dissertation

THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

VARIATIONAL DATA ASSIMILATION WITH 2-D SHALLOW WATER EQUATIONS AND 3-D FSU GLOBAL SPECTRAL MODELS

By

A Dissertation submitted to the
Department of Mathematics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Fall Semester, 1993

The members of the Committee approve the dissertation of ZHI WANG defended on November 10, 1993.

I. Michael Navon
Professor Directing Thesis

James J. O'Brien
Outside Committee Member

Christopher Hunter
Committee Member

Francois Xavier Le Dimet
Committee Member

Louis N. Howard
Committee Member

This work is dedicated to My Wife, My Parents, My Brothers and Sister.

ACKNOWLEDGEMENTS

First, and foremost, I would like to express my thanks to Dr. I. M. Navon, my major professor. He was the driving force behind the development of this study. I am fortunate and thankful to have had Dr. I. M. Navon always there to consistently point me in the right direction. I also wish to thank Dr. James J. O'Brien, Dr. C. Hunter, Dr. F. X. Le Dimet and Dr. L. Howard for serving on my committee.

The author thanks Dr. F. X. Le Dimet for contributing original ideas related to the second order adjoint analysis and optimal nudging assimilation. His help is highly appreciated.

Special thanks go to T. N. Krishnamurti at the Department of Meteorology of the Florida State University for providing me the Florida State University Global Spectral Model (FSUGSM), to the many people at the Department of Mathematics and to Dr. X. Zou of the Supercomputer Computations Research Institute of the Florida State University for her contribution towards my understanding of the subject.

My deepest appreciation to all my friends who gave me encouragement through the last few years.

The author gratefully acknowledges the support of Dr. Pamela Stephens, Director of GARP Division of the Atmospheric Sciences at NSF. This research was funded by NSF Grant ATM-9102851. Additional support was provided by the Supercom-

puter Computations Research Institute at Florida State University, which is partially funded by the Department of Energy through contract No. DE-FC0583ER250000.

TABLE OF CONTENTS

List of Tables **x**

List of Figures **xi**

Abstract **xiv**

1 INTRODUCTION **1**

1.1 Overview 1

1.2 Four dimensional variational data assimilation in meteorology 5

1.3 Preliminaries 8

1.3.1 The definition of a variational method 8

1.3.2 Description of the first order adjoint analysis 12

1.4 The theme of my dissertation 16

2 SECOND ORDER ADJOINT ANALYSIS: THEORY AND APPLICATIONS **21**

2.1 Introduction 21

2.2 The second order adjoint model 22

2.2.1 Theory of the second order adjoint model 22

2.2.2 The estimate of the condition number of the Hessian 25

2.3 The derivation of the second order adjoint model for the shallow water equations model 27

2.3.1 The verification of the correctness of the first and the second order adjoint models 31

2.4 Second order adjoint information 32
2.4.1 Calculation of the Hessian/vector product 32
2.4.2 The uniqueness of the solution 33
2.4.3 Convergence analysis 35
2.5 Sensitivity analysis for observations 37
2.5.1 Sensitivity theory for observation 37
2.5.2 Numerical results from model generated data 40
2.5.3 Numerical results using real analysis of First Global Geophysical Experiment (FGGE) data 42
2.6 Conclusions 44

3 THE ADJOINT TRUNCATED NEWTON ALGORITHM FOR LARGE-SCALE UNCONSTRAINED OPTIMIZATION 58

3.1 Introduction 58
3.2 The adjoint Newton algorithm 60
3.2.1 Description of the first and second order adjoint theories . . . 60
3.2.2 The adjoint truncated-Newton method. 62
3.3 Numerical results obtained using the adjoint truncated-Newton algorithm. 63
3.3.1 Application of the adjoint truncated-Newton algorithm to variational data assimilation 64
3.3.2 Numerical results 65
3.3.3 An accuracy analysis of the Hessian/vector product. 67
3.4 Conclusions 72

4 THE FSU GLOBAL SPECTRAL MODEL 79

4.1 Introduction 79
4.2 Model equations 82

4.3 Spectral form of the governing equations 87
4.3.1 Grid to spectral transform 89
4.3.2 Spectral to grid transform 91
4.4 The semi-implicit algorithm 91
4.5 Vertical discretization 94

