}) d\theta d\phi dt \right. \\
 & \left. - \int_t \int_{\phi} \int_{\theta} \frac{1}{\cos \theta} \frac{\partial \lambda_u}{\partial \phi} \frac{\partial U}{\partial \phi} d\theta d\phi dt \right] = \\
 & - A \frac{\partial}{\partial U} \left[ \int_t \int_{\phi} \int_{\theta} \frac{1}{\cos \theta} \frac{\partial}{\partial \phi} \left( \frac{\partial \lambda_u}{\partial \phi} U \right) d\theta d\phi dt \right. \\
 & \left. - \int_t \int_{\phi} \int_{\theta} \frac{1}{\cos \theta} U \frac{\partial^2 \lambda_u}{\partial \phi^2} d\theta d\phi dt \right] = A \int_{\Sigma} \frac{1}{a^2 \cos^2 \theta} \frac{\partial^2 \lambda_u}{\partial \phi^2} d\sigma
 \end{aligned} \tag{A-13}$$

where it is assumed that  $\lambda_u = 0$  at the eastern and the western end of the basin (where the no-slip boundary condition is used for the velocity component).

Last term in (A - 12):

$$\begin{aligned}
& A \frac{\partial}{\partial U} \left[ \int_t \int_\phi \int_\theta \lambda_u \cos\theta \frac{\partial^2 U}{\partial \theta^2} d\theta d\phi dt - \int_t \int_\phi \int_\theta \lambda_u \sin\theta \frac{\partial U}{\partial \theta} d\theta d\phi dt \right] = \\
& A \frac{\partial}{\partial U} \left[ \int_t \int_\phi \int_\theta \frac{\partial}{\partial \theta} (\lambda_u \cos\theta \frac{\partial U}{\partial \theta}) d\theta d\phi dt - \int_t \int_\phi \int_\theta \frac{\partial \lambda_u}{\partial \theta} \cos\theta \frac{\partial U}{\partial \theta} d\theta d\phi dt \right. \\
& + \left. \int_t \int_\phi \int_\theta \lambda_u \sin\theta \frac{\partial U}{\partial \theta} d\theta d\phi dt \right] - A \frac{\partial}{\partial U} \left[ \int_t \int_\phi \int_\theta \frac{\partial}{\partial \theta} (\lambda_u U \sin\theta) d\theta d\phi dt \right. \\
& - \left. \int_t \int_\phi \int_\theta \frac{\partial \lambda_u}{\partial \theta} U \sin\theta d\theta d\phi dt - \int_t \int_\phi \int_\theta \lambda_u U \cos\theta d\theta d\phi dt \right] = \\
& A \frac{\partial}{\partial U} \left[ - \int_t \int_\phi \int_\theta \frac{\partial}{\partial \theta} \left( \frac{\partial \lambda_u}{\partial \theta} U \cos\theta \right) d\theta d\phi dt + \int_t \int_\phi \int_\theta \frac{\partial^2 \lambda_u}{\partial \theta^2} U \cos\theta d\theta d\phi dt \right. \quad (A - 14) \\
& - \int_t \int_\phi \int_\theta \frac{\partial \lambda_u}{\partial \theta} U \sin\theta d\theta d\phi dt + \int_t \int_\phi \int_\theta \frac{\partial}{\partial \theta} (\lambda_u U \sin\theta) d\theta d\phi dt \\
& - \left. \int_t \int_\phi \int_\theta \frac{\partial \lambda_u}{\partial \theta} U \sin\theta d\theta d\phi dt - \int_t \int_\phi \int_\theta \lambda_u U \cos\theta d\theta d\phi dt \right] \\
& - A \frac{\partial}{\partial U} \left[ - \int_t \int_\phi \int_\theta \frac{\partial \lambda_u}{\partial \theta} U \sin\theta d\theta d\phi dt - \int_t \int_\phi \int_\theta \lambda_u U \cos\theta d\theta d\phi dt \right] = \\
& A \int_\Sigma \frac{1}{a^2} \frac{\partial^2 \lambda_u}{\partial \theta^2} d\sigma - A \int_\Sigma \frac{1}{a^2 \cos^2 \theta} \frac{\partial \lambda_u}{\partial \theta} \cos\theta \sin\theta d\sigma
\end{aligned}$$

where it has been used that  $\lambda_u = 0$  at a solid northern and southern boundary. It is also assumed that  $U = 0$  at a solid northern and southern boundary. Again the discussion of what happens at the open boundary will be delayed until the finite difference equations are derived. For the time being the extra terms are neglected.

