

**TIME-DEPENDENT ASSIMILATION OF CTD DATA
TO AN OPEN OCEAN ROSSBY WAVE MODEL**

by

James Reginald Gunson

B.Sc.(Hons.) The Flinders University of South Australia.
(1981)

Submitted in partial fulfillment of the
requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION


March 1993

© James R. Gunson 1993

The author hereby grants to MIT and to WHOI permission to reproduce
and to distribute copies of this thesis document in whole or in part.

Signature of Author 

Joint Program in Physical Oceanography
Massachusetts Institute of Technology
Woods Hole Oceanographic Institution
April 5, 1993

Certified by 

Paola Malanotte-Rizzoli
Professor of Physical Oceanography
Thesis Supervisor

Accepted by 

Lawrence J. Pratt
Chairman, Joint Committee for Physical Oceanography
Massachusetts Institute of Technology
Woods Hole Oceanographic Institution

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY
LINDGREN
MAY 12 1993

TIME-DEPENDENT ASSIMILATION OF CTD DATA TO AN OPEN OCEAN ROSSBY WAVE MODEL

by
James Reginald Gunson

Submitted in partial fulfillment of the requirements for the degree of
Master of Science at the Massachusetts Institute of Technology
and the Woods Hole Oceanographic Institution
April 5, 1993

Abstract

For two CTD surveys taken two months apart during the 1981 Ocean Tomography experiment southwest of Bermuda, a time-dependent ocean state is estimated that evolves from the first to the second survey. The mesoscale motions are modelled by quasi-geostrophic dynamics, and the ocean model state is represented by freely propagating Rossby waves in the 1st and 2nd baroclinic modes whose time-dependent amplitudes are fit to the data at the two survey times. First the linear dispersion relation is used to evolve the waves in time, then the weak nonlinear interactions between the waves are incorporated into the model. The wave-wave interactions enable the waves in the barotropic mode to enter into the model state evolution as an unknown control variable. Three techniques from optimal control theory are developed and applied to this problem: the Kalman filter/smoothen, the adjoint method and dynamic programming. These methods allow the barotropic component of the flow, which is indeterminate from the CTD data, to be estimated.

Thesis Supervisor: Paola Malanotte-Rizzoli,
Professor of Physical Oceanography
Department of Earth, Atmospheric and Planetary Sciences
Massachusetts Institute of Technology

Acknowledgements

I gratefully thank my thesis advisor, Paola Malanotte-Rizzoli, for her guidance and support along the way. Also, thanks to Bruce Cornuelle who provided the original problem and the data, and many useful comments. I would also like to thank Glenn Flierl who elucidated the formulation of wave-wave interactions for me, and Carl Wunsch for his suggestions on the application of control theory to oceanographic problems and for his financial support. This research was carried out under Office of Naval Research (ONR) University Research Initiative contract N00014-92-J-1571.

Contents

Abstract	3
Acknowledgements	4
Introduction	6
1 The model and the data	12
1.1 The Model	12
1.2 The Data	18
1.3 Stationary Fit	23
2 Time-dependent Fit	31
2.1 Formulation of the time-dependent control problem	32
2.2 The Kalman Filter/Smoothen	35
2.3 The Adjoint Method	38
2.4 Dynamic Programming	40
2.5 Results	42
3 Estimating the Barotropic Flow	49
3.1 Wave-Wave interactions	50
3.2 Formulation of the control	52
3.3 Results	55
Conclusions	62
References	66

Introduction

Our ideas and theories about the ocean circulation must be able to stand the test of observations. Observations of the ocean have been sparse and sporadic in space and time. It is a major problem to reconcile our various models of the ocean with these observations.

For this particular problem the observations are two CTD surveys taken 55 days apart in the western North Atlantic. A linear quasi-geostrophic model is used to evolve the flow between the two surveys.

There are diverse means of fitting dynamics to data. Here I am going to examine and compare three techniques taken from the field of optimal control theory: the Kalman filter/smoothen, the adjoint method and dynamic programming. These methods enable ocean models to evolve in time such that they produce observable fields consistent with real data, or equivalently, that allow the information contained in the data to be extrapolated in time and space using some model of the ocean physics. The mathematical background for these methods can be found in Bryson and Ho (1975) and Brogan (1982).

The application of these methods to physical oceanography is still in its infancy, mainly due to the large computing memory required. A strong motivation for applying these techniques to an oceanographic problem is that it is possible to estimate unknown variables that are not directly observable. Wunsch (1988) used the adjoint method to show that; given some observations in space and time of a tracer field in the ocean, one could estimate what the history of the surface boundary condi-

tions were, in order that the tracer concentrations obeyed a simple advection-diffusion model and produced fields consistent with the data.

Oceanographic problems are usually very underdetermined, there are far more model parameters to be estimated than there are independent pieces of information available. Simple models are used as a first attempt, so as to reduce the number of parameters. I have taken this approach here, via the quasi-geostrophic approximation, the oceanographic variables are all expressed in terms of the dynamic pressure streamfunction, which is then expanded as vertical and horizontal modes or Rossby waves. The model state is then just the time-dependent amplitudes of these waves.

Rossby waves have long been thought to be an important feature of the general circulation of the ocean and atmosphere. With the MODE experiment in the mid-seventies it was found that a lot of the kinetic energy of oceanic motions is in the mesoscale band, 50-150 days with horizontal scales of the order of hundreds of kilometres (Richman et al, 1977). Linear quasi-geostrophic dynamics can not adequately model these mesoscale motions, or eddies, as their short time and length scales make the nonlinear terms in the equations of motion significant. When the nonlinear terms dominate, the motions are in the regime of geostrophic turbulence (Charney, 1971 and Rhines, 1977). Linear dynamics apply for large enough scales. In between these two regimes, it should be possible to model mesoscale motions using Rossby waves that weakly nonlinearly interact with each other.

Attempts have been made to fit the Rossby waves predicted by quasi-geostrophic theory to observations. McWilliams and Flierl (1976) fit four barotropic and baroclinic waves to MODE data, and found that their linear wave model fit the data quite well, but also predicted significant weak nonlinear interactions that were not evident in their fit.

Gaspar and Wunsch (1989) fit barotropic Rossby waves to satellite altimetric data from the northwest Atlantic. They applied a Kalman filter and smoother to

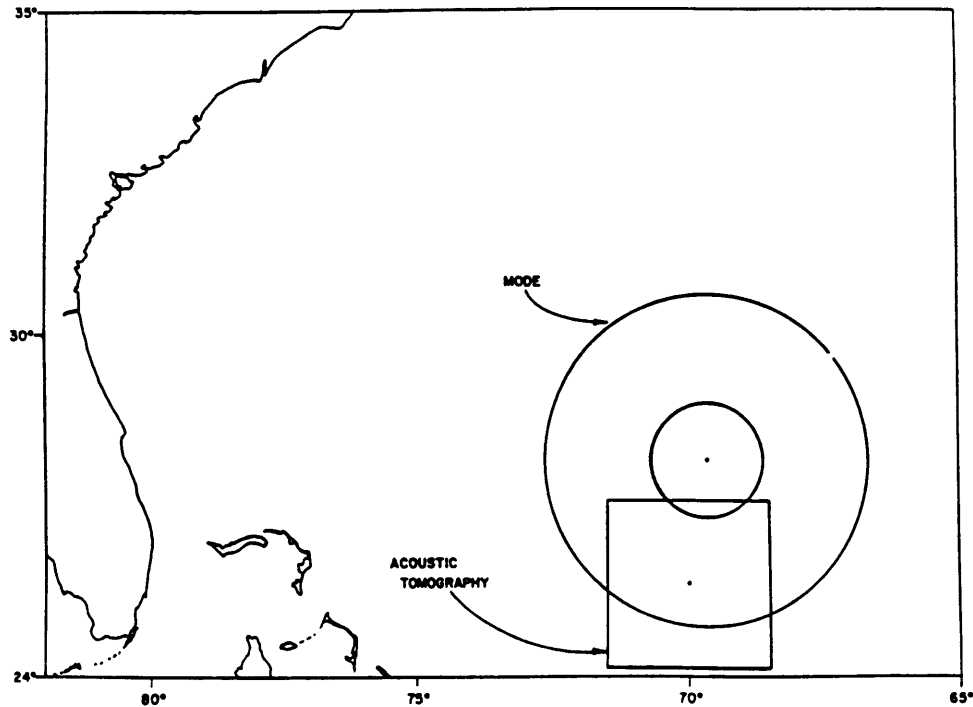


Figure 0.1: Relative locations of the 1981 Ocean Tomography Experiment and the MODE experiment in the western North Atlantic.

the time-dependent data. Fu, Vazquez and Perigaud (1991) fit equatorial planetary waves to altimetric data using the Kalman filter.

The 1981 Ocean Tomography Experiment is the source of the data I will be using here. This experiment was conducted in the MODE region (see figure 0.1) and produced a dataset consisting of acoustic tomography travel-times, temperature time series, AXBT and CTD surveys (see Ocean Tomography Group, 1982 and Cornuelle et al, 1985). Chiu and Desaubies (1987) fit three 1st mode baroclinic mode Rossby waves with a mean barotropic flow, to the time series of acoustic travel time and temperature and the first two CTD surveys, using a linear wave model. A time sequence of their estimated sound speed anomaly (which is proportional to temperature) at 700 metres depth is shown in figure 0.2. The CTD surveys were taken at the time of the first and last plots. A cold eddy is present in the middle of the domain at the first survey, then it slowly moves off to the west.

In all of these previous studies only a few long wavelength Rossby waves were used that could explain only a fraction of the variability of the data, effectively filtering out smaller scales. Also, the nonlinear interactions were not included in the fit. These fits entailed finding the wavenumbers and amplitudes of some small number of waves that fit the data and the dispersion relation for Rossby waves to some degree.

For this study I am using a model basis consisting of 130 waves in the first and second baroclinic modes so the data is fit very well. Richman, Wunsch and Hogg (1977) found that in the MODE region, most of the baroclinic energy of the mesoscale motions is in a form where the thermocline simply moves up and down, the entire water column moving together. The mesoscale variability is adequately described by the lowest vertical modes. I am here only using the first and second CTD surveys which are two months apart. The optimal control methods will be applied to estimate the time-dependent amplitudes of the waves necessary to evolve the ocean from the first to the second survey, and estimate any unknown control in the wave model.

In chapter one the quasi-geostrophic model is formulated and the dataset is described. It is shown how the model parameters are related to the observations, then the model parameters are fit to the observations using ordinary least-squares. In chapter two the time-dependent control problem is formulated and the three optimal control methods are expounded to show how to calculate an optimal state in time consistent with the model and the data. The models used are simple persistence in time and then the linear dispersion relation for Rossby waves. This is done to see if improving the model improves the time-dependent fit. A more sophisticated model is developed in chapter three that includes the weak nonlinear interactions between the waves in the model. The wave-wave interactions between the baroclinic and the barotropic waves are used to estimate the size of the barotropic waves, in terms of an unknown control that is to be estimated.

Assimilating a baroclinic wave model with CTD data, I find that it is possible to estimate a barotropic flow that is consistent with the model and the data.

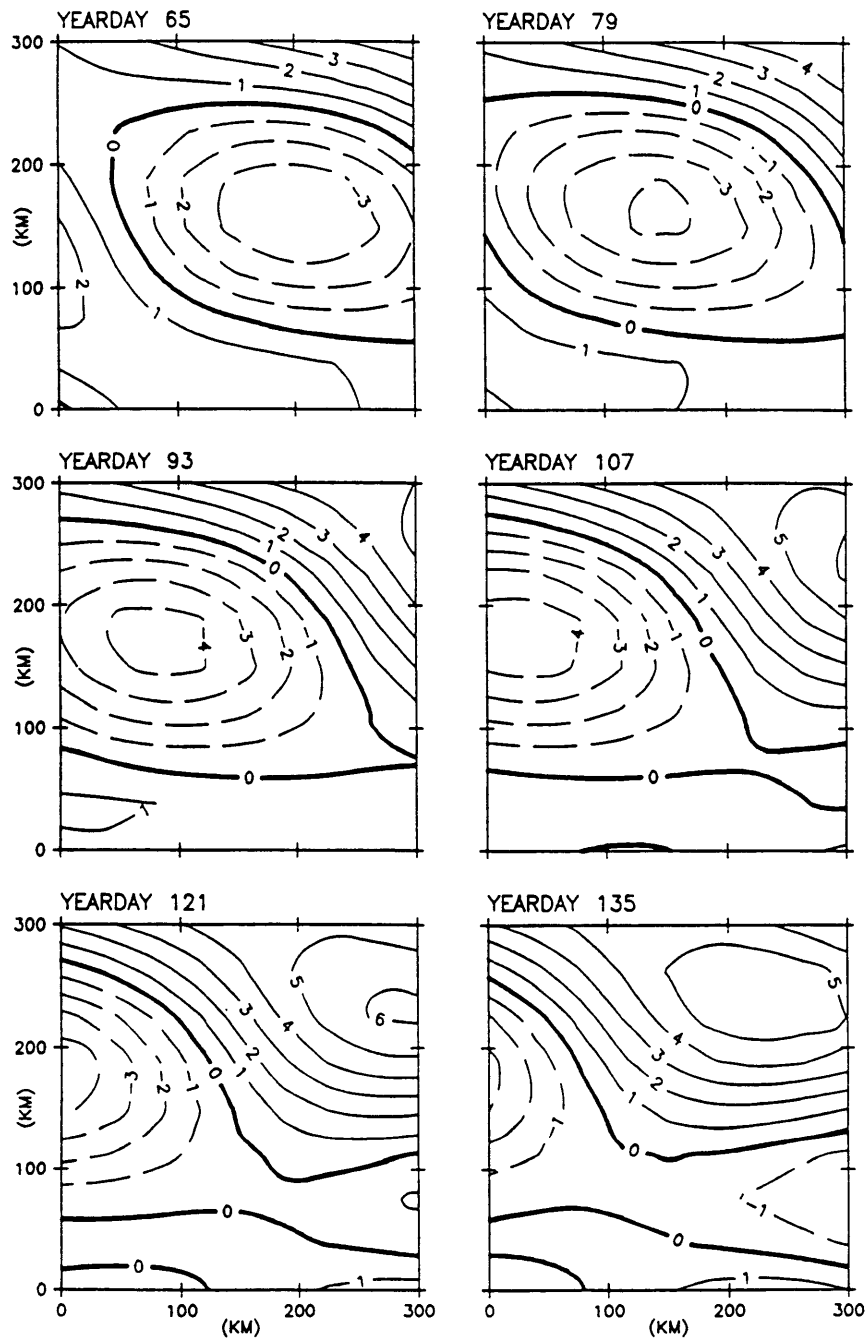


Figure 0.2: Time sequence of sound speed anomaly (in m.s^{-1}) at a depth of 700 metres from Chiu and Desaubies' estimated waves for the 1981 Ocean Tomography Experiment, this square region is the square in figure 0.1.

The determination of the barotropic flow from observations has long been a frustrating problem for oceanographers. Here is presented a new approach to solving this problem.

Chapter 1

The model and the data

1.1 The Model

Within this quiescent region of the ocean south-west of Bermuda, quasi-geostrophic dynamics are expected to model the ocean quite well. Following Pedlosky (1979) and Flierl (1978), let dynamic pressure, p' , and density, ρ' , be quasi-geostrophic perturbations to a hydrostatic basic state, ie.

$$p'(x, y, z, t) = p(x, y, z, t) - \bar{p}(z)$$

$$\rho'(x, y, z, t) = \rho(x, y, z, t) - \bar{\rho}(z)$$

and these perturbations are in hydrostatic and geostrophic balance using a beta-plane approximation for the coriolis acceleration,

$$\rho'(x, y, z, t) = \frac{-\rho_0}{g} \frac{\partial p'(x, y, z, t)}{\partial z} \tag{1.1}$$

$$f_0 u = -\frac{\partial p'}{\partial y} \tag{1.2}$$

$$f_0 v = \frac{\partial p'}{\partial x} \tag{1.3}$$

note that here p' is the dynamic pressure (\equiv pressure/density, units m^2s^{-2}). Also, for adiabatic perturbations, conservation of density is

$$\rho'_t + u\rho'_x + v\rho'_y + w\frac{d\bar{\rho}}{dz} = 0$$

ie. $w = \frac{g}{N^2\rho_0} \frac{d\rho'}{dt}$ (1.4)

Combining the above equations using the quasi-geostrophic approximations and neglecting forcing and dissipation terms, we get

$$\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right) \left[\nabla^2 + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial}{\partial z}\right)\right] p' + \beta p'_x = 0$$

or equivalently, using (1.2) and (1.3)

$$\frac{\partial}{\partial t} \left[\nabla^2 + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial}{\partial z}\right)\right] p' + \frac{1}{f_0} J \left(p, \left[\nabla^2 + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial}{\partial z}\right)\right] p\right) + \beta p'_x = 0 \quad (1.5)$$

This equation expresses the conservation of quasi-geostrophic potential vorticity for a continuously stratified ocean on a beta-plane, and governs the evolution of the dynamic pressure for mesoscale motions.

Here the mesoscale motions are not being linked to local forcing by wind stress curl or dissipation by bottom topography. The ocean bottom is flat over the region of interest and the wind stress curl acts on time scales too short (10 days) to excite baroclinic Rossby waves. Furthermore the connection between mesoscale baroclinic motions and Ekman pumping in this region is not well known, Brink (1989).

Hence the mesoscale motions are modelled here as freely propagating Rossby waves. Where these waves come from and where they get their energy is not clear. Some possible sources for these waves are: radiation into the region from somewhere energetic outside such as the Antilles Current or the Gulf Stream (Malanotte-Rizzoli et al, 1987; Hogg, 1988; Malanotte-Rizzoli et al, 1992) or they arise due to the baroclinic instability process extracting energy from the mean meridional temperature gradient as suggested by Gill et al (1974).

The surface and bottom boundary conditions for the dynamic pressure are that,

$$\begin{aligned}\frac{dp'}{dt} &= gw \text{ at } z=0 \\ w &= 0 \text{ at } z=-H\end{aligned}$$

Separation of variables by

$$p'(x, y, z, t) = \sum_m \alpha_m(x, y, t) G_m(z) \quad (1.6)$$

produces an equation for the vertical component

$$\frac{d}{dz} \left(\frac{f_0^2}{N^2} \frac{dG_m}{dz} \right) = -\lambda_m^2 G_m \quad (1.7)$$

with $G'_m(z) + \frac{N(z)^2}{g} G_m = 0$ at $z=0$ and $G'_m(z) = 0$ at $z=-H$. The solutions to this Sturm-Liouville problem are the eigenfunctions or vertical modes, $G_m(z)$, with eigenvalues, λ_m , orthonormalised by

$$\frac{1}{H} \int_{-H}^0 G_i(z) G_j(z) dz = \delta_{ij} \quad (1.8)$$

and by definition, the radius of deformation of the n th mode is $R_n = \lambda_n^{-1}$.