A similar integration of the friction term for V gives

$$A \int_\Sigma \frac{1}{a^2 \cos^2 \theta} \frac{\partial^2 \lambda_v}{\partial \phi^2} d\sigma \quad (A - 15)$$

$$A \int_\Sigma \frac{1}{a^2} \frac{\partial^2 \lambda_v}{\partial \theta^2} d\sigma - A \int_\Sigma \frac{1}{a^2 \cos^2 \theta} \frac{\partial \lambda_v}{\partial \theta} \cos\theta \sin\theta d\sigma \quad (A - 16)$$

where it is assumed that  $\lambda_v = 0$  at the eastern and the western end of the basin.

Letting the first-order variation of  $L$  with respect to  $U$ ,  $V$  and  $h$  vanish, gives the adjoint equations:

$$-\frac{\partial \lambda_u}{\partial t} + f \lambda_v - \frac{1}{a \cos \theta} \frac{\partial \lambda_h}{\partial \phi} - A \nabla^2 \lambda_u = 0 \quad (A-17)$$

$$-\frac{\partial \lambda_v}{\partial t} - f \lambda_u - \frac{1}{a} \frac{\partial \lambda_h}{\partial \theta} - A \nabla^2 \lambda_v = 0 \quad (A-18)$$

$$\begin{aligned} & -\frac{\partial \lambda_h}{\partial t} - \frac{c^2}{a \cos \theta} \left[ \frac{\partial \lambda_u}{\partial \phi} + \frac{\lambda_u}{c^2} \frac{\partial c^2}{\partial \phi} + \frac{\partial}{\partial \theta} (\lambda_v \cos \theta) \right] \\ & + K_h (h - h') = 0 \end{aligned} \quad (A-19)$$

By rescaling  $\lambda_u$  and  $\lambda_v$  in the following way:

$$\begin{aligned} \lambda'_u &= \lambda_u \cdot c^2 \\ \lambda'_v &= \lambda_v \cdot c^2 \end{aligned} \quad (A-20)$$

equations (A-17) – (A-19) can be written as (dropping the primes):

$$-\frac{\partial \lambda_u}{\partial t} + f \lambda_v - \frac{c^2}{a \cos \theta} \frac{\partial \lambda_h}{\partial \phi} - A \nabla^2 \lambda_u = 0 \quad (A-21)$$

$$-\frac{\partial \lambda_v}{\partial t} - f \lambda_u - \frac{c^2}{a} \frac{\partial \lambda_h}{\partial \theta} - A \nabla^2 \lambda_v = 0 \quad (A-22)$$

$$-\frac{\partial \lambda_h}{\partial t} - \frac{1}{a \cos \theta} \left[ \frac{\partial \lambda_u}{\partial \phi} + \frac{\partial}{\partial \theta} (\lambda_v \cos \theta) \right] + K_h (h - h') = 0 \quad (A-23)$$

The adjoint equations (A-21) – (A-23) now have the same form as the original model equations (A-2) – (A-4).

From the term resulting from differentiation with respect to the parameter  $c^2$  gives:

$$\int_t \int_\theta \left[ \frac{\lambda_u}{c^2 a \cos \theta} \frac{\partial h}{\partial \phi} + \frac{\lambda_v}{c^2 a} \frac{\partial h}{\partial \theta} \right] d\theta dt + T L_\theta K_c (c^2 - c^{2'}) = 0 \quad (A-24)$$

where the primes have been dropped.  $T$  represents the period over which the data are assimilated, while  $L_\theta$  is a length scale in the north-south direction.

## Appendix B: Derivation of the finite difference adjoint equations

The finite difference equations for the variational formalism will be derived.