An associated Sturm-Liouville problem emerges from (1.7) by substituting

$$\zeta_m \equiv \frac{f_0^2}{N^2} \frac{dG_m}{dz}$$

then we get

$$\zeta_m''(z) + \frac{\lambda_m^2 N^2}{f_0^2} \zeta_m = 0$$

with $\zeta_m'(z) - \frac{g\lambda_m^2}{f_0^2} \zeta_m = 0$ at $z=0$ and $\zeta_m(z) = 0$ at $z=-H$. The orthogonality relation here is

$$\frac{1}{H} \int_{-H}^0 \frac{N(z)^2}{f_0^2} \zeta_i(z) \zeta_j(z) dz = \lambda_j^2 \delta_{ij} \quad (1.9)$$

All dynamical quantities can now be expressed in terms of the perturbation dynamic pressure using the above eigenfunction expansions. (1.1),(1.2),(1.3) and (1.4) become

$$\rho' = \sum_m \alpha_m \left(\frac{-N^2 \rho_0}{f_0^2 g} \zeta_m \right) \quad (1.10)$$

$$u = \frac{-1}{f_0} \sum_m \frac{\partial \alpha_m}{\partial y} G_m \quad (1.11)$$

$$v = \frac{1}{f_0} \sum_m \frac{\partial \alpha_m}{\partial x} G_m \quad (1.12)$$

$$w = \sum_m \frac{\partial \alpha_m}{\partial t} \left(\frac{1}{f_0^2 \rho_0} \zeta_m \right) \quad (1.13)$$

Combining (1.6) and (1.8) we can write

$$\alpha_m(x, y, t) = \frac{1}{H} \int_{-H}^0 p'(x, y, z, t) G_m(z) dz.$$

Equation (1.5) is now multiplied by $G_m(z)$ and vertically integrated using (1.6) to give

$$\left(\frac{\partial}{\partial t} [\nabla^2 - \lambda_{m_a}^2] + \beta \frac{\partial}{\partial x} \right) \alpha_{m_a} + \frac{1}{f_0} \sum_{m_b, m_c} J(\alpha_{m_b}, [\nabla^2 - \lambda_{m_c}^2] \alpha_{m_c}) \xi_{m_a m_b m_c} = 0 \quad (1.14)$$

where $\xi_{m_a m_b m_c} \equiv \frac{1}{H} \int_{-H}^0 G_{m_a}(z) G_{m_b}(z) G_{m_c}(z) dz$ is a vertical interaction term representing the size of the coupling between vertical modes m_a, m_b and m_c .

Now this is the equation governing the evolution of the horizontal components of the perturbation dynamic pressure. The Jacobian term represents the change in mode m_a potential vorticity due to the advection of mode m_c potential vorticity by mode m_b motions.

We now expand $\alpha_m(x, y, t)$ as a Fourier Series in x and y , periodic with respect to a square model domain of dimension $(L \times L)$

$$\alpha_m(x, y, t) = \Re \left\{ \sum_{k, l} \phi_{klm}(t) e^{-i(kx + ly)} \right\}. \quad (1.15)$$

where $k, l = \pm \frac{2\pi 1}{L}, \pm \frac{2\pi 2}{L} \dots$ $\phi_{klm}(t)$ is complex and is hereafter referred to as the state vector $\phi(t) \equiv \{\phi_{k, l, m}(t)\}$, the elements of this state vector being the real and imaginary components of the amplitude of the wave (k, l, λ_m) . It is this state vector which we want to estimate from the data using the dynamics. \Re denotes the real part of the complex sum.

It is useful and simple at this stage to linearise the evolution equation by discarding the Jacobian term; then upon substituting the above expansion for $\alpha_m(x, y, t)$ into (1.14) we get the linear dispersion relation for planetary waves, ie.

$$\frac{d\phi_{klm}}{dt} = \frac{-i\beta k \phi_{klm}(t)}{k^2 + l^2 + \lambda_m^{-2}} \equiv i\omega_{klm} \phi_{klm}(t). \quad (1.16)$$

However, from scaling, the relative sizes of the local and advective accelerations are, for typical values of $L = 1 \times 10^5 m$, $U = 0.1 m s^{-1}$, $f_0 = 10^{-4} s^{-1}$, $\beta = 2 \times 10^{-11} m^{-1} s^{-1}$

$$\frac{\partial}{\partial t} \sim O(\beta L) \quad u \frac{\partial}{\partial x} \sim O(UL^{-1})$$

$$\text{thus} \quad \frac{\text{advective}}{\text{local}} \sim O\left(\frac{U}{\beta L^2}\right) \sim 0.5$$

So the non-linear advective term in the evolution equation is probably significant and should not be neglected. This is also to say that the horizontal gradients of planetary and relative vorticity are comparable. Physically this means that the phase speeds of the Rossby waves are comparable to the water velocity.

Equation (1.16) is still an approximation to the full dynamics and due to this, and its simple form, is used as a model for state evolution in chapter 2. In chapter 3 we return to address the non-linear wave-wave interactions.

The model domain is a $(L \times L)=600 \times 600$ km square region centred at $26^\circ N$, $70^\circ W$. I use $k = \pm \frac{2\pi 1}{L}, \dots, \pm \frac{2\pi 5}{L}$ $l = \frac{2\pi 1}{L}, \dots, \frac{2\pi 5}{L}$ as shown in figure 1.1, and $m = 1, 2$, that is, horizontal wavelengths from 600km down to 120km in the first and second baroclinic modes. The size of the state space is reduced by only using $l > 0$. The wave basis is still complete but there are no southward propagating waves. The shortest wave has a length scale of 19 km. The radii of deformation are 48 km for the 1st mode and 19 km for the second mode. The mesoscale motions could be more accurately modelled by expanding to higher wavenumbers but the state vector becomes quite large. Here there are 2×65 unknown complex wave amplitudes to be estimated, so the state vector has dimension (260×1) , which is of a comfortable size for the assimilation.

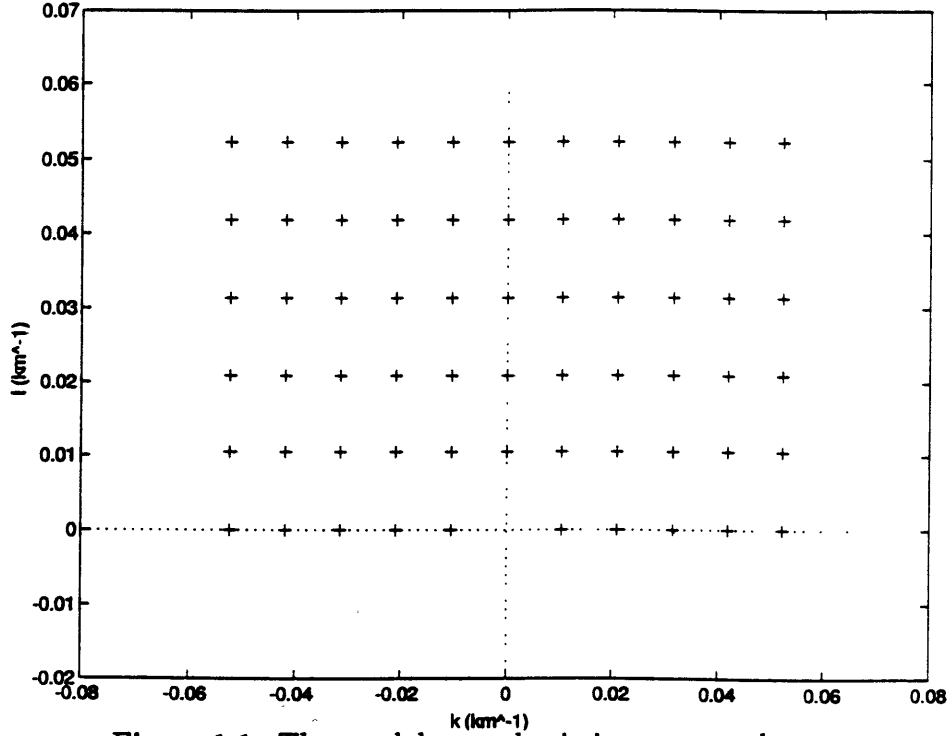


Figure 1.1: The model wave basis in wavenumber space.

The first and second baroclinic mode Rossby waves have generally low frequencies, with periods ranging from 146 days for $|k| = \frac{2\pi}{L} = 0.029 \text{ km}^{-1}$, $l = 0$ to well over 2000 days for the higher wavenumbers. Hence, time-scales shorter than about 150 days and length-scales shorter than about 20 km are not being modelled. It is known from observational studies (Richman, Wunsch and Hogg, 1974) and from modelling (Rhines, 1975) that there is strong variability at scales shorter than these. The variability at short time and length scales is dominated by strong nonlinear interactions. Here I am ignoring these scales of variability and in chapter 3 model the flow as being due to Rossby waves with weakly nonlinear interactions.

Equation (1.16) can be rewritten in matrix form as

$$\phi(t+1) = F\phi(t) \quad (1.17)$$

where the transition matrix, F , has the frequencies along the main diagonal with the time derivative expressed in Crank-Nicholson form. The time step I am using is 1 day, which gives 54 time-steps between survey A and survey B. Apart from the

neglected nonlinear interactions, another uncertainty in this formulation is due to truncating the Fourier Series expansion for the dynamic pressure. This gives rise to Gibbs phenomena; where spatial changes too short to be modelled by the waves, such as around the edges of the model domain, are picked up by the higher wavenumbers and propagated in the model domain. Hopefully this will not be a problem during the short (relative to the wave periods) time of the model run.

1.2 The Data

The data used for this time-dependent assimilation are two CTD surveys taken as part of the 1981 Tomography Demonstration Experiment (see Ocean Tomography Group, 1982). The first is denoted survey A at March 16, 1981, consisting of 65 stations, and the second is survey B at May 9, 1981, consisting of 75 stations, as shown in figure 1.2. Temperature and salinity profiles with two metre depth spacing, at each station, were made available to me from these surveys. All of the temperature and salinity profiles were smoothed to fifty metre depth spacing then the density was calculated for each profile. Horizontal mean profiles of density and buoyancy frequency were calculated from both surveys, from which were calculated the first and second baroclinic modal structure (see figure 1.3). The mean profile was subtracted from all of the density profiles to produce the density anomaly profiles for each station as shown in figure 1.4, ie.

$$\rho'(x, y, z, t) = \rho(x, y, z, t) - \bar{\rho}(z).$$

These anomalies are assumed to be caused by adiabatic perturbations to the mean density profile, these perturbations being the quasi-geostrophic planetary waves described in the previous section, ie.

$$\rho'(x, y, z, t) = \frac{-\rho_0}{g} \frac{\partial p'(x, y, z, t)}{\partial z}$$

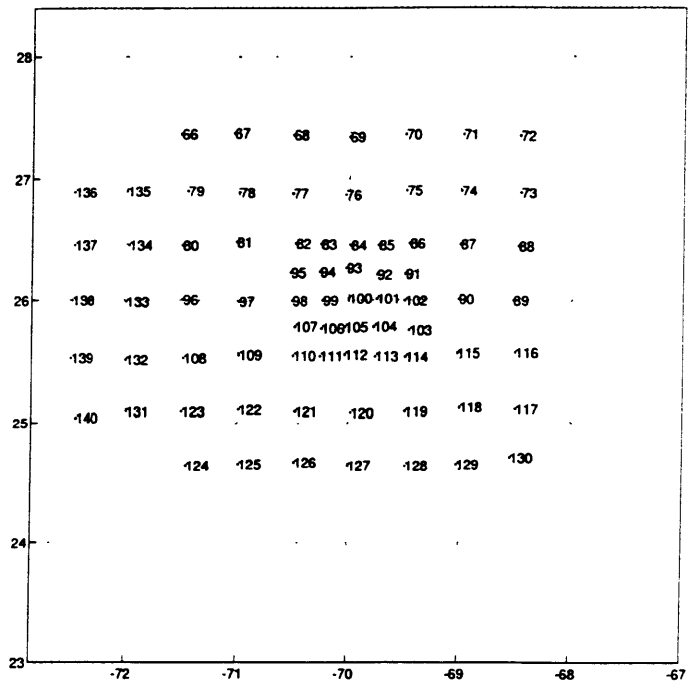
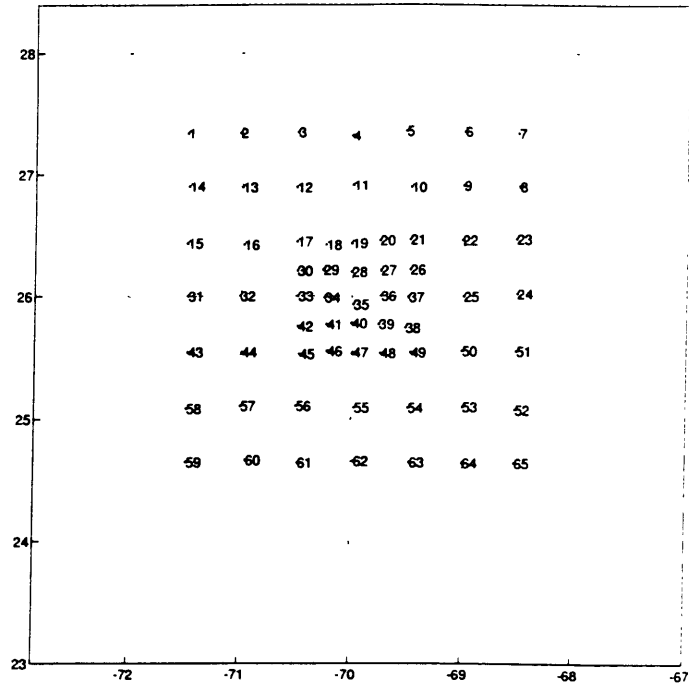


Figure 1.2: CTD station locations (top) survey A 3/16/81 (bottom) survey B 5/9/81.

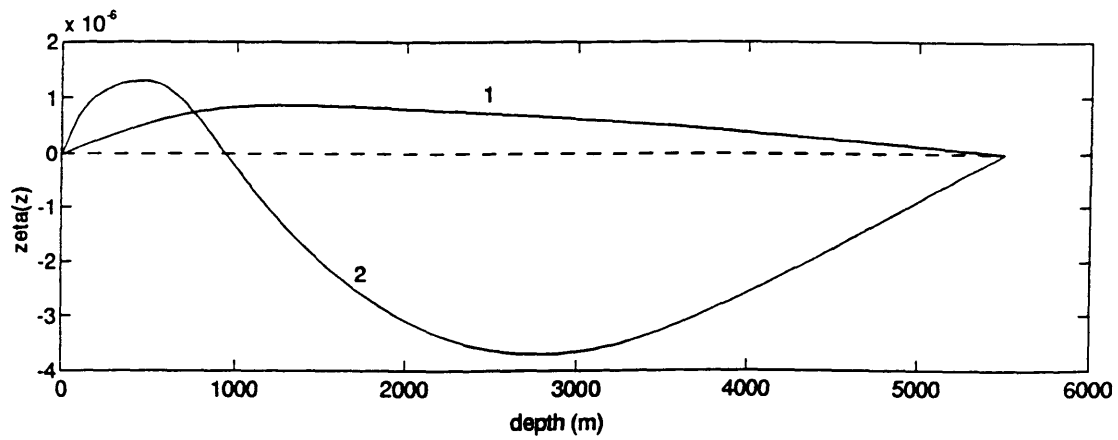
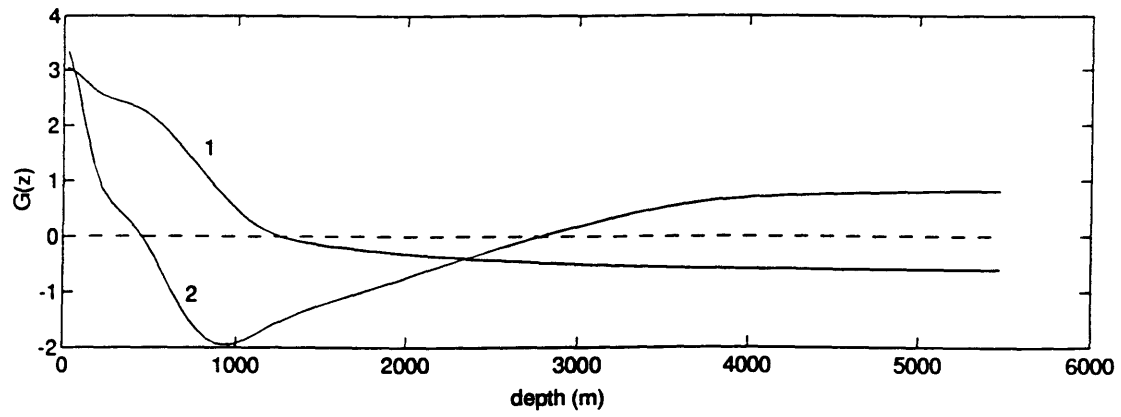
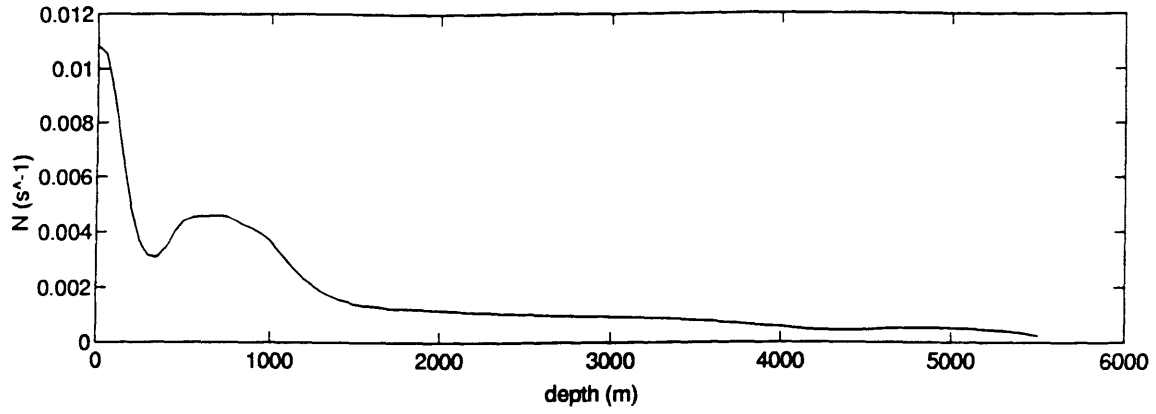


Figure 1.3: Plots with depth (metres) of (a) buoyancy frequency (b) Vertical modes, $G_1(z)$ and $G_2(z)$ (c) Vertical displacement modes, $\zeta_1(z)$ and $\zeta_2(z)$