The following finite differencing operators can be defined:

$$D_{p+}\alpha_p = \frac{\alpha_{p+1} - \alpha_p}{h} \quad (B-1)$$

$$D_{p-}\alpha_p = \frac{\alpha_p - \alpha_{p-1}}{h} \quad (B-2)$$

$$D_{p0}\alpha_p = \frac{1}{2}[D_{p+}\alpha_p + D_{p-}\alpha_p] = \frac{\alpha_{p+1} - \alpha_{p-1}}{2h} \quad (B-3)$$

where for  $p = i \Rightarrow h = \Delta\phi$ ,  $p = j \Rightarrow h = \Delta\theta$  and  $p = n \Rightarrow h = \Delta t$ .  $\Delta\phi$ ,  $\Delta\theta$  are the spatial gridspacing in the  $\phi$  and  $\theta$  direction and  $\Delta t$  is the timestep. Two useful relationships between  $D_{p+}$  and  $D_{p-}$  are

$$D_{p+}\alpha_{p-1} = D_{p-}\alpha_p \quad (B-4)$$

$$D_{p-}\alpha_{p+1} = D_{p+}\alpha_p \quad (B-5)$$

The following formulae for summation by parts are going to be used when the first order variation of the Lagrangian with respect to  $\lambda_{u,i,j}^n$ ,  $\lambda_{v,i,j}^n$  and  $\lambda_{h,i,j}^n$  is evaluated.

$$D_{p+}(\alpha\gamma)_p = \alpha_p D_{p+}\gamma_p + \gamma_{p+1} D_{p+}\alpha_p \quad (B-6)$$

$$D_{p-}(\alpha\gamma)_p = \alpha_p D_{p-}\gamma_p + \gamma_{p-1} D_{p-}\alpha_p \quad (B-7)$$

A summation by parts will result in boundary terms and the range of the summation may have to be modified. The following notation is convenient, Richtmyer and Morton (1967)

$$\alpha_i \gamma_j \Big|_r^s = \alpha_{s+i} \gamma_{s+j} - \alpha_{r-i} \gamma_{r-j} \quad (B-8)$$

The formulae for summation by parts may take different forms, but they can be derived by summing the identities (B-6) – (B-7) to get

$$\sum_{j=r}^s \alpha_j D_{j+} \gamma_j = - \sum_{j=r+1}^{s+1} \gamma_j D_{j-} \alpha_j + \alpha_0 \gamma_0 \Big|_r^{s+1} \quad (B-9)$$

$$\sum_{j=r}^s \alpha_j D_{j-} \gamma_j = - \sum_{j=r-1}^{s-1} \gamma_j D_{j+} \alpha_j + \alpha_0 \gamma_0 \Big|_{r-1}^s \quad (B-10)$$

The finite difference equation for the Lagrangian in (5-6) can be written as

$$\begin{aligned} L(U_{i,j}^n, V_{i,j}^n, h_{i,j}^n, \lambda_{u,i,j}^n, \lambda_{v,i,j}^n, \lambda_{h,i,j}^n, c_i^2) &= \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=1}^N \lambda_{u,i,j}^n \left[ D_{n0} U_{i,j}^n \right. \\ &\quad - \frac{f}{4} (V_{i,j}^n + V_{i+1,j}^n + V_{i,j+1}^n + V_{i+1,j+1}^n) \\ &\quad + \frac{c_i^2}{a \cos \theta_j} D_{i+} h_{i,j}^n - \frac{\tau_{i,j}^\phi}{\rho} - \frac{A}{a^2} \left( \frac{1}{\cos^2 \theta_j} (D_{i+} D_{i-}) U_{i,j}^n \right. \\ &\quad \left. \left. + (D_{j+} D_{j-}) U_{i,j}^n - \frac{\sin \theta_j}{\cos \theta_j} D_{j0} U_{i,j}^n \right) \right] a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \\ &\quad + \sum_{i=2}^I \sum_{j=2}^J \sum_{n=1}^N \lambda_{v,i,j}^n \left[ D_{n0} V_{i,j}^n \right. \\ &\quad + \frac{f}{4} (U_{i-1,j-1}^n + U_{i,j-1}^n + U_{i-1,j}^n + U_{i,j}^n) \\ &\quad + \frac{c_i^2}{a} D_{j-} h_{i,j}^n - \frac{\tau_{i,j}^\theta}{\rho} - \frac{A}{a^2} \left( \frac{1}{\cos^2 \theta_j} (D_{i+} D_{i-}) V_{i,j}^n \right. \\ &\quad \left. \left. + (D_{j+} D_{j-}) V_{i,j}^n - \frac{\sin \theta_j}{\cos \theta_j} (D_{j0} V_{i,j}^n) \right) \right] a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \\ &\quad + \sum_{i=2}^I \sum_{j=1}^{J-1} \sum_{n=1}^N \left[ \lambda_{h,i,j}^n \left[ D_{n0} h_{i,j}^n \right. \right. \\ &\quad \left. \left. + \frac{1}{a \cos \theta_j} \left( D_{i-} U_{i,j}^n + D_{j+} (V_{i,j}^n \cos \theta_j) \right) \right] \right. \\ &\quad \left. + \frac{K_h}{2} (h_{i,j}^n - h_{i,j}^{n'})^2 + \frac{K_c}{2} (c_i^2 - c_i^{2'})^2 \right] a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \end{aligned} \quad (B-11)$$