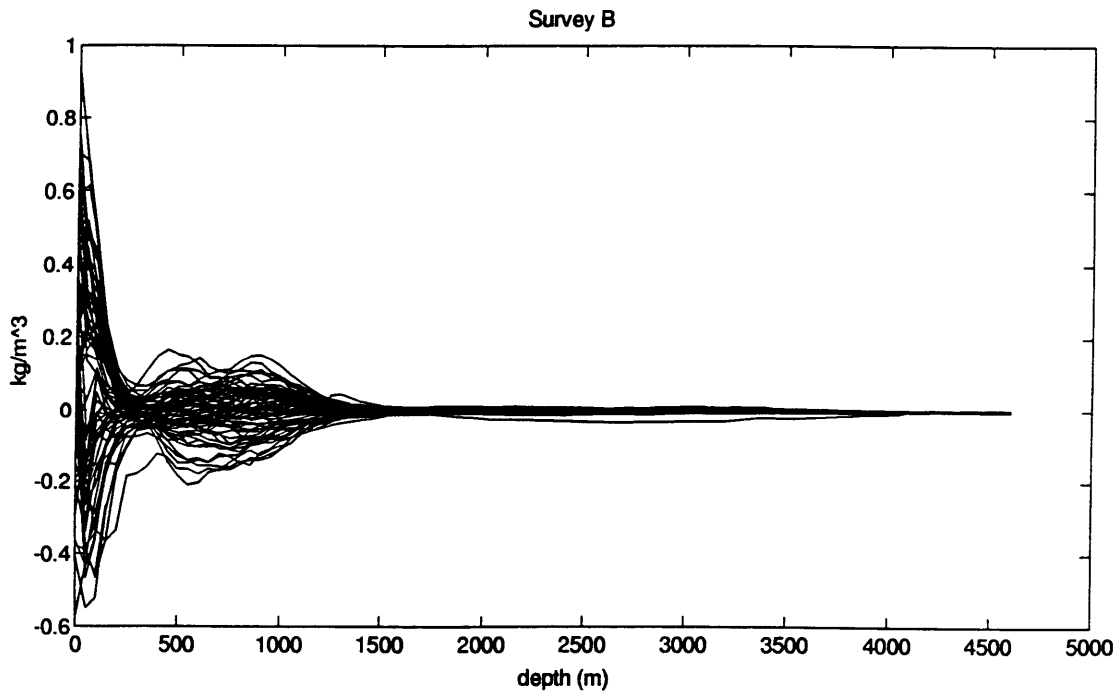
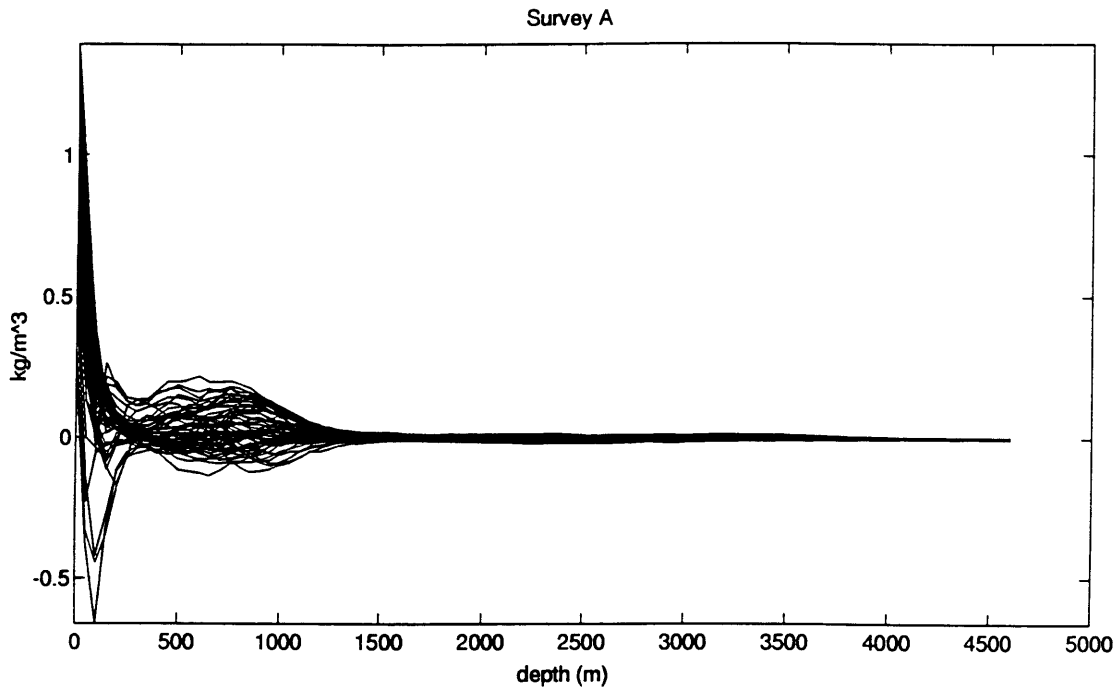


Figure 1.4: Density anomalies for all stations (top) survey A 3/16/81 (bottom) survey B 5/9/81 .

$$= \sum_{k,l,m} \phi_{klm}(t) e^{i(kx+ly)} \left(\frac{-N^2 \rho_0}{f_0^2 g} \zeta_m \right) \quad (1.18)$$

With $\rho'(x, y, z, t)$ as just described, there are about 7000 data points for each survey. To effectively reduce the number of data points whilst keeping as much information as possible, each density profile was projected onto the vertical baroclinic modes in the following way: Multiply both sides of (1.18) by $\zeta_n(z)$ and integrate in the vertical to give

$$\begin{aligned} \frac{1}{H} \int_{-H}^0 \rho'(x, y, z, t) \zeta_n(z) dz &= \sum_{k,l,m} \phi_{klm}(t) e^{i(kx+ly)} \frac{1}{H} \int_{-H}^0 \frac{-N^2 \rho_0}{f_0^2 g} \zeta_m(z) \zeta_n(z) dz \\ &= \sum_{k,l} \phi_{kln}(t) e^{i(kx+ly)} \left(\frac{-\rho_0 \lambda_n^2}{g} \right) \end{aligned}$$

where the orthogonality of the ζ_m , equation (1.9), was used. Thus

$$\sum_{k,l} \phi_{kln}(t) e^{i(kx+ly)} = d_n(x, y, t) \quad (1.19)$$

where
$$d_n(x, y, t) \equiv \frac{-g}{\rho_0 \lambda_n^2 H} \int_{-H}^0 \rho'(x, y, z, t) \zeta_n(z) dz \quad (1.20)$$

$d_n(x, y, t)$ is the modal data that the wave amplitudes are fit to. For each mode and each station there is one number giving the vertical structure instead of one number every fifty metres, reducing the number of data points for each survey to about 140 for two modes. The $d_n(x, y)$ were calculated for surveys A and B for the first and second baroclinic modes, using (1.20).

One can see from figure 1.4 that there is a large signal in the surface layer and around the main thermocline, where N is large. Quasi-geostrophic theory breaks down in the surface mixed layer since the first and second vertical displacement modes go close to zero at the surface (see figure 1.3), but the data does not. So the modes will fit the data poorly near the surface. The variability in the main thermocline, however, should be well modelled. Figure 1.5 shows the variance of the variability of the perturbation density that is captured by projecting onto the first and second modes. The modal expansion does well around the thermocline but poorly near the surface.

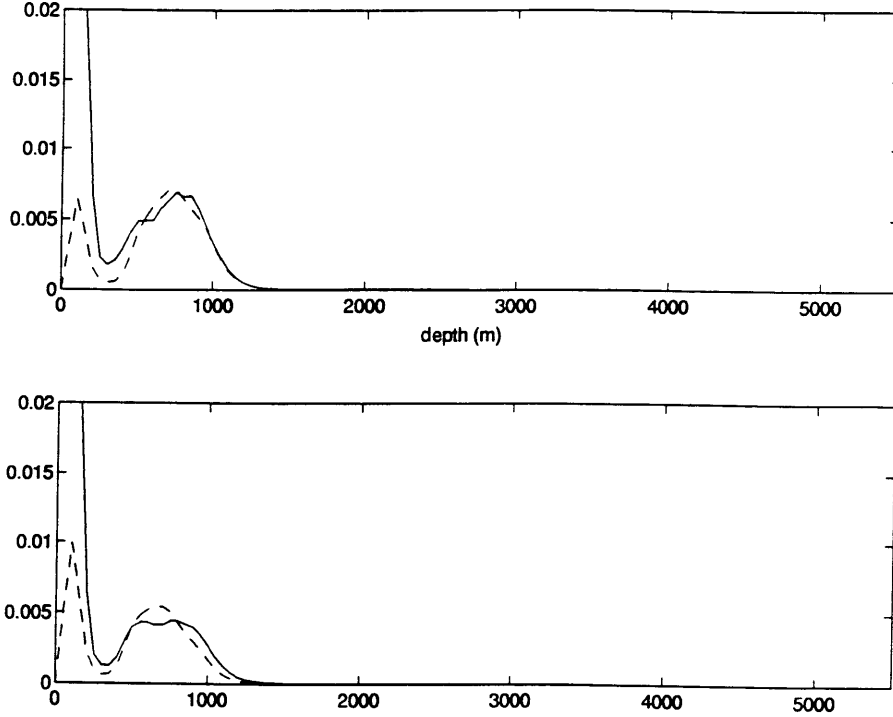


Figure 1.5: Explained variance $(\text{kg/m}^3)^2$ vs. depth: variance of the density data (solid) variance of the data expanded in the first and second modes (dashed) for survey A (top) and survey B (bottom).

1.3 Stationary Fit

Here the waves in the model basis are separately fit to the two surveys. The duration of each survey is about two weeks so the surveys are treated as being synoptic with respect to the Rossby waves. The amplitudes of the Rossby waves are estimated from the modal density data. They are related by the sampling equation (1.19) which is written in matrix/vector form as

$$\mathbf{d} = \mathbf{E}\boldsymbol{\phi} + \mathbf{n} \quad (1.21)$$

where $\mathbf{d} = \{d_n(x, y)\}$, $\boldsymbol{\phi} = \{\phi_{k,l,n}\}$ are column vectors, and the elements of \mathbf{E} are the sines and cosines in (1.19). The rows of \mathbf{E} correspond to fixed values of x, y and mode number, the columns of \mathbf{E} to fixed values of k, l and mode number. The vector \mathbf{n} is the noise vector representing uncertainty in the data and inadequacies in the model.

The best fit of ϕ to the data is found using ordinary least-squares and the Gauss-Markov theorem, that is, find the ϕ that minimises the cost,

$$(\mathbf{d} - \mathbf{E}\phi)^T \mathbf{R}^{-1}(\mathbf{d} - \mathbf{E}\phi) + (\phi - \phi_0)^T \mathbf{P}_0^{-1}(\phi - \phi_0)$$

where \mathbf{R}^{-1} and \mathbf{P}_0^{-1} are positive definite weighting matrices. ϕ_0 is the best *a priori* estimate of ϕ , here set to zero. This cost is a positive definite quadratic form and hence has a minimum with respect to ϕ that can readily be found.

\mathbf{R} and \mathbf{P}_0 are set as *a priori* estimates of the covariance matrices of the noise and the state respectively, ie.

$$\mathbf{R} \equiv \langle \mathbf{nn}^T \rangle \quad \mathbf{P}_0 \equiv \langle (\phi - \phi_0)(\phi - \phi_0)^T \rangle$$

which are to be estimated beforehand. For \mathbf{R} , measurement error due to instrumental and navigation error is small, the main source of error in this sampling equation is internal waves acting on the vertical density gradient, these waves are not modelled by quasi-geostrophic dynamics. A first estimate of \mathbf{n} is constructed by finding the $\Delta\rho'$ due to some typical internal wave vertical displacement on the buoyancy frequency as a function of depth, this typical wave displacement is taken from the Garrett-Munk spectrum for internal waves. The $\Delta\rho'(z)$ is then projected onto the modes, treating this error as being independent of horizontal location,

$$\text{ie.} \quad \Delta\rho' = \frac{d\bar{\rho}}{dz} \Delta z$$

$$\begin{aligned} \text{then} \quad \Delta d_n &= \frac{\partial d}{\partial \rho'} \Delta\rho' \\ &= \frac{-g}{\rho_0 \lambda_n^2 H} \int_{-H}^0 \Delta\rho' \zeta_n dz \end{aligned}$$

then set $\mathbf{R} = (\Delta d_n)(\Delta d_n)^T$. In figure 1.6 is plotted the *a priori* density error with depth, and the residuals (rms difference between the density data and that produced by the least-squares fit) for the fits to surveys A and B. The residuals of the fit are large near the surface as expected. Around the thermocline the fit is good but the

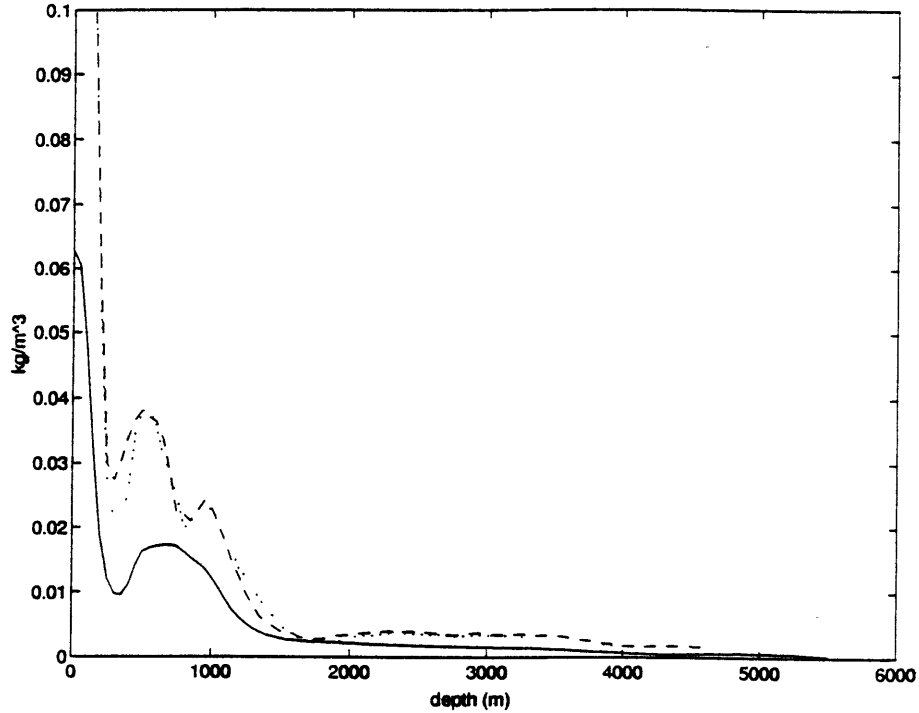


Figure 1.6: rms density residuals from least-squares fit to survey A (dashed), and survey B (dotted), and the *a priori* density error, $\Delta\rho'(z)$ (solid).

residuals are somewhat larger than the *a priori* error due to the fit being biased by the surface signal.

P_0 is constructed as a diagonal matrix representing the expected square of the amplitudes of the waves. This is set as a Gaussian with respect to total wavenumber, ie.

$$\begin{aligned} \langle (\phi_{k,l,1})^2 \rangle &\propto e^{-\frac{\kappa^2}{\gamma_1^2}} + e^{-\frac{\kappa^2}{\lambda_1^2}} \\ \langle (\phi_{k,l,2})^2 \rangle &\propto e^{-\frac{\kappa^2}{\gamma_2^2}} + e^{-\frac{\kappa^2}{\lambda_2^2}} \end{aligned}$$

where: γ_m is a wavenumber representing the spatial correlation length scale for mode m , I have set $\gamma_1 = \gamma_2 = 2\pi/100 \text{ km}^{-1}$, and $\lambda_n = R_n^{-1}$. The magnitude of P_0 , $|P_0|$ is set quite large so as to give more weight to the data than to the *a priori* guess for ϕ_0 .

There is usually difficulty in setting the *a priori* covariances in oceanographic problems. The important thing is that the estimated covariances found after the estimation be consistent with the *a priori* estimates.

The state that minimises the cost is found by setting the derivative of the cost with respect to ϕ to zero, then we get

$$\hat{\phi} = \phi_0 + \mathbf{P}_0 \mathbf{E}^T (\mathbf{E} \mathbf{P}_0 \mathbf{E}^T + \mathbf{R})^{-1} (\mathbf{d} - \mathbf{E} \phi_0)$$

with expected state error covariance, that is, expected variance of the estimated wave amplitudes about their true value, of

$$\begin{aligned} \hat{\mathbf{P}} &= \langle (\hat{\phi} - \phi)(\hat{\phi} - \phi)^T \rangle \\ &= \mathbf{P}_0 - \mathbf{P}_0 \mathbf{E}^T (\mathbf{E} \mathbf{P}_0 \mathbf{E}^T + \mathbf{R})^{-1} \mathbf{E} \mathbf{P}_0 \end{aligned}$$

These were calculated for surveys A and B. ϕ had 260 elements (65 1st and second mode waves with complex amplitudes), \mathbf{d} had 130 elements (65 stations x 2 modes) for survey A and 150 elements for survey B. There are more state variables to be estimated than there are observations available, the *a priori* statistics for the state provide extra information. Fortunately for this least-squares problem, the matrix inverses in the above expressions are not singular.

The least-squares method of fitting unknowns to observations with a model in the presence of uncertainty as described here, is at the heart of the optimal estimation methods to be used in the next chapters.

Figure 1.7 shows $\hat{\phi}$, $\hat{\mathbf{P}}$ and \mathbf{P}_0 for the fits to surveys A and B. As can be seen, the underdeterminedness of the state with respect to the data has meant that there is not enough information from the observations to determine all of the wave amplitudes so the higher wavenumbers are not significantly different from their *a priori* value of zero.

In figures 1.8 and 1.9 are shown the modal data, \mathbf{d} , the least-squares fit, $\mathbf{E} \hat{\phi}$, and the residuals $\mathbf{d} - \mathbf{E} \hat{\phi}$, plotted versus station number (see figure 1.2) for the first and second baroclinic modes. The fit is very good due to there being a relatively large number of state variables to fit the variability of the modal data. The least-squares fit is practically indistinguishable from the data, the dashed curve being coincident with the solid curve, hence the residuals (dotted curve) are small.

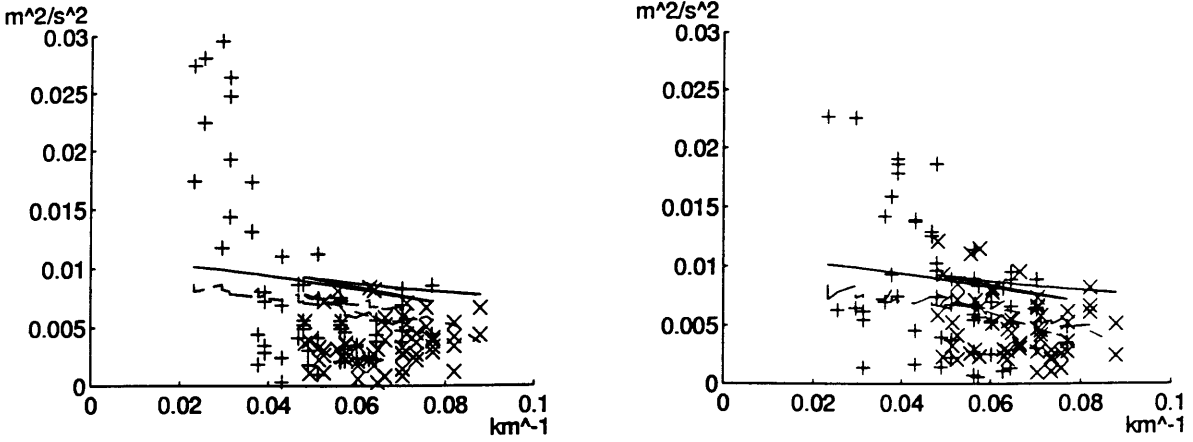


Figure 1.7: Estimated amplitudes of the 1st (+) and 2nd (x) mode waves, *a priori* standard deviation (solid) and estimated standard deviation (dashed) vs. total wavenumber for Left: survey A and Right: survey B.

The quality of the fit can be expressed by how much of the variance of the modal data is modelled by the wave fit, that is, the fractional explained variance given by

$$V = \frac{\langle \mathbf{d}^T \mathbf{d} \rangle - \langle (\mathbf{d} - \mathbf{E}\hat{\phi})^T (\mathbf{d} - \mathbf{E}\hat{\phi}) \rangle}{\langle \mathbf{d}^T \mathbf{d} \rangle}$$

For the waves fit to survey A, $V=0.9939$, and for survey B, $V=0.9885$, indicating that the waves adequately describe the data. If we estimate the modal data for those waves whose estimated amplitudes are significantly different from zero to one standard deviation, we get $V=0.9533$ for survey A and $V=0.9253$. To two standard deviations we get $V=0.8733$ for survey A and $V=0.7552$ for survey B.

In figure 1.10 are shown the estimated density anomaly for surveys A and B at 700 metres depth, mapped onto a regular grid (every 50 km) covering the model domain and contoured. The area covered by the CTD surveys is within the dotted line.

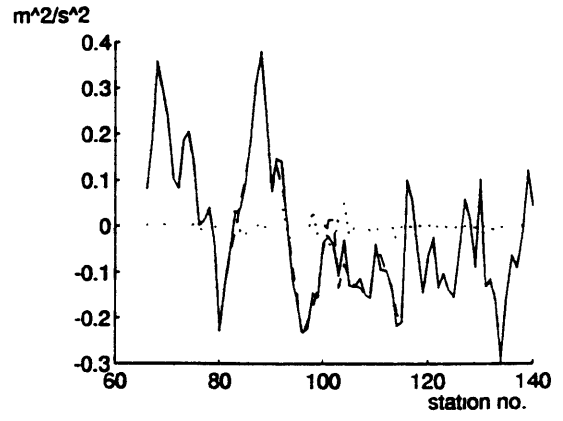
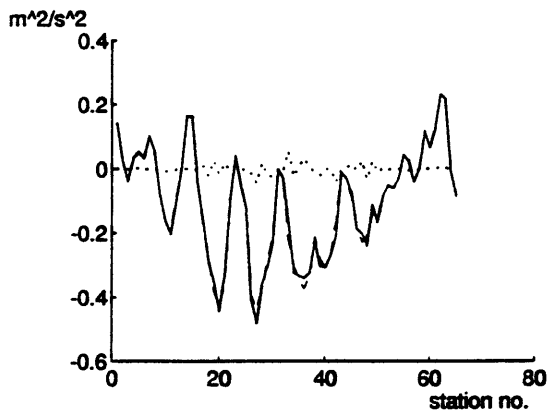


Figure 1.8: 1st modal data (solid), the least-squares fit (dashed), and the residuals (dotted) vs. station number.

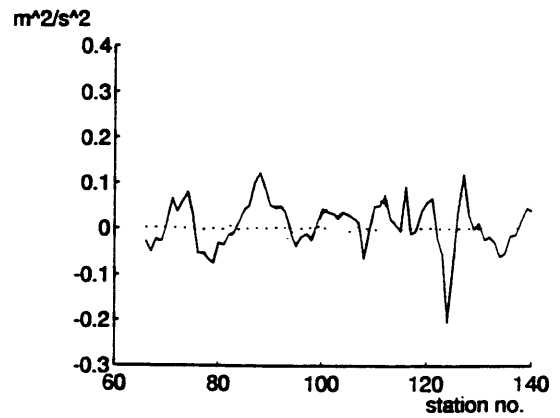
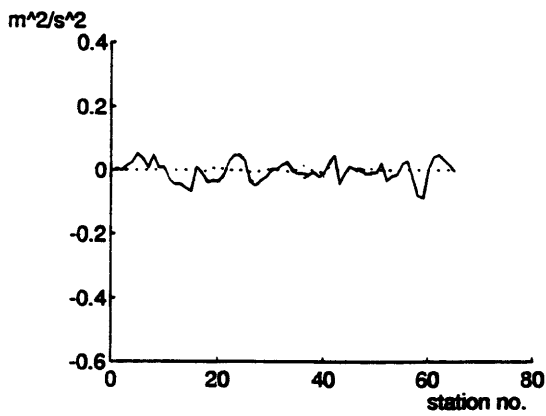


Figure 1.9: 2nd modal data (solid), the least-squares fit (dashed), and the residuals (dotted) vs. station number.

The estimated density anomaly at 700 metres depth for the two surveys can be compared to the first and last frames in figure 0.2. Here I am picking up more of the spatial variability of the data, as I have more waves with a wide range of scales available to fit to the data than had Chiu and Desaubies. The same cold (dense) eddy is seen in survey A. A cold eddy is on the western edge, and a warm eddy is in the northeast corner of the data domain, in survey B. How the flow evolves from survey A to survey B will be shown in the next chapters.

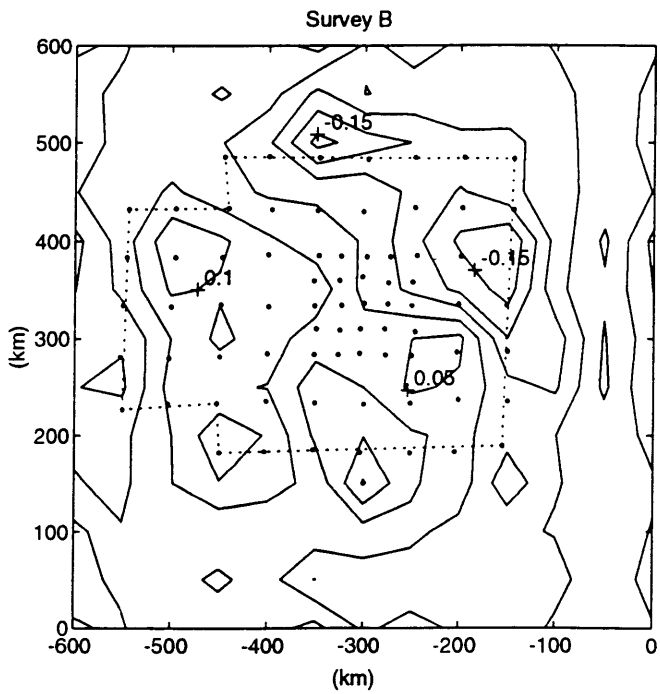
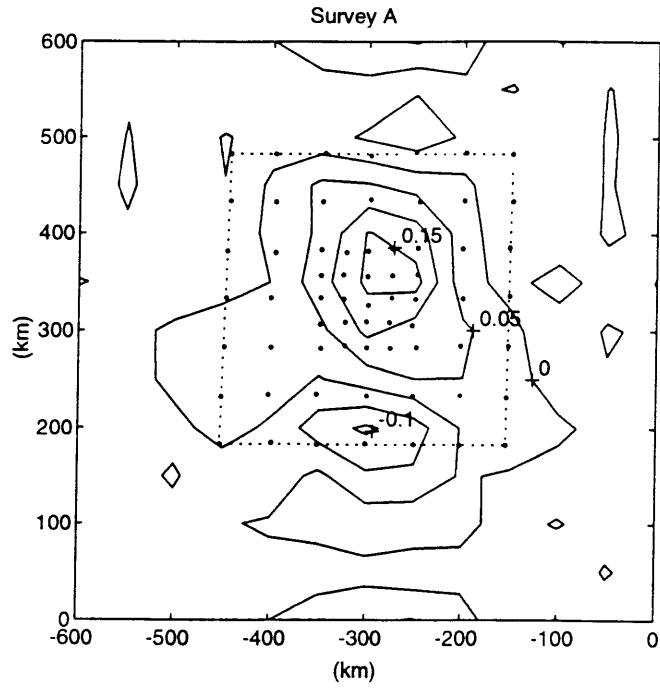


Figure 1.10: Estimated density anomaly at 700m for Top: Survey A and Bottom: Survey B, from fitting the waves separately to each survey (c.i.=0.05 kg/m³).

Chapter 2

Time-dependent Fit

Having fit the waves to surveys A and B separately, the waves from survey A are evolved forward in time using some dynamical model to see if they match the data at survey B. Three methods from optimal control theory are then formulated so as to calculate an optimal state trajectory that takes the state from the initial time to the final time such that it fits the data at both times as well as evolving in time according to the dynamical model. A hierarchy of models is used to see how including more and more dynamics improves the fit. To start with, simple persistence in time is used to evolve the state,

$$\phi_{t+1} = \phi_t + \mathbf{e}_t$$

which does not model any of the dynamics of the flow. Secondly, the linear dispersion relation for Rossby waves is used, equation (1.17),

$$\phi_{t+1} = \mathbf{F}\phi_t + \mathbf{e}_t$$

In both equations, \mathbf{e}_t is the process noise and represents the uncertainty in each state variable in predicting ϕ_{t+1} from ϕ_t due to ignored dynamical effects.

Figure 2.1 shows the time sequence of the density anomaly at 700 metres over the model domain every 11 days, from evolving the state forward in time from $\hat{\phi}_1$ using the linear model (1.17). Here $\hat{\phi}_1$ denotes the state vector of wave amplitudes

fit to survey A, the initial time in the time-dependent fit. In the 55 days between surveys A and B, the waves move very slowly toward the west, and comparing the state at $t=55$ with the stationary fit to survey B (see figure 1.7) it is apparent that the state does not evolve from $\hat{\phi}_1$ using the dispersion relation to match survey B. The fractional explained variance of the evolved final state is $V=0.0626$. The waves in the model basis cannot propagate quickly enough in the time between the surveys to render the ocean state as seen in survey A to the form as seen in survey B. In chapter three a third, more sophisticated state evolution equation is formulated to take into account the nonlinear interactions.

2.1 Formulation of the time-dependent control problem

The optimal control problem is as follows, this general form is taken from Bryson and Ho, p.395, (1975).

Minimise

$$J = \frac{1}{2}(\phi_1 - \hat{\phi}_1)^T \hat{\mathbf{P}}_1^{-1}(\phi_1 - \hat{\phi}_1) + \frac{1}{2} \sum_{t=1}^{N-1} \mathbf{u}_t^T \mathbf{Q}^{-1} \mathbf{u}_t + \frac{1}{2}(\mathbf{d}_N - \mathbf{E}\phi_N)^T \mathbf{R}^{-1}(\mathbf{d}_N - \mathbf{E}\phi_N) \quad (2.1)$$

with respect to ϕ_t , $t = 1 \dots N$ and \mathbf{u}_t , $t = 1 \dots N - 1$, subject to

$$\phi_{t+1} = \mathbf{F}\phi_t + \mathbf{B}\mathbf{u}_t \quad (2.2)$$

which is the general form of the state evolution equation.

This is a mathematical form of the statement: Find an optimal state and control trajectory that obey the model evolution equation such that the state goes from the initial state, $\hat{\phi}_1$, forward in time to match the data at the final time, \mathbf{d}_N , whilst minimising some unknown control vector along the way. The terms “optimal” and “match” mean in a least-squares sense.

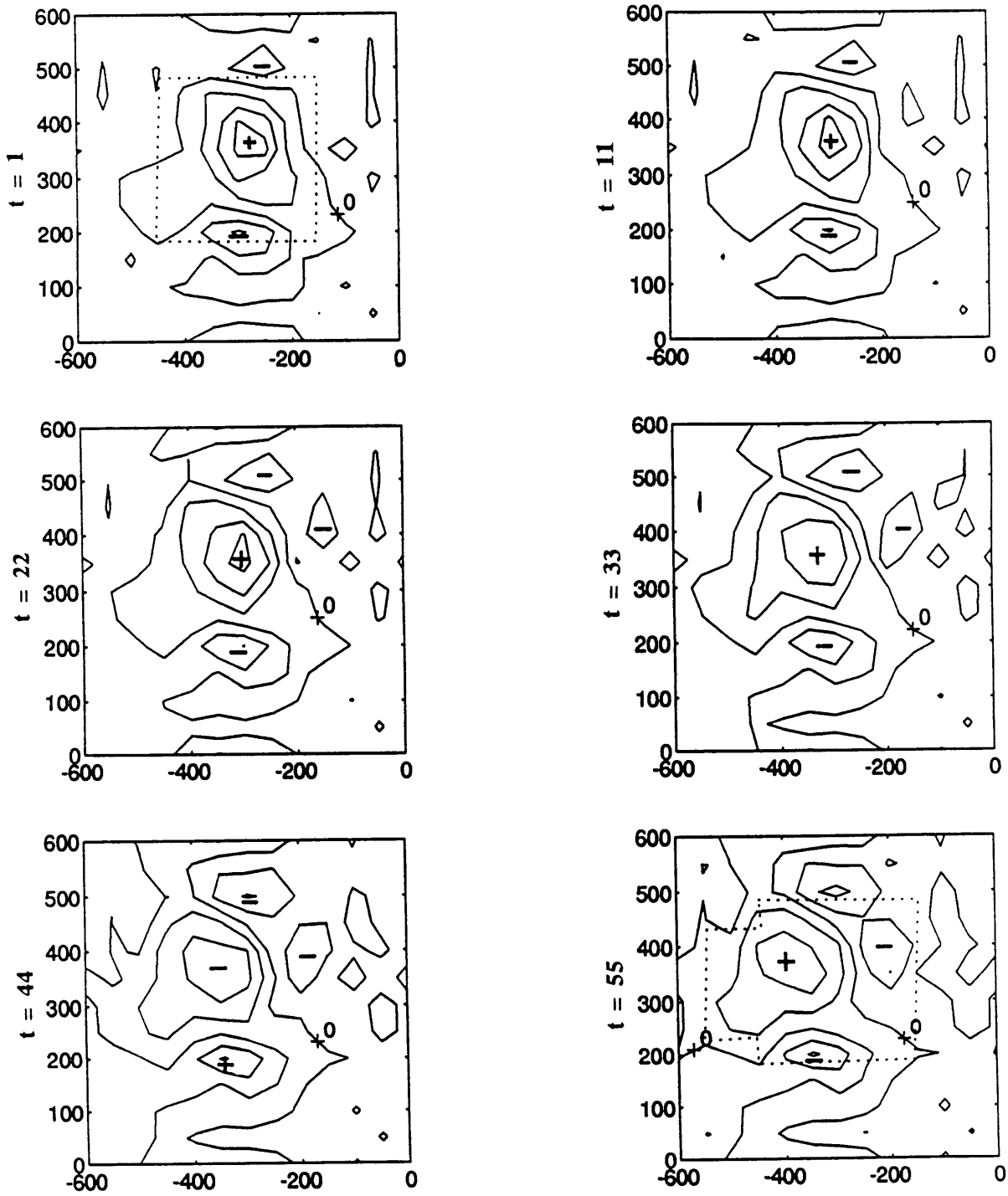


Figure 2.1: Time sequence of density anomaly at 700 metres obtained by evolving the state estimated from survey A forward in time using linear dynamics (c.i.=0.05 kg/m³).

Here \mathbf{u}_t is some unknown vector of control variables with covariance $\mathbf{Q} = \langle \mathbf{u}_t \mathbf{u}_t^T \rangle$, that is mapped onto the state vector space by \mathbf{B} . For the two models used in this chapter, the control vector to be estimated is the process noise at each time-step, assumed to be Gaussian and uncorrelated in time. If linear Rossby wave dynamics can model the oceanic flow between the surveys to some degree, there should be less estimated process noise using the linear model than that estimated using the simple persistence model. For the results in this chapter, \mathbf{u}_t is set as the process noise, \mathbf{e}_t , and \mathbf{B} set as the identity matrix.

The control problem will be formulated here using the more general and versatile form, (2.2), so that the resulting equations can be readily applied in the next chapter. As was done in the stationary fit, $\mathbf{R} = \langle \mathbf{n}_t \mathbf{n}_t^T \rangle$ is the observation noise variance at $t=N$. $\hat{\phi}_1, \hat{\mathbf{P}}_1$ are the best estimates of ϕ_1 and its covariance from the data at $t=1$, that is, the least squares estimates calculated from the stationary fit.

For both the Smoother and the Adjoint method, we apply the so-called sweep method (see Bryson and Ho, chapter 5, 1975). Lagrange multipliers are used in the following way. Define an augmented cost function

$$J' \equiv J + \sum_{t=1}^{N-1} \lambda_{t+1}^T [\mathbf{F}\phi_t + \mathbf{B}\mathbf{u}_t - \phi_{t+1}]$$

thus when the model constraint is satisfied, $J' = J$, and λ_t is arbitrary. Also define

$$H_t \equiv \frac{1}{2} \mathbf{u}_t^T \mathbf{Q}^{-1} \mathbf{u}_t + \lambda_{t+1}^T [\mathbf{F}\phi_t + \mathbf{B}\mathbf{u}_t]$$

Then we get

$$J' = \frac{1}{2} (\phi_1 - \hat{\phi}_1)^T \hat{\mathbf{P}}_1^{-1} (\phi_1 - \hat{\phi}_1) + \frac{1}{2} (\mathbf{d}_N - \mathbf{E}\phi_N)^T \mathbf{R}^{-1} (\mathbf{d}_N - \mathbf{E}\phi_N) + \sum_{t=1}^{N-1} [H_t - \lambda_{t+1}^T \phi_{t+1}]$$

The optimal state and control are given by the minimum of this positive-definite quadratic form. Setting $dJ' = 0$, J' is minimised with respect to ϕ_t and \mathbf{u}_t for $t=1, \dots, N$

$$dJ' = d\phi_1^T \frac{\partial J'}{\partial \phi_1} + d\phi_N^T \frac{\partial J'}{\partial \phi_N} + \sum_{t=1}^{N-1} \left[d\phi_t^T \frac{\partial J'}{\partial \phi_t} + d\mathbf{u}_t^T \frac{\partial J'}{\partial \mathbf{u}_t} + d\lambda_{t+1}^T \frac{\partial J'}{\partial \lambda_{t+1}} \right] = 0$$

the λ are now defined by:

$$\begin{aligned}\lambda_t &\equiv \frac{\partial H_t}{\partial \phi_t} = \mathbf{F}^T \lambda_{t+1} &\Rightarrow \frac{\partial J'}{\partial \phi_t} &= 0 \\ \lambda_N &\equiv -\mathbf{E}^T \mathbf{R}^{-1}(\mathbf{d}_N - \mathbf{E}\phi_N) &\Rightarrow \frac{\partial J'}{\partial \phi_N} &= 0 \\ \mathbf{F}^T \lambda_2 &\equiv -\hat{\mathbf{P}}_1^{-1}(\phi_1 - \hat{\phi}_1) &\Rightarrow \frac{\partial J'}{\partial \phi_1} &= 0\end{aligned}$$

therefore

$$dJ' = \sum_{t=1}^{N-1} d\mathbf{u}_t^T \frac{\partial J'}{\partial \mathbf{u}_t} = 0 \Rightarrow \mathbf{u}_t = -\mathbf{Q}\mathbf{B}^T \lambda_{t+1}$$

In this way the minimisation problem reduces to the two-point boundary value problem:

$$\phi_{t+1} = \mathbf{F}\phi_t - \mathbf{B}\mathbf{Q}\mathbf{B}^T \lambda_{t+1} \quad , \quad \phi_1 = \hat{\phi}_1 - \hat{\mathbf{P}}_1 \mathbf{F}^T \lambda_2 \quad (2.3)$$

$$\lambda_t = \mathbf{F}^T \lambda_{t+1} \quad , \quad \lambda_N = -\mathbf{E}^T \mathbf{R}^{-1}(\mathbf{d}_N - \mathbf{E}\phi_N) \quad (2.4)$$

These two equations can be readily solved by numerically integrating them from their boundary conditions on the right-hand sides. The Kalman filter/smoothen and the Adjoint method do this in different ways. Dynamic Programming starts from (2.1) and (2.2) but instead uses a cost function based on (2.1) that expresses the cost to go from state ϕ_t to the final state, and finds the control sequence from t to N that minimises that cost. Since all these methods minimise the same cost function constrained by the same state evolution equation, they should all produce the same optimal estimates of the state and the control.

2.2 The Kalman Filter/Smoothen

The Kalman filter is the most well-known method of assimilating model variables with observations in time, and there is much literature on its application (see, e.g.: Wunsch, 1989; Sorenson, 1985; Ghil and Malanotte-Rizzoli, 1991).

First the Kalman filter is run forward in time. The data available at $t=1$ is used to make estimates of $\hat{\phi}_1, \hat{P}_1$ as in section 1.3. From these estimates the Kalman filter is run forward by: forecast using the model

$$\tilde{\phi}_{t+1} = F\hat{\phi}_t \quad (2.5)$$

$$\tilde{P}_{t+1} = F\hat{P}_tF^T + BQB^T \quad (2.6)$$

then whenever data becomes available,

$$\hat{\phi}_{t+1} = \tilde{\phi}_t + K_t(d_t - E_t\tilde{\phi}_t)\hat{P}_t = \tilde{P}_t - K_tE_t\tilde{P}_t$$

where

$$K_t = \tilde{P}_tE_t^T(E_t\tilde{P}_tE_t^T + R)^{-1}$$

For this problem, however, there is data only at $t = 1$ and $t = N$ so $\hat{\phi}_t = \tilde{\phi}_t, \hat{P}_t = \tilde{P}_t$ for all other times.