where a leap frog time differencing scheme is used. To start up the integration a forward timestep is used. This will be represented in the above equation by changing the operator  $D_{n0}$  to  $D_{n+}$ . This forward timestep is also used every  $99^{th}$  timestep in order to eliminate the computational mode. The spatial derivatives are estimated by centered differences using an Arakawa C-grid, Arakawa (1966). Superscript  $n$  denotes the timelevel, while subscripts  $(i, j)$  represent the gridpoint.  $\lambda_{u,i,j}^n$ ,  $\lambda_{v,i,j}^n$  and  $\lambda_{h,i,j}^n$  are the Lagrangian multipliers for  $U_{i,j}^n, V_{i,j}^n$  and  $h_{i,j}^n$ , respectively.  $K_h$  and  $K_c$  are specified (constant) validity coefficients. The parameter  $c_i^2$  is taken to be a function of longitude only. For a description of the other parameters and variables, see section 6.

Letting the first-order variation of (B - 11) with respect to  $\lambda_{u,i,j}^n$ ,  $\lambda_{v,i,j}^n$  and  $\lambda_{h,i,j}^n$  vanish, give the following equations

$$\begin{aligned} & D_{n0}U_{i,j}^n - \frac{f}{4}(V_{i,j}^n + V_{i+1,j}^n + V_{i,j+1}^n + V_{i+1,j+1}^n) \\ & + \frac{c_i^2}{a \cos \theta_j} D_{i+} h_{i,j}^n - \frac{\tau_{i,j}^\phi}{\rho} - \frac{A}{a^2} \left( \frac{1}{\cos^2 \theta_j} (D_{i+} D_{i-}) U_{i,j}^n \right. \\ & \left. + (D_{j+} D_{j-}) U_{i,j}^n - \frac{\sin \theta_j}{\cos \theta_j} D_{j0} U_{i,j}^n \right) = 0 \end{aligned} \quad (B - 12)$$

$$\begin{aligned} & D_{n0}V_{i,j}^n + \frac{f}{4}(U_{i-1,j-1}^n + U_{i,j-1}^n + U_{i-1,j}^n + U_{i,j}^n) \\ & + \frac{c_i^2}{a} D_{j-} h_{i,j}^n - \frac{\tau_{i,j}^\theta}{\rho} - \frac{A}{a^2} \left( \frac{1}{\cos^2 \theta_j} (D_{i+} D_{i-}) V_{i,j}^n \right. \\ & \left. + (D_{j+} D_{j-}) V_{i,j}^n - \frac{\sin \theta_j}{\cos \theta_j} (D_{j0} V_{i,j}^n) \right) = 0 \end{aligned} \quad (B - 13)$$

$$D_{n0}h_{i,j}^n + \frac{1}{a \cos \theta_j} \left( D_{i-} U_{i,j}^n + D_{j+} (V_{i,j}^n \cos \theta_j) \right) = 0 \quad (B - 14)$$

These three equations are identical to the original model equations (5-1) - (5-3).

The equations for the Lagrangian multipliers, or the adjoint equations, are found by letting the first-order variation of (B - 11) with respect to  $U_{i,j}^n$ ,  $V_{i,j}^n$

and  $h_{i,j}^n$  vanish. In order to find the appropriate equations one must do several summations by parts using the formulae in (B-9) – (B-10). The terms resulting from differentiation with respect to  $U_{i,j}^n$  will be considered first. The first term of (B-11) gives using the forward timestep  $D_{n+}$

$$\begin{aligned}
& \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=1}^{N-1} \frac{\partial}{\partial U_{i,j}^n} (\lambda_{u,i,j}^n D_{n+} U_{i,j}^n) a^2 \cos\theta_j \Delta\phi \Delta\theta \Delta t = \\
& \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=2}^N \left[ \frac{\partial}{\partial U_{i,j}^n} \left( (\lambda_{u,i,j}^N U_{i,j}^N - \lambda_{u,i,j}^1 U_{i,j}^1) \right. \right. \\
& \left. \left. - U_{i,j}^n D_n - \lambda_{u,i,j}^n \right) \right] a^2 \cos\theta_j \Delta\phi \Delta\theta \Delta t = \\
& \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=2}^N \left( -D_n - \lambda_{u,i,j}^n \right) a^2 \cos\theta_j \Delta\phi \Delta\theta \Delta t
\end{aligned} \tag{B-15}$$