For the smoother, estimates of the state are made by using the data at $t=N$ to improve the filtered estimates at $t < N$ which only used the information from the data at $t = 1$ carried forward by the model. Also, estimates can be made of the control vector. Denoting these smoothed estimates by ϕ_t^+, u_t^+ , explicit expressions for these estimates can be derived from (2.3) and (2.4), see Bryson and Ho, chapter 13, (1975) for details.

Using the forecasts, (2.5) and (2.6), equation (2.3) becomes

$$\phi_t^+ = \tilde{\phi}_t - \tilde{P}_t\lambda_t \quad (2.7)$$

and also

$$u_t^+ = -QB^T\lambda_{t+1} \quad (2.8)$$

At the final time, $t=N$, the filtered state, ϕ_N , is the best estimate of ϕ_N from all the data, so

$$\phi_N^+ = \hat{\phi}_N$$

$$\mathbf{P}_N^+ = \hat{\mathbf{P}}_N$$

The smoothing is then done by integrating (2.4) backwards from ϕ_N^+ to produce the sequence of λ_t ,

$$\begin{aligned}\lambda_N &= -\mathbf{E}^T \mathbf{R}^{-1} (\mathbf{d}_N - \mathbf{E} \phi_N^+) \\ \lambda_t &= \mathbf{F}^T \lambda_{t+1}\end{aligned}$$

The smoothed state and control sequences are then given by (2.7) and (2.8). The attractive feature of the filter/smoothener is that the covariance matrices of the state and control are calculated at each time. With the above sweep method, the covariance of λ_t , Λ_t , is given by

$$\begin{aligned}\Lambda_N &= \mathbf{E}^T (\mathbf{E} \tilde{\mathbf{P}}_N \mathbf{E}^T + \mathbf{R})^{-1} \mathbf{E} \\ \Lambda_t &= \mathbf{F}^T \Lambda_{t+1} \mathbf{F}\end{aligned}$$

Then the covariances of the smoothed state and control vectors are readily calculated,

$$\begin{aligned}\mathbf{P}_t^+ &= \hat{\mathbf{P}}_t - \hat{\mathbf{P}}_t \mathbf{F}^T \Lambda_t \mathbf{F} \hat{\mathbf{P}}_t \\ \mathbf{Q}_t^+ &= \mathbf{Q}_t - \mathbf{Q}_t \mathbf{B}^T \Lambda_t \mathbf{B} \mathbf{Q}_t\end{aligned}$$

The method so far described necessitates the storage of $\tilde{\mathbf{P}}$ at each time-step, as the Kalman filter is run forward in time.

As with most oceanographic problems these covariance matrices are quite large, (260x260) for this model basis, and there might not be enough computer memory to store them at each time step. An alternative algorithm that requires less memory but involves more computations is the so-called RTS algorithm, after Rauch, Tung and Streibel (1965). They developed the following expressions using maximum likelihood estimates. These expressions can also be derived algebraically from the sweep method expressions described above (see Bryson and Ho, sections 13.1, 13.2, 1975).

Here the smoothed state and control vectors and their covariance matrices are calculated recursively without the λ_t , Λ_t , and without storing the filtered covariances.

$$\phi_t^+ = \hat{\phi}_t + \mathbf{L}_t (\phi_{t+1}^+ - \tilde{\phi}_{t+1})$$

$$\begin{aligned}
\mathbf{u}_t^+ &= \mathbf{M}_t(\phi_{t+1}^+ - \tilde{\phi}_{t+1}) \\
\mathbf{P}_t^+ &= \hat{\mathbf{P}}_t + (\mathbf{P}_{t+1}^+ - \tilde{\mathbf{P}}_t)\tilde{\mathbf{P}}_t^{-1}\mathbf{F}\hat{\mathbf{P}}_t \\
\mathbf{Q}_t^+ &= \mathbf{Q}_t + \mathbf{Q}_t\mathbf{F}^T\tilde{\mathbf{P}}_t^{-1}(\mathbf{P}_{t+1}^+ - \tilde{\mathbf{P}}_t)\tilde{\mathbf{P}}_t^{-1}\mathbf{F}\mathbf{Q}_t
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{L}_t &\equiv \hat{\mathbf{P}}_t\mathbf{F}^T\tilde{\mathbf{P}}_t^{-1} \\
\mathbf{M}_t &\equiv \mathbf{Q}_t\mathbf{B}^T\tilde{\mathbf{P}}_t^{-1}
\end{aligned}$$

$\hat{\mathbf{P}}_t$ is calculated at each time step by inverting equation (2.6).

To summarise: The state is run forward from the initial estimate at $t = 1$, making forecasts at each time-step. When data is encountered at $t = N$, a best estimate of the state at $t = N$ is made by taking a weighted average of the forecast and the least-squares fit to the data. The misfit between this best estimate and the data, is evolved back in time, and is used to make best estimates of the states at $t < N$ using all the available data. The misfit at the final time is also used to estimate an optimal control sequence \mathbf{u}_t^+ , that drives the smoothed state from $t = 1$ to $t = N$ so that the state evolution equation is satisfied.

2.3 The Adjoint Method

This method as formulated here was that used by Wunsch (1988) to estimate boundary conditions as optimal control in a model of tracer flow in a box model of the ocean. More theory and applications of the adjoint method, otherwise known as Pontryagin's Minimum Principle, can be found in Bryson and Ho (1975) and Brogan (1982).

With this method equations (2.3) and (2.4) are manipulated to produce explicit expressions for the optimal estimates of ϕ_t^+ and \mathbf{u}_t^+ , expressing these in terms of the so-called adjoint state, λ_t . Start by rewriting (2.4) as

$$\lambda_t = -\mathbf{F}^{(N-t)T}\mathbf{E}^T\mathbf{R}^{-1}(\mathbf{d}_N - \mathbf{E}\phi_N) \quad (2.9)$$

then the boundary condition for (2.3) becomes

$$\phi_1^+ = \hat{\phi}_1 - \hat{\mathbf{P}}_1 \mathbf{F}^{(N-1)T} \mathbf{E}^T \mathbf{R}^{-1} (\mathbf{d}_N - \mathbf{E} \phi_N)$$

where, as before, ϕ_1^+ and ϕ_N^+ are the optimal estimates of ϕ_1 and ϕ_N . Now (2.3) is integrated forward from ϕ_1^+ , using the above two expressions to yield an expression for ϕ_N^+ in terms of ϕ_1^+ ,

$$\phi_N^+ = \mathbf{F}^{(N-1)T} \phi_1^+ + \mathbf{H}_1 \mathbf{E}^T \mathbf{R}^{-1} (\mathbf{d}_N - \mathbf{E} \phi_N^+)$$

where

$$\mathbf{H}_1 \equiv \sum_{i=0}^{N-i-1} \mathbf{F}^i \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{F}^{iT}$$

ie.

$$\phi_N^+ = (1 + \mathbf{H}_1 \mathbf{E}^T \mathbf{R}^{-1} \mathbf{E})^{-1} [\mathbf{F}^{(N-1)T} \phi_1^+ + \mathbf{H}_1 \mathbf{E}^T \mathbf{R}^{-1} \mathbf{d}] \quad (2.10)$$

Substituting this into the above expression for ϕ_1^+ we get, after some manipulation

$$\phi_1^+ = (1 + \hat{\mathbf{P}}_1 \mathbf{F}^{(N-1)T} \mathbf{V} \mathbf{W}^{-1} \mathbf{F}^{(N-1)})^{-1} [\hat{\phi}_1 + \hat{\mathbf{P}}_1 \mathbf{F}^{(N-1)T} (1 - \mathbf{V} \mathbf{W}^{-1} \mathbf{H}_1) \mathbf{E}^T \mathbf{R}^{-1} \mathbf{d}] \quad (2.11)$$

$$\text{where} \quad \mathbf{W} = 1 + \mathbf{H}_1 \mathbf{V}, \quad \mathbf{V} = \mathbf{E}^T \mathbf{R}^{-1} \mathbf{E}$$

Then, as in the Kalman Smoother, the λ_t are readily calculated using (2.4)

$$\lambda_N = -\mathbf{E}^T \mathbf{R}^{-1} (\mathbf{d}_N - \mathbf{E} \phi_N^+)$$

$$\lambda_t = \mathbf{F}^T \lambda_{t+1}$$

and \mathbf{u}_t^+ is calculated using (2.3). Finally ϕ_1^+ is evolved forward using the optimal control sequence to produce the optimal state estimates.

Hence the best state estimate at $t=1$ is calculated as a weighted sum of the initial estimate and the final data (mapped back N time-steps under an adjoint transformation), which are both known. The best estimate at $t=N$ is then calculated as a weighted sum of the best estimate from $t=1$ (evolved forward to $t=N$) and the final

data. As in the Kalman Smoother; the λ_t represent the misfit between the optimal estimate and the data at $t=N$ evolved back in time with the adjoint model equation (2.4), and the optimal control sequence and state estimates are readily calculated from the λ_t . This method is faster and clearly requires much less memory than the smoother as the covariance matrices are not calculated, but there is no measures of the uncertainty in the optimal estimates.

Wunsch (1988) and Tziperman and Thacker (1989) applied the adjoint method to oceanographic problems. The latter showed a way of calculating error covariances of the optimal estimates using the Hessian matrix of the cost function (see also Thacker, 1989).

2.4 Dynamic Programming

The idea here is to dissect the cost function, J , equation (2.1), and construct an "optimal return function" that expresses the cost to go from some state ϕ_t , to the final state ϕ_N in terms of finding the optimal control sequence that minimises J , while satisfying the model (2.2). For more details and examples of this technique see Brogan (1982) and Bryson and Ho (1975). Starting at the final time, define

$$g(\phi_N) \equiv \frac{1}{2}(\mathbf{d}_N - \mathbf{E}\phi_N)^T \mathbf{R}^{-1}(\mathbf{d}_N - \mathbf{E}\phi_N)$$

then

$$\begin{aligned} g(\phi_{N-1}) &\equiv \underset{\mathbf{u}_{N-1}}{Min} \left[\frac{1}{2} \mathbf{u}_{N-1}^T \mathbf{Q}^{-1} \mathbf{u}_{N-1} + \frac{1}{2} (\mathbf{d}_N - \mathbf{E}\phi_N)^T \mathbf{R}^{-1} (\mathbf{d}_N - \mathbf{E}\phi_N) \right] \\ &= \underset{\mathbf{u}_{N-1}}{Min} \left[\frac{1}{2} \mathbf{u}_{N-1}^T \mathbf{Q}^{-1} \mathbf{u}_{N-1} + g(\phi_N) \right] \end{aligned} \quad (2.12)$$

then at any time t ,

$$g(\phi_t) \equiv \underset{\mathbf{u}_t}{Min} \left[\frac{1}{2} \mathbf{u}_t^T \mathbf{Q}^{-1} \mathbf{u}_t + g(\phi_{t+1}) \right]$$

This is the optimal return function, which is saying that for any state at time t , the optimal control sequence, from time t to time N , must also be optimal for any state resulting from that at time t (optimal in the sense of minimising J). At the initial time, $t=1$, we have

$$J = \frac{1}{2}(\phi_1 - \hat{\phi}_1)^T \hat{\mathbf{P}}_1^{-1}(\phi_1 - \hat{\phi}_1) + g(\phi_1) \quad (2.13)$$

The above expressions are used in a backward then a forward sweep to get the ϕ_t^+ and the \mathbf{u}_t^+ . Starting at $t=N-1$, we have from (2.12) and (2.1)

$$g(\phi_{N-1}) \equiv \underset{\mathbf{u}_t}{Min} \left[\frac{1}{2} \mathbf{u}_{N-1}^T \mathbf{Q}^{-1} \mathbf{u}_{N-1} + \frac{1}{2} (\mathbf{d}_N - \mathbf{E}\mathbf{F}\phi_{N-1} - \mathbf{E}\mathbf{B}\mathbf{u}_{N-1})^{-1} \mathbf{R}^{-1} (\mathbf{d}_N - \mathbf{E}\mathbf{F}\phi_{N-1} - \mathbf{E}\mathbf{B}\mathbf{u}_{N-1}) \right]$$

The term in the brackets is a positive-definite quadratic form in \mathbf{u}_{N-1} , and thus has a minimum for

$$\mathbf{u}_{N-1}^+ = -[\mathbf{Q}^{-1} + \mathbf{B}^T \mathbf{E}^T \mathbf{R}^{-1} \mathbf{E}\mathbf{B}]^{-1} \mathbf{B}^T \mathbf{E}^T \mathbf{R}^{-1} (\mathbf{d}_N - \mathbf{E}\mathbf{F}\phi_{N-1})$$

and thus,

$$g(\phi_{N-1}) = \frac{1}{2} (\mathbf{d}_N - \mathbf{E}\mathbf{F}\phi_{N-1})^T [\mathbf{R} + \mathbf{E}\mathbf{B}\mathbf{Q}\mathbf{B}^T \mathbf{E}^T]^{-1} (\mathbf{d}_N - \mathbf{E}\mathbf{F}\phi_{N-1})$$

Continuing back in this way we get the optimal control for time t , in terms of ϕ_t ,

$$\mathbf{u}_t^+ = \mathbf{S}\mathbf{B}^T \mathbf{F}^{(N-t-1)T} \mathbf{E}^T \mathbf{G}_{t+1}^{-1} (\mathbf{d}_N - \mathbf{E}\mathbf{F}^{N-t} \phi_t) \quad (2.14)$$

where $\mathbf{S}_t = -[\mathbf{Q}^{-1} + \mathbf{B}^T \mathbf{F}^{(N-t-1)T} \mathbf{E}^T \mathbf{G}_{t+1}^{-1} \mathbf{E}\mathbf{F}^{N-t-1} \mathbf{B}]^{-1}$. The cost to go from any state ϕ_t to a final state at $t=N$, which is arbitrarily close to \mathbf{d}_N in a least-squares sense, is

$$g(\phi_t) = \frac{1}{2} (\mathbf{d}_N - \mathbf{E}\mathbf{F}^{N-t} \phi_t)^T \mathbf{G}_t^{-1} (\mathbf{d}_N - \mathbf{E}\mathbf{F}^{N-t} \phi_t)$$

where $\mathbf{G}_t = \mathbf{G}_{t+1} + \mathbf{E}\mathbf{F}^{N-t-1} \mathbf{B}\mathbf{Q}\mathbf{B}^T \mathbf{F}^{(N-t-1)T} \mathbf{E}^T$ is calculated recursively starting from $\mathbf{G}_N \equiv \mathbf{R}$.

Now (2.13) can be used to get the optimal state at the initial time, ϕ_1^+ ,

$$J = \frac{1}{2}(\phi_1 - \hat{\phi}_1)^T \hat{\mathbf{P}}_1^{-1}(\phi_1 - \hat{\phi}_1) + \frac{1}{2}(\mathbf{d}_N - \mathbf{E}\mathbf{F}^{N-1}\phi_1)^T \mathbf{G}_1^{-1}(\mathbf{d}_N - \mathbf{E}\mathbf{F}^{N-1}\phi_1)$$

Setting $\frac{\partial J}{\partial \phi_1} = 0$ gives

$$\phi_1^+ = (\hat{\mathbf{P}}_1^{-1} + \mathbf{F}^{(N-1)T} \mathbf{E}^T \mathbf{G}_1^{-1} \mathbf{E} \mathbf{F}^{N-1})^{-1} [\hat{\mathbf{P}}_1 \hat{\phi}_1 + \mathbf{F}^{(N-1)T} \mathbf{E}^T \mathbf{G}_1^{-1} \mathbf{d}]$$

This expression is similar to (2.11) in that ϕ_1^+ is calculated as a weighted sum of the estimate, and the final data \mathbf{d}_N (mapped back to $t=1$ under an adjoint transformation), which are both known. \mathbf{u}_1^+ is then calculated using (2.14), and the model (2.2) is marched forward in time to produce all of the ϕ_t^+ and the \mathbf{u}_t^+ using (2.14). This method involves the storage of the \mathbf{G}_t at each time-step.

The essential difference between dynamic programming and the other two methods is that the smoother and the adjoint required that the optimal final state, and some estimate of the initial state, be known before the optimal control sequence could be calculated, this is so-called ‘‘closed loop control’’. With dynamic programming, the optimal control at time t is only a function of: the state at time t and some estimate of the state at $t=N$, so-called open loop control. This is not an important difference for the present problem as all of the above solutions are in closed loop form since here we only have initial and final states. Dynamic programming has potential for solving other sorts of problems where one would like to know all possible state and control trajectories passing through some state that minimise some cost whilst satisfying a model evolution equation.

2.5 Results

Each of the three methods were run using the simple persistence model and the linear dynamical model. It was found necessary to use the RTS algorithm for the smoother so as to reduce the memory requirements. $\hat{\phi}_1$ and $\hat{\mathbf{P}}_1$ were set as the least-squares

estimates for survey A from section 1.3, \mathbf{d}_N , \mathbf{E} and \mathbf{R} are the modal data, the sampling matrix and the noise covariance for survey B. The variance of the process noise was rather arbitrarily set as $\mathbf{Q} = 0.01|\mathbf{P}_0|$.

It is to be expected that the smoothed initial and final states will be slightly different from the stationary fits, and that there will be some process noise estimated at each time-step.

Figure 2.2 shows the optimal wave amplitudes ϕ^+ for the initial (left) and final (right) times calculated using (top) Kalman Smoother with standard deviation $\sqrt{\text{diag}(\mathbf{P}_t^+)}$ of ϕ^+ , (middle) Adjoint method, and (bottom) Dynamic programming, using the linear model.

As expected the methods all produce the same optimal states at the initial and final times, since they all minimise the same linear cost function with the linear model as a constraint.

In figure 2.3 are the contour maps at 700 metres depth of the density anomaly produced by the optimal state trajectory, ϕ_t^+ , for the linear model using the Adjoint method. Surveys A and B correspond to times $t=1$ and $t=55$. Comparing these to the stationary fits from section 1.3, figure 1.10, we can see that the optimal estimators have estimated initial and final states that resemble each other more, than do the stationary fits to surveys A and B. For those waves whose amplitudes are significantly different from zero to one standard deviation, the quality of their fit to the modal data at the initial and final times is such that they give a fractional explained variance of $V=0.9640$ to survey A and $V=0.9283$ to survey B. Thus the fit to the data is good, while the process noise variance, \mathbf{Q} is low enough so that the model can interpolate the state between the two surveys.

Since there are uncertainties in the data passed to the optimal estimators, due to internal waves and to projecting the data onto modes, the methods do not force the state to fit the initial and final data exactly. The misfit allowed by the uncertainties

in the data is used to estimate initial and final states more in agreement with the state evolution equation.

Figure 2.4 shows the process noise estimated as an optimal control trajectory when the simple persistence model is used, and figure 2.5 shows the estimated process noise for the linear dynamical model. For both models the process noise is an order of magnitude less than the state wave amplitudes. The estimated process noise required for the optimal state trajectory using the linear model is substantially less than that required for the simple persistence model, indicating that the linear dynamics are better than no dynamics at all.

However, we saw that when left to themselves the waves do not propagate from survey A to survey B by the linear dynamical model, so the linear dynamics are not perfect. Some added control is needed at each time step. This estimated control has here been treated as some process noise for each wave in the model basis. This noise represents errors in the model state evolution equation. From the scaling in chapter one, we assume that the biggest error in the model is due to the ignored dynamical effect of the nonlinear wave-wave interactions. In the next chapter, this effect will be included in the state evolution equation and formulated in the control vector.

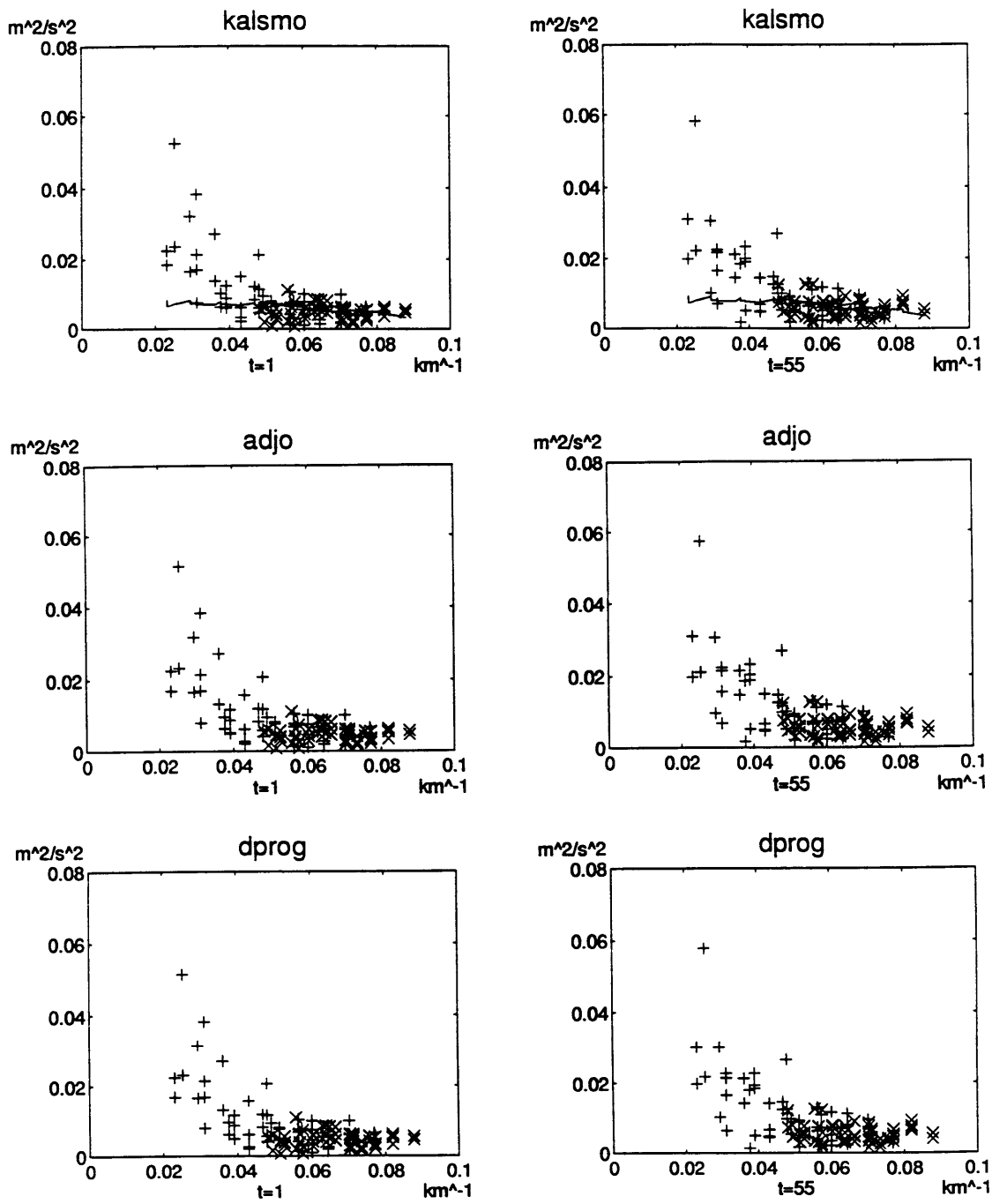


Figure 2.2: Optimal amplitudes of the 1st (+) and 2nd (x) mode waves as a function of total wavenumber for the initial (left) and final (right) times calculated using (top) Kalman Smoother with the standard deviation of the estimate shown, (middle) Adjoint method, and (bottom) Dynamic programming.

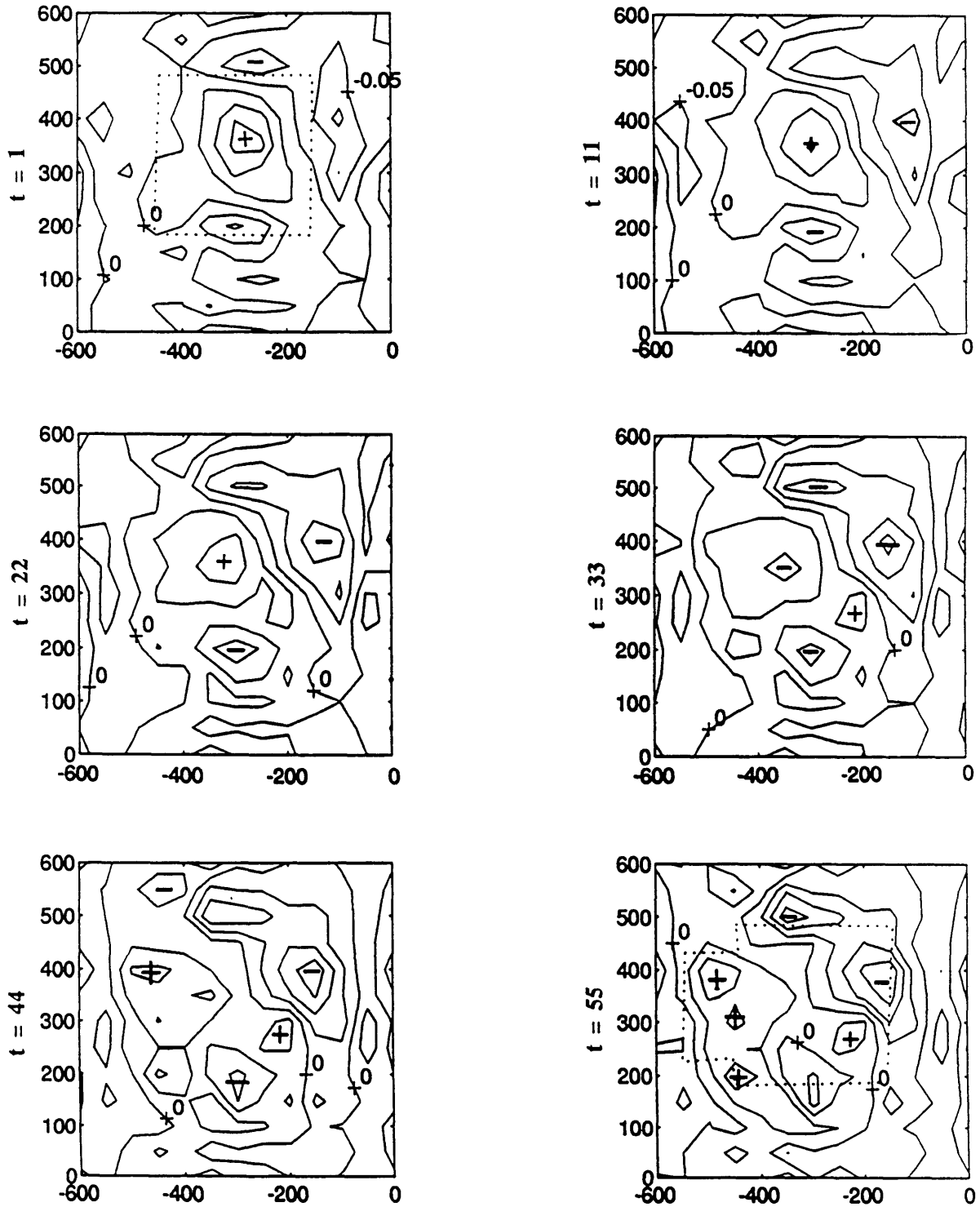


Figure 2.3: Time sequence of estimated density anomaly at 700m using the optimal wave amplitudes from the linear dynamical model (c.i.= 0.05 kg/m^3).

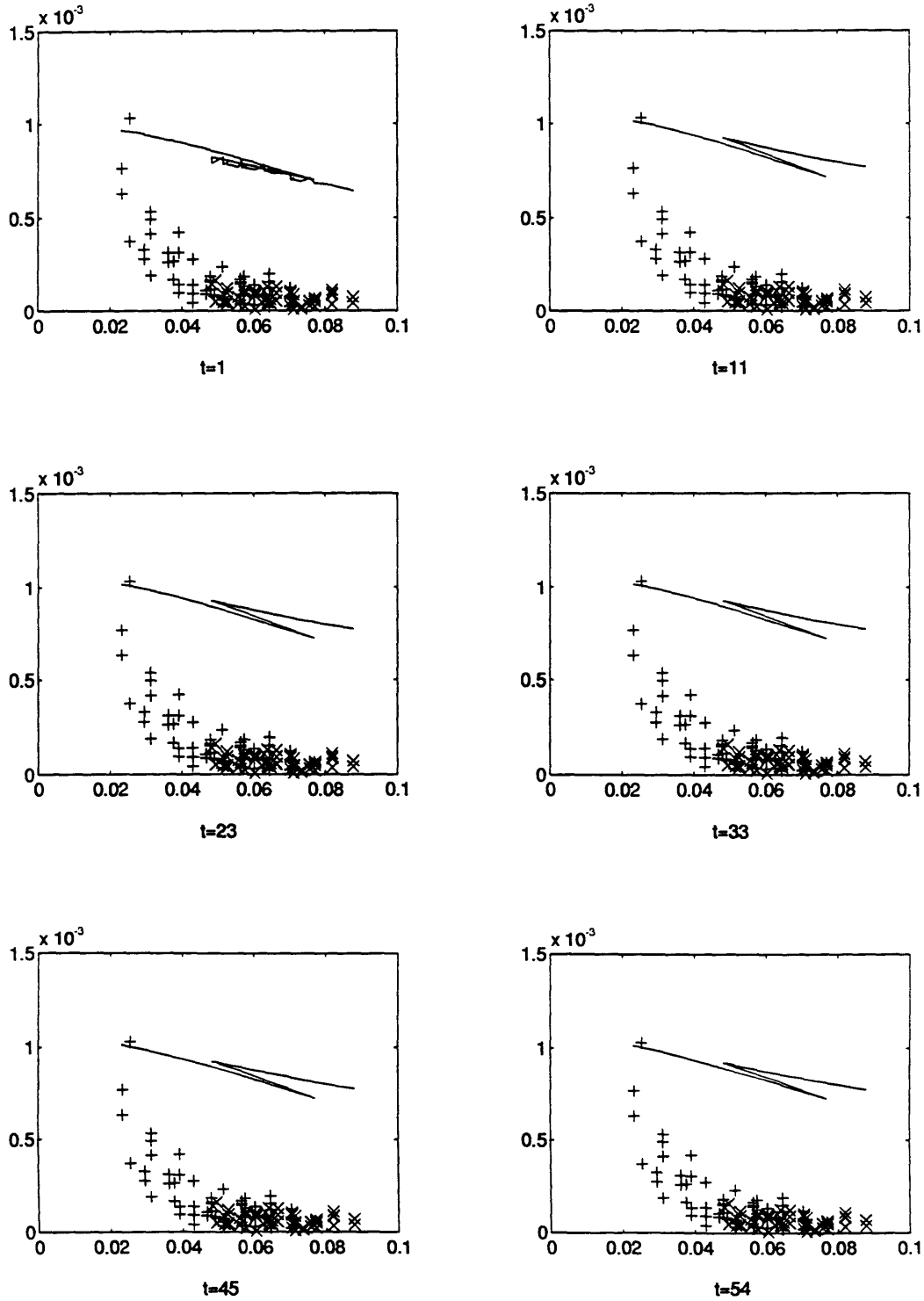


Figure 2.4: Process noise (m^2/s^2) for each wave with one standard deviation error curve, as a function of total wavenumber (km^{-1}) at 11 day intervals using the simple persistence model.

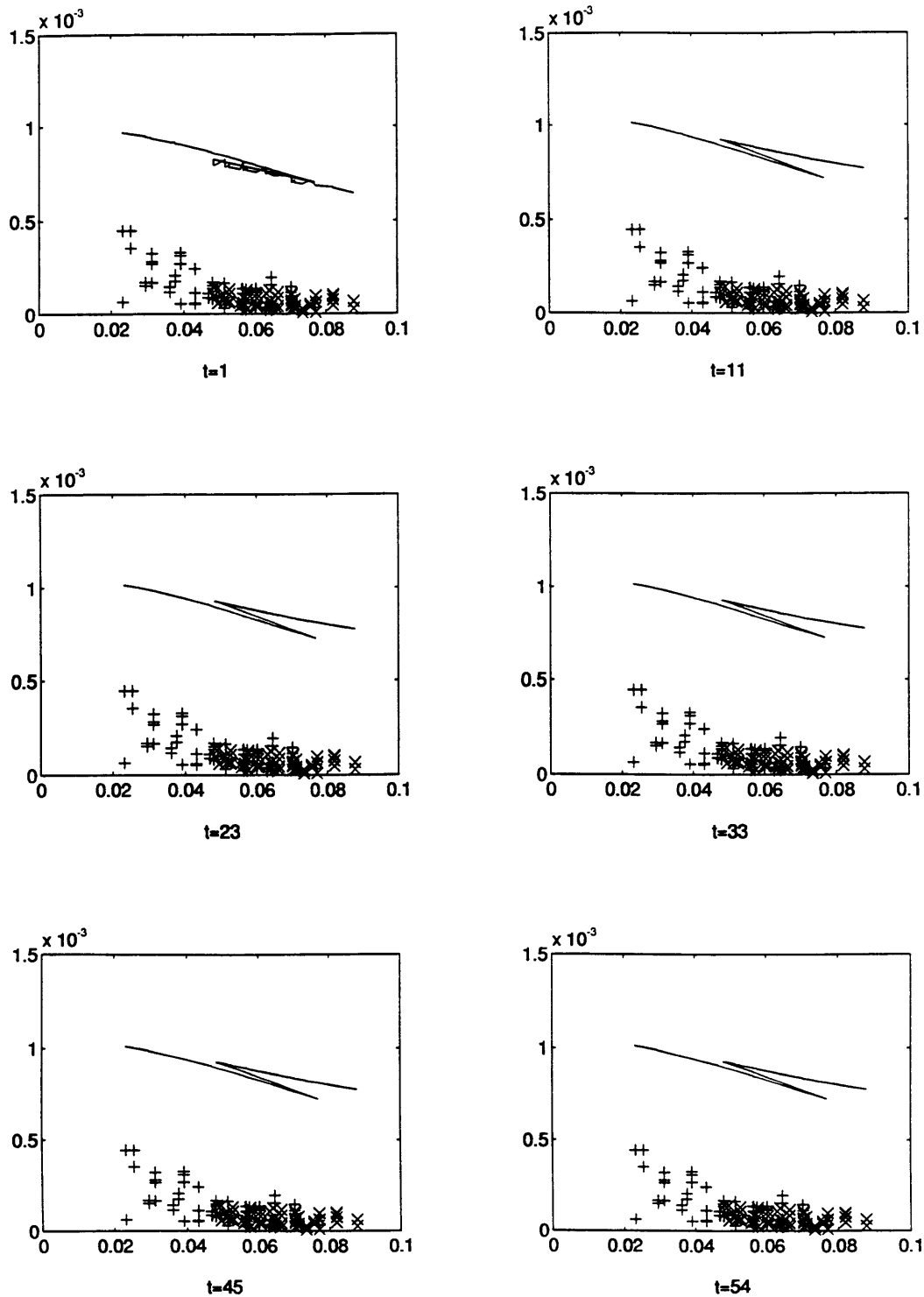


Figure 2.5: Process noise (m^2/s^2) for each wave with one standard deviation error curve, as a function of total wavenumber (km^{-1}) at 11 day intervals using the linear dynamical model.

Chapter 3

Estimating the Barotropic Flow

So far in this study simplified dynamics for the model state evolution have been used. At the simplest level we had persistence in time, then we saw that linear dynamics did a better job of taking the state from the initial to the final surveys. The nonlinear advective term in the conservation of potential vorticity equation has been ignored and from scaling, this term is the next most important effect to be included.

For freely propagating Rossby waves in a periodic domain the nonlinear term is represented as weakly nonlinear wave-wave interactions between waves that form resonant triads. Here I follow Fu and Flierl (1980), to formulate these interactions. The interactions for the waves in the model basis are then calculated, and included in the state evolution equation. The evolution equation is linearised so that the control methods so far developed can be applied. As in chapter two, optimal state and control trajectories are estimated that take the state from the initial to the final surveys.

Through the nonlinear interactions the 1st and 2nd baroclinic mode waves in the model basis not only interact with each other but also with waves in other modes. The control vector is now formulated as the unknown amplitudes of the barotropic waves that interact with the baroclinic waves in the model basis. In this way the barotropic component of the flow can be estimated, even though it is not able to be determined directly from the density data.

3.1 Wave-Wave interactions

Considering only motions in the 1st and 2nd baroclinic modes, the conservation of quasi-geostrophic potential vorticity, equation (1.14) becomes

$$\left(\frac{\partial}{\partial t} [\nabla^2 - \lambda_n^2] + \beta \frac{\partial}{\partial x} \right) \alpha_n + \frac{1}{f_0} \sum_{p,q=1}^2 J(\alpha_p, [\nabla^2 - \lambda_q^2] \alpha_q) \xi_{npq} = 0$$

To examine how the nonlinear term affects the evolution of these waves, substitute in a triplet of waves from (1.15), the Fourier Series over the periodic model domain,

$$\alpha_n(x, y, t) = \Re \left\{ \phi_{an}(t) e^{i(k_a x + l_a y)} + \phi_{bn}(t) e^{i(k_b x + l_b y)} + \phi_{cn}(t) e^{i(k_c x + l_c y)} \right\}.$$

and then integrate over the periodic model domain. The Jacobian term vanishes except for when $k_a = k_b + k_c$ and $l_a = l_b + l_c$ then we get

$$\begin{aligned} \kappa_{an}^2 \frac{d\phi_a}{dt} &= -i\beta k_a \phi_a - \frac{1}{2f_0} (k_a l_b - k_b l_a) \sum_{p,q=1}^2 \left(\kappa_{bq}^2 \phi_{bq} \phi_{cp} - \kappa_{cq}^2 \phi_{bp} \phi_{cq} \right) \xi_{npq} \\ \kappa_{bn}^2 \frac{d\phi_b}{dt} &= -i\beta k_b \phi_b - \frac{1}{2f_0} (k_a l_b - k_b l_a) \sum_{p,q=1}^2 \left(\kappa_{cq}^2 \phi_{ap} \phi_{cq}^* - \kappa_{aq}^2 \phi_{aq} \phi_{cp}^* \right) \xi_{npq} \\ \kappa_{cn}^2 \frac{d\phi_c}{dt} &= -i\beta k_c \phi_c + \frac{1}{2f_0} (k_a l_b - k_b l_a) \sum_{p,q=1}^2 \left(\kappa_{bq}^2 \phi_{ap} \phi_{bq}^* - \kappa_{aq}^2 \phi_{aq} \phi_{bp}^* \right) \xi_{npq} \end{aligned} \quad (3.1)$$

where $\kappa_{an}^2 \equiv k_a^2 + l_a^2 + \lambda_n^2$, and ϕ_{an} is an abbreviation for $\phi(k_a, l_a, \lambda_n)$, ie. horizontal wavenumber (k_a, l_a) , and vertical mode n . ϕ^* is the complex conjugate of ϕ .

These equations express the conservation of potential vorticity of each wave in the triplet. The rate of change of the potential vorticity of each wave, $-(k^2 + l^2 + \lambda_n^2)\phi(t)$, is due to two effects: advection of planetary vorticity north and south, and to nonlinear interactions with the other two waves. The former contribution gives rise to planetary waves and the latter are contributions from other waves present that form triad interactions, ie. that have horizontal wavenumbers such that

$$k_a = \pm k_b \pm k_c \quad \text{and} \quad l_a = \pm l_b \pm l_c.$$

For such waves the advection of the potential vorticity of wave c by wave b , say, affects the rate of change of the potential vorticity of wave a .