where it has been assumed that the initial condition for  $U_{i,j}^n$  is  $U_{i,j}^1 = 0$  and also that the Lagrangian multiplier  $\lambda_{u,i,j}^n$  is equal to zero at the end of the integration,  $\lambda_{u,i,j}^N = 0$ . The latter is the “initial” condition for the backward integration of the adjoint equations.  $\lambda_{u,i,j}^N = 0$  because there are no data after the last computational time level and the Lagrangian multipliers are forced by the data misfits.

For the leapfrog timestep one gets

$$\begin{aligned}
& \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=2}^{N-1} \frac{\partial}{\partial U_{i,j}^n} (\lambda_{u,i,j}^n D_{n0} U_{i,j}^n) a^2 \cos\theta_j \Delta\phi \Delta\theta \Delta t = \\
& \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=2}^{N-1} \left[ \frac{\partial}{\partial U_{i,j}^n} \left( (\lambda_{u,i,j}^N U_{i,j}^N - \lambda_{u,i,j}^1 U_{i,j}^1) \right. \right. \\
& \left. \left. - U_{i,j}^n D_{n0} \lambda_{u,i,j}^n \right) \right] a^2 \cos\theta_j \Delta\phi \Delta\theta \Delta t = \\
& \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=2}^{N-1} \left( -D_{n0} \lambda_{u,i,j}^n \right) a^2 \cos\theta_j \Delta\phi \Delta\theta \Delta t
\end{aligned} \tag{B-16}$$

where (B-3) has been used and the initial conditions  $U_{i,j}^1 = 0$  and  $\lambda_{u,i,j}^N = 0$ .

Similarly, differentiation of  $L$  with respect to  $V_{i,j}^n$  and  $h_{i,j}^n$  gives for the forward timestep

$$\sum_{i=2}^I \sum_{j=2}^J \sum_{n=2}^N \left( -D_n - \lambda_{v,i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \quad (B-17)$$

$$\sum_{i=2}^I \sum_{j=1}^{J-1} \sum_{n=2}^N \left( -D_n - \lambda_{h,i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \quad (B-18)$$

and for the leapfrog timestep

$$\sum_{i=2}^I \sum_{j=2}^J \sum_{n=2}^{N-1} \left( -D_{n0} \lambda_{v,i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \quad (B-19)$$

$$\sum_{i=2}^I \sum_{j=1}^{J-1} \sum_{n=2}^{N-1} \left( -D_{n0} \lambda_{h,i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \quad (B-20)$$

As for  $\lambda_{u,i,j}^N$ , it is assumed that  $\lambda_{v,i,j}^N = 0$  and  $\lambda_{h,i,j}^N = 0$ .

Differentiation of the Coriolis term with respect to  $U_{i,j}^n$  gives

$$\begin{aligned} & \sum_{i=2}^I \sum_{j=2}^J \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} \lambda_{v,i,j}^n \frac{f}{4} (U_{i-1,j-1}^n + U_{i,j-1}^n \\ & + U_{i-1,j}^n + U_{i,j}^n) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t = \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=1}^N \frac{f}{4} (\lambda_{v,i+1,j+1}^n \\ & + \lambda_{v,i,j+1}^n + \lambda_{v,i+1,j}^n + \lambda_{v,i,j}^n) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \end{aligned} \quad (B-21)$$

Differentiation of the first term of the friction term with respect to  $U_{i,j}^n$  gives

$$\begin{aligned} & \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} \left( \frac{A}{a^2 \cos^2 \theta_j} \lambda_{u,i,j}^n (D_{i+} + D_{i-}) U_{i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t = \\ & \sum_{i=3}^I \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} \left( \left( \frac{A}{\cos \theta_j} (\lambda_{u,I,j}^n D_{i-} - \lambda_{u,2,j}^n D_{i-}) \right. \right. \\ & \left. \left. - (D_{i-} - U_{i,j}^n) (D_{i-} - \lambda_{u,i,j}^n) \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t = \end{aligned}$$