For the model basis of 130 1st and 2nd mode waves, as shown in figure 1.1, there are many possible triad interactions for almost all of the waves. These can be computed beforehand. The evolution of the state vector of all the wave amplitudes, can be put in matrix form as

$$\frac{d\phi}{dt} = \mathbf{F}\phi(t) + C(\phi(t)) \quad (3.2)$$

where $C(\phi_t)$ is the vector of non-linear contributions to the evolution of the state vector, the second terms on the right in (3.1). This term involves products of wave amplitudes. Each row of this matrix equation represents one of the equations in (3.1), with the nonlinear contributions summed over all triad interactions. As was done in chapter 2, this matrix equation is discretised by rendering the time derivative in Crank-Nicholson form. However the size of the time step must be reduced significantly from that used in the linear model since the non-linear term is not calculated at exactly the same time that the derivative is calculated, and hence numerical instabilities can develop. The equation can be evolved forward in time from some initial state, calculating $C(\phi(t))$ at each time step.

The model state was evolved forward from, $\hat{\phi}_1$. $C(\phi(t))$ was calculated at each time step from the state at that time. It was found that when a time-step greater than one day was used, the wave amplitudes blew up over the 54 day time period. For time-steps of less than one day the model was stable but very slow and used a lot of memory. Since now the transition matrix from one state to the next is time-dependent, for the control methods it must be stored at each time step. A time-step of one day was used. In figure 3.1 is shown the time sequence of density anomaly at 700 metres obtained by evolving the state forward from $\hat{\phi}_1$ using equation (3.2).

Comparing these snapshots to those of figure 2.1 where the linear dynamics were used, it is clear that the nonlinearities in the model are important. The state trajectories shown at 11 day intervals for the two models do not resemble each other.

Shorter time scale motions are present using the nonlinear model, due to the wave-wave interactions moving amplitude information around the model basis.

Also comparing the field at $t=55$ with the stationary fit to survey B, we can see that the two are quite different. The explained variance of the waves from this forward run with the data at survey B is $V=-.88$, indicating that the waves fit to survey A do not evolve on their own to match the data well at survey B. Some control may be needed and the fit to survey A can be adjusted so that the waves evolve toward survey B.

3.2 Formulation of the control

The barotropic component of the flow is indeterminate from the density data, and is thus not represented in the model state. However, examination of the full wave equation, (1.14), shows that it is possible for motions in the barotropic mode to interact with motions in the baroclinic modes. Hence it is possible for the barotropic dynamic pressure to enter into the state evolution equation for the baroclinic flow as an unknown control variable.

When the barotropic dynamic pressure mode is expanded in the same horizontal Fourier Series as the baroclinic modes, ie.

$$\begin{aligned}\alpha_0(x, y, t) &= \sum_{k,l} \phi_{kl0}(t) e^{i(kx+ly)} \\ &\equiv \sum_{k,l} \pi_{kl}(t) e^{i(kx+ly)}.\end{aligned}$$

the barotropic waves can interact with the baroclinic waves via the same triad interactions as described in the last section. Equation (1.14) becomes

$$\begin{aligned}\left(\frac{\partial}{\partial t} [\nabla^2 - \lambda_n^2] + \beta \frac{\partial}{\partial x}\right) \alpha_n &+ \frac{1}{f_0} \sum_{p,q=1}^2 J(\alpha_p, [\nabla^2 - \lambda_q^2] \alpha_q) \xi_{npq} \\ &+ \frac{1}{f_0} J(\alpha_n, \nabla^2 \alpha_0) + \frac{1}{f_0} J(\alpha_0, [\nabla^2 - \lambda_1^2] \alpha_n) = 0\end{aligned}$$

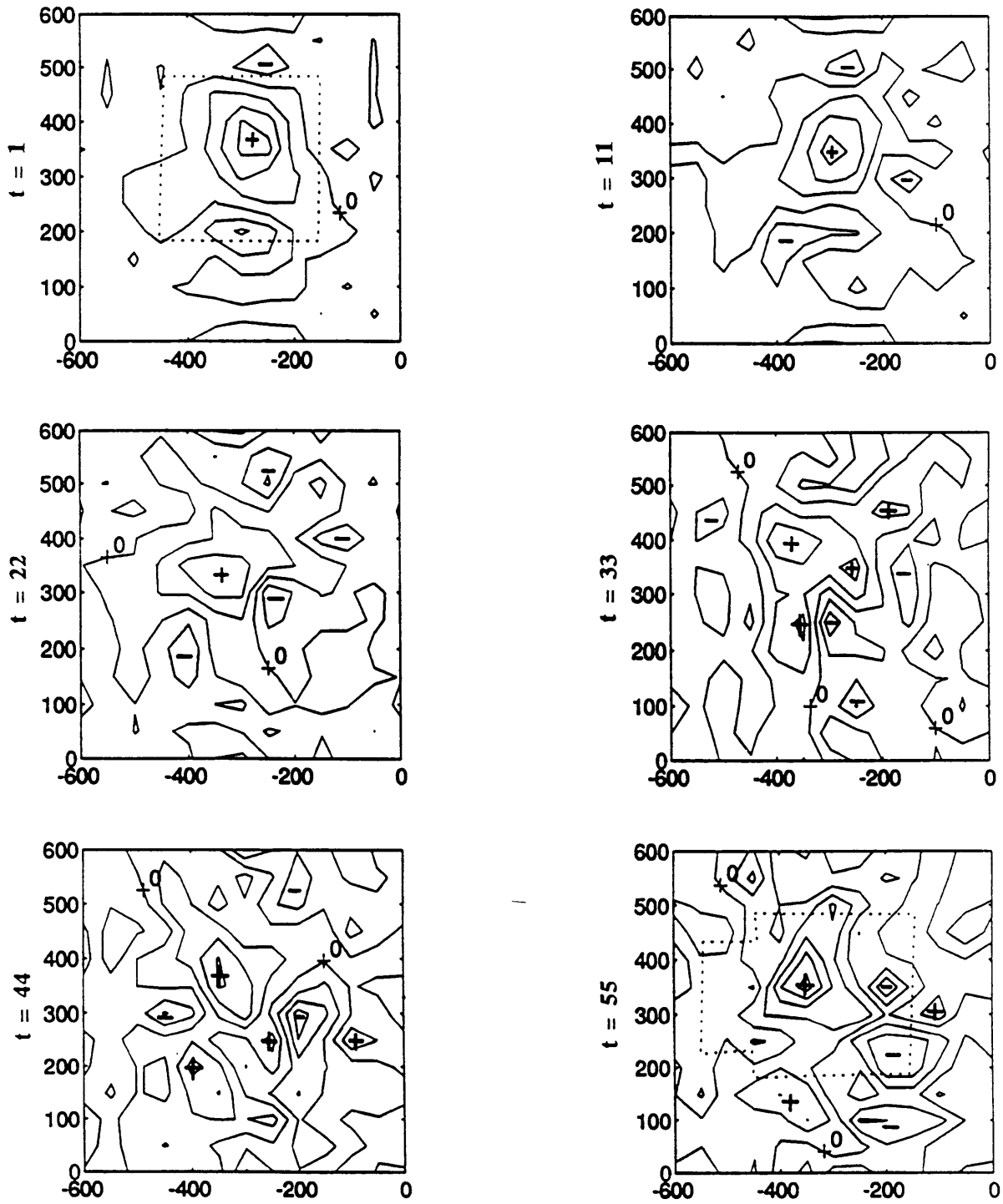


Figure 3.1: Time sequence of density anomaly at 700 metres from evolving the state estimated from survey A forward in time including wave-wave interactions (c.i.=0.05 kg/m³).

where $n=1$ or 2 and $\xi_{101} = \xi_{110} = \xi_{202} = \xi_{220} = 1$. In each Jacobian term is the unknown barotropic component, α_0 .

This equation can be put in a discrete matrix form like (3.2),

$$\frac{d\phi}{dt} = \mathbf{F}\phi(t) + C(\phi(t)) + B(\phi_t, \pi_t) \quad (3.3)$$

which is still nonlinear. C and B are functions of products of wave amplitudes. For the linear methods of control theory developed in chapter 2, the state evolution equation needs to be in the linear form

$$\phi_{t+1} = \mathbf{D}_t\phi_t + \mathbf{B}_t\pi_t$$

where the state and control transition matrices, \mathbf{D} and \mathbf{B} can be time-dependent but are not to be functions of the present state or control vectors.

Equation (3.3) is linearised by linearising C and B about some known state trajectory. $C(\phi(t))$ is formulated as $\mathbf{C}_t\phi(t)$ by setting

$$\begin{aligned} C(\phi_{an}(t)) &= -\frac{1}{2f_0} \sum_{b,c} (k_a l_b - k_b l_a) \sum_{p,q=1}^2 (\kappa_{bq}^2 \phi_{bq}(t) \phi_{cp}(t) - \kappa_{cq}^2 \phi_{bp}(t) \phi_{cq}(t)) \xi_{npq} \\ &= -\frac{1}{2f_0} \sum_{b,c} (k_a l_b - k_b l_a) \sum_{p,q=1}^2 (\kappa_{bq}^2 \phi'_{bq}(t) \phi_{cp}(t) - \kappa_{cq}^2 \phi'_{bp}(t) \phi_{cq}(t)) \xi_{npq} \\ &\equiv \mathbf{C}_t\phi(t) \end{aligned}$$

where \mathbf{C}_t is a function of $\phi'(t)$, which here denotes some state trajectory previously calculated.

Similarly, $B(\phi_t, \pi_t)$ is approximated as

$$\begin{aligned} B(\phi_{an}(t), \pi(t)) &= -\frac{1}{2f_0} \sum_{b,c} (k_a l_b - k_b l_a) \sum_{p,q=0}^1 (\kappa_{bq}^2 \phi_{bq}(t) \phi_{cp}(t) - \kappa_{cq}^2 \phi_{bp}(t) \phi_{cq}(t)) \\ &\equiv \mathbf{B}_t\pi(t) \end{aligned}$$

where \mathbf{B}_t is also a function of $\phi'(t)$. Also $\phi_{c0} \equiv \pi_c$ and $\phi_{b0} \equiv \pi_b$ within the summation.

In this way, \mathbf{C}_t and \mathbf{B}_t are made independent of the present state, and we can write

$$\phi_{t+1} = \mathbf{D}_t\phi_t + \mathbf{B}_t\pi_t$$

$$\equiv (\mathbf{F}_t + \mathbf{C}_t)\phi_t + \mathbf{B}_t\pi_t$$

as the model to use in the control methods. The time-dependent transition matrices, \mathbf{D}_t and \mathbf{B}_t , are calculated and stored for each of the 54 time-steps. The optimal control methods can now be applied using this model.

It is important to note that here π_t represents any unknown control affecting the state evolution. As well as the time-dependent amplitudes of the barotropic waves that interact with the baroclinic waves, π_t also contains any process noise due to errors in the model such as truncation of the model basis.

I am going to try two different state trajectories to linearise about, the first being the smoothed state from chapter two using the linear model, the second being the state from the nonlinear model of the previous section evolved forward in time from the initial survey. The optimally estimated state and control, that is the baroclinic and barotropic wave amplitudes, will probably be different for the two linearisations since the two state trajectories look quite different.

One could possibly determine a best state trajectory to linearise about by iteratively calculating an optimal state trajectory then linearising the transition matrices about the new optimal state trajectory.

3.3 Results

These assimilations were run using the methods formulated in chapter two. The three methods all produced the same results and I am here showing the results using the Kalman filter/smoothen. The time-step I am using is one day so there are 54 time-steps between the two surveys.

Case 1: The nonlinear model is linearised about the smoothed state trajectory obtained in chapter two using the linear dynamical model. The optimal state and control trajectories are calculated. The time series of the density anomaly at 700 metres every eleven days as produced by the optimal state, is shown in figure 3.2.

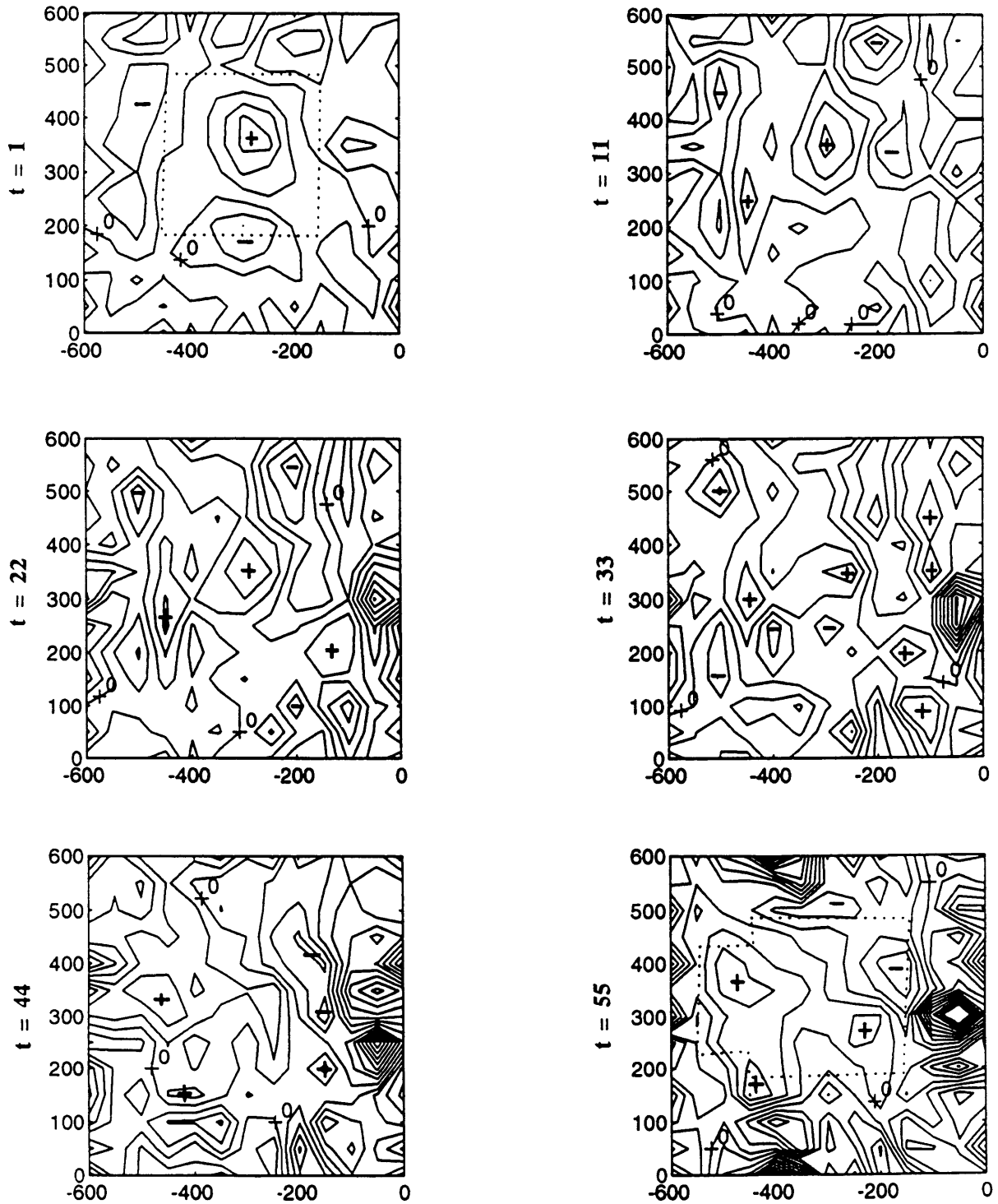


Figure 3.2: Time sequence of estimated density anomaly at 700m every eleven days from the optimal state for case 1 ($c.i.=0.05 \text{ kg/m}^3$).

Case 2: The nonlinear model is linearised about the state trajectory obtained by evolving the state forward in time from survey A using the nonlinear model. The time series of the density anomaly every eleven days as produced by this optimal state, is shown in figure 3.3.

Both cases produce initial and final states that are consistent with the data. This can be seen by comparing the agreement of the area within the dotted region of the first and last frames of figures 3.2 and 3.3 with the stationary fits to surveys A and B as shown in figure 1.10. There is high underdeterminedness of the model parameters with respect to the available data so the estimated field within the data domain is reproduced well by the optimal estimators.

Outside of the data domain in space and time the two cases produce quite different states, due to the models used in each case having different transition matrices. There is energy going into high wavenumbers in the part of the model domain outside of the data domain. This is due to the wave-wave interactions transferring energy to higher wavenumbers as most of the energy is initially at low wavenumbers. These short scale features are amplified by Gibbs phenomena: Since the Fourier Series is truncated (the shortest wavelength in the model basis is 120 km), small scale spatial changes in the field such as at the boundaries are not well modelled. Any sudden changes in the field are taken up by the higher wavenumbers.

Assuming that there is no process noise, the control vector then represents only the barotropic waves. The time sequence of barotropic dynamic pressure can be computed from the optimal control trajectory for cases 1 and 2, and is shown in figures 3.4 and 3.5. The estimated barotropic flows for both cases are quite different from each other, illustrating that the optimal control trajectory is dependent on which state trajectory we linearise the wave-wave interactions about.

Qualitatively, case 1 produces a smoother time-dependent fit. There is less energy at high wavenumbers in the optimal state from the linear dynamics, and it has been smoothed by both surveys. For case 2, the trajectory is linearised about the

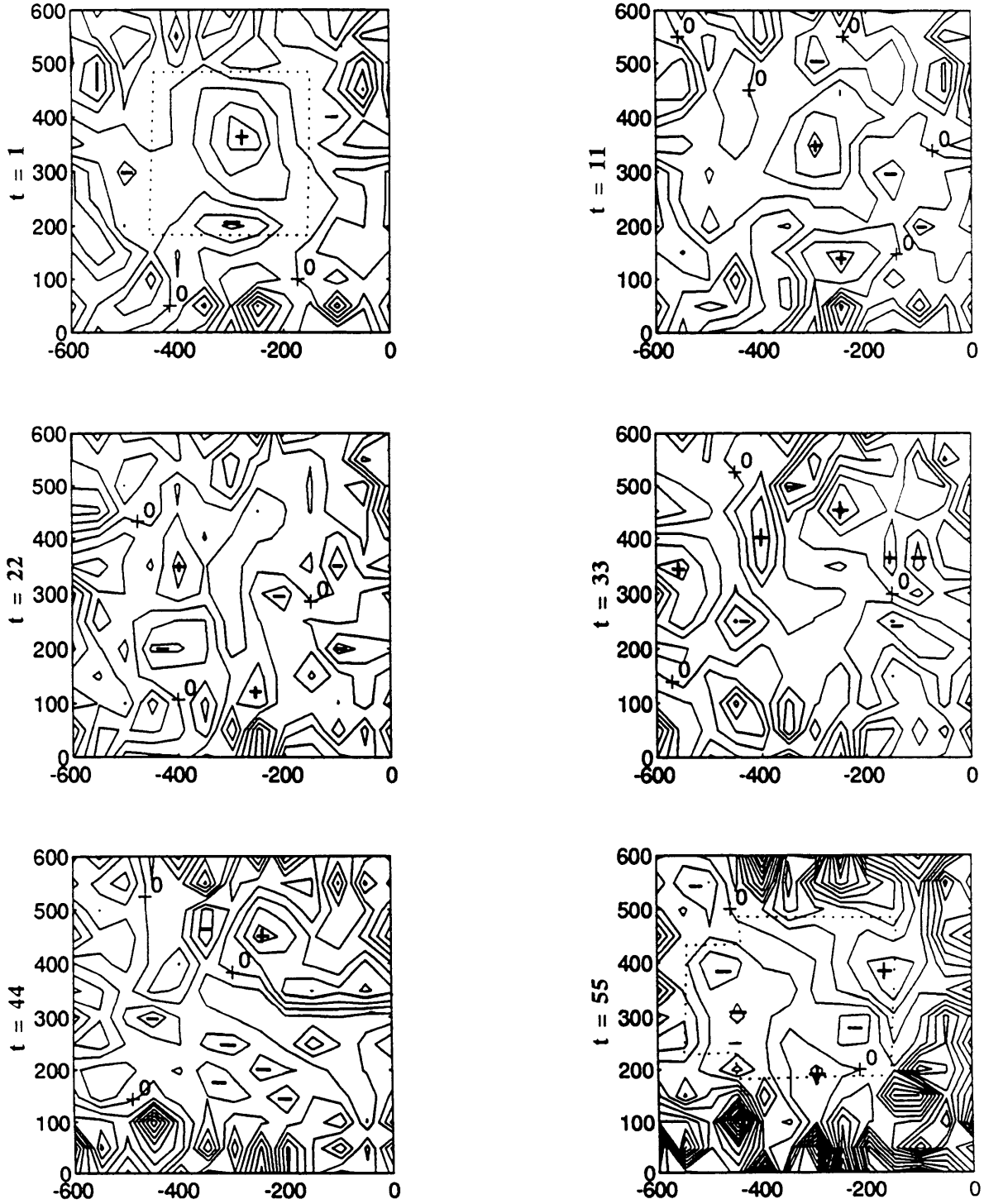


Figure 3.3: Time sequence of estimated density anomaly at 700m every eleven days from the optimal state for case 2 ($c.i.=0.05 \text{ kg/m}^3$).

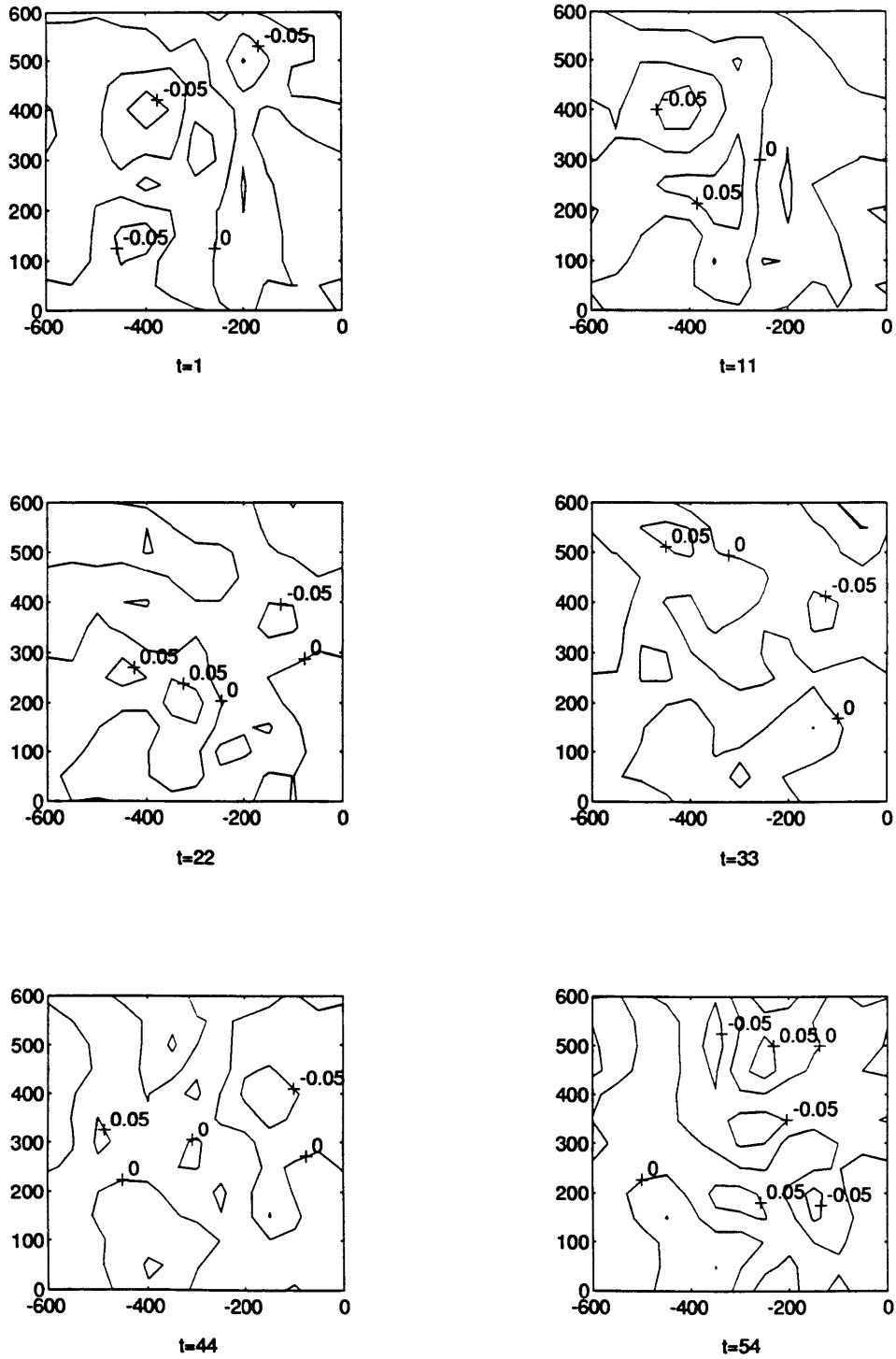


Figure 3.4: Estimated barotropic dynamic pressure time sequence for case 1, calculated from the optimal control trajectory (c.i.= $0.05 \text{ m}^2/\text{s}^2$).

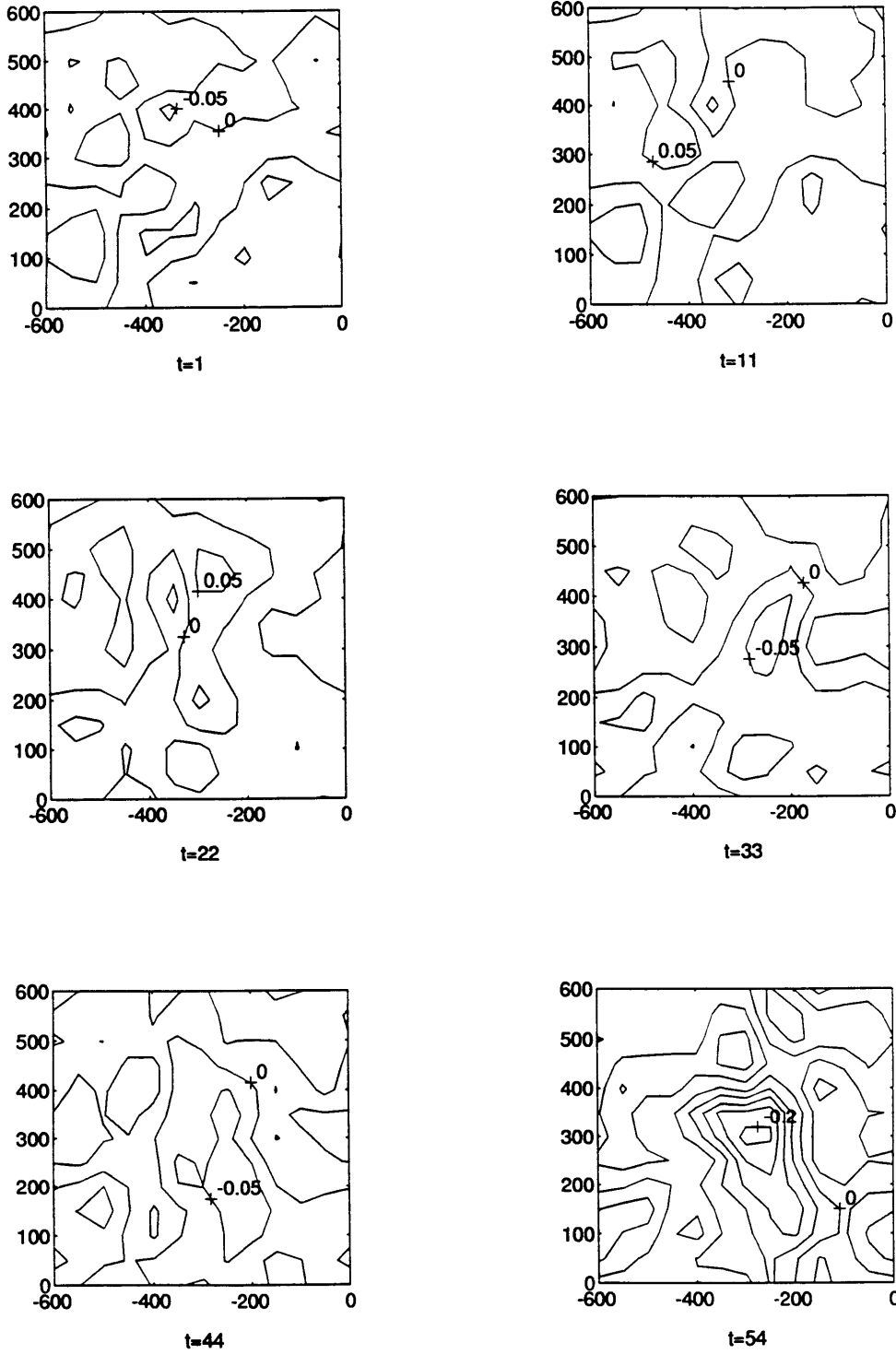


Figure 3.5: Estimated barotropic dynamic pressure time sequence for case 2, calculated from the optimal control trajectory (c.i.=0.05 m²/s²).

state trajectory containing the wave-wave interactions, but it only evolves forward from survey A and does not contain any information about survey B. Consequently, the model state is “surprised” by the data at survey B. This is illustrated in the time sequence of barotropic dynamic pressure for case 2 as shown in figure 3.5. Its amplitude increases suddenly at the last times, so as to force the state evolution away from the linearised trajectory toward the state at survey B (see figure 1.10) which does not resemble the state evolved forward by the nonlinear model starting from survey A (see figure 3.1).

These results illustrate how there can be different flows between the two surveys by linearising the nonlinear model about different state trajectories.

Conclusions

I have here shown an effective means of combining oceanographic data that is sparse in space and time with a dynamical model using optimal control theory. The size of the model state space was reduced by representing all dynamical quantities in terms of the amplitudes of quasi-geostrophic Rossby waves. The size of the data space was also reduced by projecting the data onto the vertical modes. With the model and data equations of a comfortable size and with a linear model and cost function, the Kalman smoother, the adjoint method and dynamic programming could be manipulated to give explicit forms for the optimal state and control trajectory estimates. One of the first pleasing results of this study was that all of the three methods produced the same optimal state and control sequences, which was to be expected.

The three methods were applied here to examine their relative merits. All of the computations were performed in MATLAB on a SPARC2 workstation, which takes advantage of the matrix/vector formulation. The calculation of the wave-wave interactions in the transition matrices, however, were done element by element and could have been computed more quickly in FORTRAN. Comparing the times taken for the three methods:

Using linear dynamics (chapter 2),	
Kalman filter/smoothen	66 minutes
Adjoint	1.5 minutes
Dyn. Prog.	28 minutes
Including wave-wave interactions (chapter 3),	

Kalman filter/smoothen	137 minutes
Adjoint	45 minutes
Dyn. Prog.	53 minutes

Out of the three optimal control methods used, the adjoint performs best in terms of speed. When the wave-wave interactions are included the transition matrices for the state and control vectors are updated at each time step, thus increasing the time taken for all of the methods. The adjoint and dynamic programming both use much less memory than the filter/smoothen but do not produce error estimates. A next step is to calculate error estimates of the optimal state and control for the adjoint method and dynamic programming by examining the Hessian of the cost function.

In chapter two the linear dispersion relation was shown to be a better model for state evolution between the two surveys than simple persistence in time. The process noise needed as a control at each time step to drive the model state between the surveys was found to be greater using the simple persistence model than that needed for the linear model.

The nonlinear wave-wave interactions of the model state were addressed in chapter three and the interactions with the barotropic waves were formulated as an unknown control. The wave-wave interaction terms were incorporated into the state and control transition matrices which were then linearised about a known state trajectory. For case 1, the transition matrices were linearised about the optimally estimated trajectory from chapter two using the linear dynamics. For case 2, they were linearised about the trajectory of the state evolved forward in time from survey A using the full nonlinear model for the baroclinic waves. Case 2 involves a state trajectory that contains more complete dynamics. The nonlinear interactions give rise to motions of shorter time scale than the period of the fastest baroclinic wave, 160 days. These shorter time scales of 50 to 150 days are energetic in this part of the ocean (Richman et al, 1977), and linear Rossby waves cannot model these motions. Thus one would expect the nonlinear state trajectory to be more realistic. However

the trajectory used in case 2 to linearise the model about did not evolve towards the final state for survey B, so large amounts of control were needed at the last time steps to correct the state. The dynamically linear trajectory used in case 1 smoothly transferred the state from the initial to the final states consistently with the data, but was lacking the short time scale motions known from observations and theory to be significant. The result is two realisations of the barotropic flow between the two surveys, corresponding to the two linearisations.

Iteratively one could find better trajectories to linearise about by running the assimilation repeatedly, each time linearising the transition matrices about the last optimal state trajectory.

In chapter 3 the barotropic flow was estimated assuming there was no process noise in the model. A further step is to run the assimilation including the barotropic waves already estimated and estimate the process noise as the unknown control. Comparing this process noise those estimated in chapter 2 one could test whether the model has been improved by the inclusion of the estimated barotropic waves.

To address these issues the assimilation needs to be run on a faster computer with more memory such as a CRAY then the time step can be reduced so that there is less numerical error in the time integrations.

Implicit in the cost function that I started from, equation (2.1), is that the unknown control vector to be determined is uncorrelated in time and has zero mean. Thus a test of the optimal control estimates is that they have no temporal structure and resemble a white noise processes. The temporal structure of the optimal control trajectories computed in this project should be examined to see if they violate this assumption. Correlation in time of the control vector can be incorporated into the optimal estimation (see Bryson and Henrikson, 1968).

An interesting prospect is the inclusion of the time series of travel times from the acoustic tomography array. This array spans the region covered by the CTD surveys shown in figure 1.2, and provide a spatially integrated measure of the sound-

speed perturbations. This data would provide much more precise information about mesoscale motions than point measurements. The travel times could be incorporated into the assimilation presented here by explicitly calculating the Fourier Series representation of the spatial slices covered by the travel times, then relating this to the amplitudes of the waves in the model basis.

REFERENCES

- Brink, K. H., 1989: Evidence for wind-driven current fluctuations in the western North Atlantic, *J. Geophys. Res.*, **94**, 2029-2044.
- Brogan, W. L., 1982: *Modern Control Theory*, Prentice-Hall, Englewood Cliffs, N. J., 393pp.
- Bryson, A. E., and L. J. Henrikson, 1968: Estimation using sampled data containing sequentially correlated noise, **5**, 662-665.
- Bryson, A. E. Jr., and Y-C Ho, 1975: *Applied Optimal Control*, Hemisphere, New York, 481 pp.
- Charney, J. G., 1971: Geostrophic turbulence, *J. Atmos. Sci.*, **28**, 1087-1095.
- Chiu, C-S, and Y. Desaubies, 1987: A planetary wave analysis using the acoustic and conventional arrays in the 1981 ocean tomography experiment, *J. Phys. Oceanogr.*, **17**, 1270-1287.
- Cornuelle, B., C. Wunsch, D. Behringer, T. Birdsall, M. Brown, R. Heinmiller, R. Knox, K. Metzger, W. Munk, J. Spiesberger, R. Spindel, D. Webb and P. Worcester, 1985: Tomographic maps of the ocean mesoscales. Part I: Pure Acoustics, *J. Phys. Oceanogr.*, **15**, 133-152.
- Fjørtoft, R., 1953: On the changes in the spectral distribution of kinetic energy for two-dimensional, non-divergent flow, *Tellus*, **5**, 225-230.
- Flierl, G. R., 1978: Models of vertical structure and the calibration of two-layer models, *Dyn. Atm. Oceans*, **2**, 341-381.
- Fu, L-L., and G. R. Flierl, 1980: Nonlinear energy and enstrophy transfers in a realistically stratified ocean, *Dyn. Atmos. Oc.*, **4**, 219-246.
- Fu, L-L, J. Vazquez, and C. Perigaud, 1991: Fitting dynamic models to the Geosat sea level observations in the tropical Pacific ocean. Part I: A free wave model, *J. Phys. Oceanogr.*, **21**, 798-809.
- Gaspar, P., and C. Wunsch, 1989: Estimates from altimeter data of barotropic Rossby waves in the northwestern Atlantic ocean, *J. Phys. Oceanogr.*, **19**, 1821-1844.
- Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography, *Adv. Geophys.*, **33**, 141-266.

- Gill, A. E., J. S. A. Green and A. J. Simmons, 1974: Energy partition in the large-scale ocean circulation and the production of mid-ocean eddies, *Deep-Sea Res.*, **21**, 499-528.
- Hogg, N. G., 1988: Stochastic wave radiation by the Gulf Stream, *J. Phys. Oceanogr.*, **18**, 1687-1701.
- Malanotte-Rizzoli, D. B. Haidvogel and R. E. Young, 1987: Numerical simulation of transient boundary-forced radiation. Part I: The linear regime, *J. Phys. Oceanogr.*, **17**, 1439-1457.
- Malanotte-Rizzoli, P., N. G. Hogg and R. E. Young, 1992: Stochastic wave radiation by the Gulf Stream: numerical experiments, (in preparation).
- McWilliams, J. C., and G. R. Flierl, 1976: Optimal, quasi-geostrophic wave analyses of MODE array data, *Deep-Sea Res.*, **23**, 285-300.
- Ocean Tomography Group, The, 1982: A demonstration of ocean acoustic tomography, *Nature*, **299**, 121-125.
- Pedlosky, J., 1979: *Geophysical Fluid Dynamics*, Springer-Verlag, New York, 624pp.
- Rauch, H. E., F. Tung and C. T. Strebil, 1965: Maximum likelihood estimates of linear dynamic systems, *AIAA Journal*, **3**, 1445-1450.
- Rhines, P. B., 1975: Waves and turbulence on a beta-plane, *J. Fluid Mech.*, **69**, 417-443.
- Rhines, P. B., 1977: The dynamics of unsteady currents, in *The Sea*, vol. 6, edited by E. D. Goldberg et al., John Wiley, New York.
- Richman, J. G., C. Wunsch and N. G. Hogg, 1977: Space and time scales of mesoscale motion in the western North Atlantic, *Rev. Geoph. Space Phys.*, **15**, 385-420.
- Sorensen, H. W., 1985: *Kalman Filtering: Theory and Applications*, IEEE press, 457pp.
- Thacker, W. C., 1989: The role of the Hessian matrix in fitting models to measurements, *J. Geophys. Res.*, **94**, 6177-6196.
- Tziperman, E., and W. C. Thacker, 1989: An optimal-control/adjoint-equations approach to studying the oceanic general circulation, *J. Phys. Oceanogr.*, **19**, 1471-1485.
- Wunsch, C., 1988: Transient tracers as a problem in control theory, *J. Geophys. Res.*, **93**, 8099-8110.

Wunsch, C., 1989: Using data with models. Ill-posed and time-dependent ill-posed problems. *Proceedings Ecole de Physique Theorique. on Geophysical Tomography*, Les Houches France, Y. Desaubies, A. Tarantola and J. Zinn-Justin, Eds., Elsevier, Amsterdam, 203-246.