$$\begin{aligned}
& \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} \left( \frac{A}{\cos\theta_j} (-U_{I,j}^n D_{i-} \lambda_{u,I,j}^n - U_{2,j}^n D_{i-} \lambda_{u,2,j}^n) \right. \\
& \left. + U_{i,j}^n D_{i-} D_{i+} \lambda_{u,i,j}^n \right) \Delta\phi\Delta\theta\Delta t = \\
& \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{A}{a^2 \cos^2\theta_j} \left( D_{i+} D_{i-} \lambda_{u,i,j}^n \right) a^2 \cos\theta_j \Delta\phi\Delta\theta\Delta t
\end{aligned} \quad (B-22)$$

where it has been used that  $\lambda_{u,i,j}^n = 0$  and  $U_{i,j}^n = 0$  at the eastern and western end of the basin.

The second term gives

$$\begin{aligned}
& \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} \left( \frac{A}{a^2} \lambda_{u,i,j}^n (D_{j+} D_{j-}) U_{i,j}^n \right) a^2 \cos\theta_j \Delta\phi\Delta\theta\Delta t = \\
& \sum_{i=2}^{I-1} \sum_{j=3}^J \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} A \left( (\lambda_{u,i,J}^n \cos\theta_J D_{j-} U_{i,J}^n - \lambda_{u,i,2}^n \cos\theta_2 D_{j-} U_{i,2}^n) \right. \\
& \left. - (D_{j-} U_{i,j}^n) (D_{j-} (\lambda_{u,i,j}^n \cos\theta_j)) \right) \Delta\phi\Delta\theta\Delta t = \\
& \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} A \left( (\lambda_{u,i,J}^n \cos\theta_J D_{j-} U_{i,J}^n - \lambda_{u,i,2}^n \cos\theta_2 D_{j-} U_{i,2}^n) (B-23) \right. \\
& \left. - (U_{i,J}^n D_{j-} (\lambda_{u,i,J}^n \cos\theta_J) - U_{i,2}^n D_{j-} (\lambda_{u,i,2}^n \cos\theta_2)) \right. \\
& \left. + U_{i,j}^n D_{j+} D_{j-} (\lambda_{u,i,j}^n \cos\theta_j) \right) \Delta\phi\Delta\theta\Delta t = \\
& \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{A}{a^2} \left( D_{j+} D_{j-} \lambda_{u,i,j}^n \right. \\
& \left. - \frac{2\sin\theta_j}{\cos\theta_j} D_{j0} \lambda_{u,i,j}^n - \lambda_{u,i,j}^n \right) a^2 \cos\theta_j \Delta\phi\Delta\theta\Delta t
\end{aligned}$$

where it has been used that  $\lambda_{u,i,j}^n = 0$  and  $U_{i,j}^n = 0$  at a solid northern and southern boundary. Expanding the boundary terms in (B-23) and using the open boundary conditions for  $U_{i,j}^n$ , it can be shown that at the open boundaries a condition identical to the open boundary condition used for  $U_{i,j}^n$ , Camerlengo and O'Brien (1980), must be used for  $\lambda_{u,i,j}^n$ .

The last term gives

$$\begin{aligned}
& \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} \left( \lambda_{u,i,j}^n \frac{A \sin \theta_j}{a^2 \cos \theta_j} D_{j0} U_{i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t = \\
& \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} A \left( (\lambda_{u,i,J}^n \sin \theta_J U_{i,J}^n - \lambda_{u,i,2}^n \sin \theta_2 U_{i,2}^n) \right. \\
& \left. - U_{i,j}^n D_{j0} (\lambda_{u,i,j}^n \sin \theta_j) \right) \Delta \phi \Delta \theta \Delta t = \\
& \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{A}{a^2} \left( -\frac{\sin \theta_j}{\cos \theta_j} D_{j0} \lambda_{u,i,j}^n - \lambda_{u,i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t
\end{aligned} \tag{B-24}$$

The last two terms for the friction can then be written as

$$\sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \sum_{n=1}^N \frac{A}{a^2} \left( D_{j+} D_{j-} \lambda_{u,i,j}^n - \frac{\sin \theta_j}{\cos \theta_j} D_{j0} \lambda_{u,i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \tag{B-25}$$

Differentiation of  $L$  with respect to  $\lambda_{v,i,j}^n$  gives similar expressions as (B-22) and (B-25) for the friction terms involving  $\lambda_{v,i,j}^n$ .

From the second term in the continuity equation one has

$$\begin{aligned}
& \sum_{i=2}^I \sum_{j=1}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial U_{i,j}^n} \left( \frac{1}{a \cos \theta_j} \lambda_{h,i,j}^n D_{i-} U_{i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t = \\
& \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=1}^N \left[ \frac{\partial}{\partial U_{i,j}^n} a \left( (\lambda_{h,I,j}^n U_{I,j}^n - \lambda_{h,2,j}^n U_{2,j}^n) \right. \right. \\
& \left. \left. - U_{i,j}^n D_{i+} \lambda_{h,i,j}^n \right) \right] \Delta \phi \Delta \theta \Delta t = \\
& \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=1}^N \left( -\frac{1}{a \cos \theta_j} D_{i+} \lambda_{h,i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t
\end{aligned} \tag{B-26}$$

where it has been used that  $U_{i,j}^n = 0$  at the solid boundaries at the eastern and western end of the basin.

Differentiation of the Coriolis term with respect to  $V_{i,j}^n$  gives

$$\begin{aligned} & \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial V_{i,j}^n} \left( -\lambda_{u,i,j}^n \frac{f}{4} (V_{i,j}^n + V_{i+1,j}^n \right. \\ & \left. + V_{i,j+1}^n + V_{i+1,j+1}^n) \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t = \sum_{i=2}^I \sum_{j=2}^J \sum_{n=1}^N \left( -\frac{f}{4} (\lambda_{u,i,j}^n \right. \\ & \left. + \lambda_{u,i-1,j}^n + \lambda_{u,i,j-1}^n + \lambda_{u,i-1,j-1}^n) \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \end{aligned} \quad (B-27)$$

From the third term in the continuity equation one has

$$\begin{aligned} & \sum_{i=2}^I \sum_{j=1}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial V_{i,j}^n} \left( \frac{1}{a \cos \theta_j} \lambda_{h,i,j}^n D_{j+} (V_{i,j}^n \cos \theta_j) \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t = \\ & \sum_{i=2}^I \sum_{j=2}^J \sum_{n=1}^N \left[ \frac{\partial}{\partial V_{i,j}^n} a \left( (\lambda_{h,i,J}^n V_{i,J}^n \cos \theta_J - \lambda_{h,i,2}^n V_{i,2}^n \cos \theta_2) \right. \right. \\ & \left. \left. - V_{i,j}^n \cos \theta_j D_{j-} \lambda_{h,i,j}^n \right) \right] \Delta \phi \Delta \theta \Delta t = \\ & \sum_{i=2}^I \sum_{j=2}^J \sum_{n=1}^N \left( -\frac{1}{a} D_{j-} \lambda_{h,i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \end{aligned} \quad (B-28)$$

where it has been used that  $V_{i,j}^n = 0$  at a solid northern or southern boundary.

Differentiation of  $L$  with respect to  $h_{i,j}^n$  gives from the third term in (B-9)

$$\begin{aligned} & \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \sum_{n=1}^N \frac{\partial}{\partial h_{i,j}^n} \left( \lambda_{u,i,j}^n \frac{c_i^2}{a \cos \theta_j} D_{i+} h_{i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t = \\ & \sum_{i=2}^I \sum_{j=1}^{J-1} \sum_{n=1}^N \left[ a \frac{\partial}{\partial h_{i,j}^n} \left( (c_I^2 \lambda_{u,I,j}^n h_{I,j}^n - c_2^2 \lambda_{u,2,j}^n h_{2,j}^n) \right. \right. \\ & \left. \left. - h_{i,j}^n D_{i-} (c_i^2 \lambda_{u,i,j}^n) \right) \right] \Delta \phi \Delta \theta \Delta t = \\ & \sum_{i=2}^I \sum_{j=1}^{J-1} \sum_{n=1}^N \left( -\frac{1}{a \cos \theta_j} D_{i-} (c_i^2 \lambda_{u,i,j}^n) \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t \end{aligned} \quad (B-29)$$

where it again has been used that  $\lambda_{u,i,j}^n$  is equal to zero at the solid boundaries at the eastern and western end of the basin.

From the eight term

$$\begin{aligned}
& \sum_{i=2}^I \sum_{j=2}^J \sum_{n=1}^N \frac{\partial}{\partial h_{i,j}^n} \left( \lambda_{v,i,j}^n \frac{c_i^2}{a} D_{j-} h_{i,j}^n \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t = \\
& \sum_{i=2}^I \sum_{j=1}^{J-1} \sum_{n=1}^N \left[ c_i^2 a \frac{\partial}{\partial h_{i,j}^n} \left( (\lambda_{v,i,J}^n \cos \theta_J h_{i,J}^n - \lambda_{v,i,2}^n \cos \theta_2 h_{i,2}^n) \right. \right. \\
& \left. \left. - h_{i,j}^n D_{j+} (\lambda_{v,i,j}^n \cos \theta_j) \right) \right] \Delta \phi \Delta \theta \Delta t = \\
& \sum_{i=2}^I \sum_{j=1}^{J-1} \sum_{n=1}^N \left( -\frac{c_i^2}{a \cos \theta_j} D_{j+} (\lambda_{v,i,j}^n \cos \theta_j) \right) a^2 \cos \theta_j \Delta \phi \Delta \theta \Delta t
\end{aligned} \tag{B-30}$$

where it has been assumed that  $\lambda_{v,i,j}^n = 0$  at a solid northern and southern boundary.

If  $\lambda_{u,i,j}^n$  and  $\lambda_{v,i,j}^n$  are rescaled by

$$\begin{aligned}
\lambda_{u,i,j}^{n'} &= \lambda_{u,i,j}^n \cdot c_i^2 \\
\lambda_{v,i,j}^{n'} &= \lambda_{v,i,j}^n \cdot c_i^2
\end{aligned} \tag{B-31}$$

the equations for the Lagrangian multipliers or the adjoint equations can be written in the following way

$$\begin{aligned}
& -\bar{D}_{n0} \lambda_{u,i,j}^n + \frac{f}{4} (\lambda_{v,i+1,j+1}^n + \lambda_{v,i,j+1}^n + \lambda_{v,i+1,j}^n + \lambda_{v,i,j}^n) \\
& -\frac{c_i^2}{a \cos \theta_j} D_{i+} \lambda_{h,i,j}^n - \frac{A}{a^2} \left( \frac{1}{\cos^2 \theta_j} (D_{i+} D_{i-}) \lambda_{u,i,j}^n \right. \\
& \left. + (D_{j+} D_{j-}) \lambda_{u,i,j}^n - \frac{\sin \theta_j}{\cos \theta_j} D_{j0} \lambda_{u,i,j}^n \right) = 0
\end{aligned} \tag{B-32}$$

$$\begin{aligned}
& -\bar{D}_{n0} \lambda_{v,i,j}^n - \frac{f}{4} (\lambda_{u,i,j}^n + \lambda_{u,i-1,j}^n + \lambda_{u,i,j-1}^n + \lambda_{u,i-1,j-1}^n) \\
& -\frac{c_i^2}{a} D_{j-} \lambda_{h,i,j}^n - \frac{A}{a^2} \left( \frac{1}{\cos^2 \theta_j} (D_{i+} D_{i-}) \lambda_{v,i,j}^n \right. \\
& \left. + (D_{j+} D_{j-}) \lambda_{v,i,j}^n - \frac{\sin \theta_j}{\cos \theta_j} (D_{j0} \lambda_{v,i,j}^n) \right) = 0
\end{aligned} \tag{B-33}$$

$$\begin{aligned}
& -\bar{D}_{n0} \lambda_{h,i,j}^n - \frac{1}{a \cos \theta_j} \left( D_{i-} (\lambda_{u,i,j}^n) + D_{j+} (\lambda_{v,i,j}^n \cos \theta_j) \right) \\
& + K_h (h_{i,j}^n - h_{i,j}^{n'}) = 0
\end{aligned} \tag{B-34}$$

where the primes have been dropped.

At last from the differentiation of  $L$  with respect to  $c_i^2$  one finds

$$\sum_{j=1}^{J-1} \sum_{n=1}^N \left[ \frac{\lambda_{u,i,j}^n}{c_i^2 a \cos \theta_j} D_{i+h_{i,j}^n} + \frac{\lambda_{v,i,j}^n}{c_i^2 a} D_{j-h_{i,j}^n} \right] \Delta \theta \Delta t \quad (B-35)$$

$$+ T L_\theta K_c (c_i^2 - c_i^{2'}) = 0$$

where again the primes have been dropped.  $T$  represents the period over which the data are assimilated, while  $L_\theta$  is a length scale in the north-south direction.