5 4-D VARIATIONAL DATA ASSIMILATION WITH THE FSU GLOBAL SPECTRAL MODEL 101

5.1 Introduction 101
5.2 Model developments and numerical validation 102
5.2.1 The cost function and its gradient with respect to control variables 102
5.2.2 The accuracy of the tangent linear model 104
5.2.3 The accuracy of the first order adjoint model 105
5.3 Descent algorithms and scaling 108
5.3.1 The LBFGS algorithm of Liu and Nocedal 108
5.3.2 Weighting and scaling 109
5.4 Numerical results of the variational data assimilation 111
5.4.1 Numerical results with shifted initial conditions 112
5.4.2 Numerical results with randomly perturbed initial conditions . 114
5.4.3 Impact of horizontal diffusion 115
5.4.4 Impact of observations distributed in time 117
5.5 Conclusions 118

6 VARIATIONAL NUDGING DATA ASSIMILATION 134

6.1 Introduction 134
6.2 Nudging data assimilation 137
6.2.1 Some simple examples 137

6.2.2 The basic theory of nudging data assimilation 140
6.2.3 Variational nudging data assimilation 142
6.2.4 Identifiability and stability of parameter estimation 145
6.3 Numerical results of the variational data assimilation 148
6.3.1 Variational nudging data assimilation with the adiabatic version of the FSU global spectral model 149
6.3.2 Variational nudging data assimilation with the FSU global spectral model in its full-physics operational form 153
6.3.3 Comparisons of estimated nudging, optimal nudging and variational data assimilation 155
6.4 Conclusions 156

7 SUMMARY AND CONCLUSIONS 172

A A simple example with exact solutions 178

B A nonlinear model with exact solutions. 182

C TN: Nash's Truncated Newton Method. 187

D The continuous form of the adjoint equations 192

E Kalman filtering, variational data assimilation and variational nudging data assimilation 194

F Physical processes 199

F.0.1 Radiative processes 199
F.0.2 Cumulus convection and large scale condensation 201
F.0.3 Boundary layer processes 204
F.0.4 Dry convective adjustment 207

F.0.5 Surface topography 208
F.0.6 Computational requirements 208

REFERENCES **209**

LIST OF TABLES

2.1 A comparison between the actual and predicted changes due to identical perturbations at specified locations at the beginning of the assimilation window. 57

3.1 Numerical results of minimization algorithms: adjoint truncated Newton, truncated Newton and LBFGS for the minimization of the cost function in the variational data assimilation problem when observations are model generated and control variables are the initial conditions. 76

3.2 Rms errors between the Hessian/vector products obtained by using the second order adjoint and the finite-difference techniques for various differencing parameters at the end of 15 iterations of the adjoint truncated Newton minimization algorithm, respectively. The maximum number of conjugate gradient inner-iterations is 4 and h is the differencing parameter used in the original truncated Newton algorithm of Nash minimization algorithm. 77

4.1 Gaussian latitudes of the T-42 model. + indicates north, and − south. 92

5.1 Rms errors of temperature 126

6.1 Optimal nudging coefficients obtained by VNDA where the 158

B.1 Errors between the true evolution given by 185

B.2 The absolute errors between the exact and approximate obtained by second order adjoint integration Hessian/vector product. 186

LIST OF FIGURES

1.1 Schematic representation of the 4-D variational data assimilation . . . 8

1.2 Schematic representation of a minimization process, where ellipses are the isolines of the cost function J, $U0$ is the initial guess, U is the optimal solution and each arrow represents an iteration of the minimization process. 12

2.1 Schematic flowchart of the relationships between the model equations and the adjoint equations 46

2.2 Verifications of the correctness of the gradient calculation 47

2.3 The first 50 components of the Hessian/vector products at the optimal solution obtained by the finite-difference approach (dash line), and the second order adjoint method (solid line) when $\alpha = 10$ 48

2.4 Same as Figure 2.3 except $\alpha = 3$ 48

2.5 Same as Figure 2.3 except $\alpha = 0.01$ 49

2.6 Same as Figure 2.5 except at the initial guess point 49

2.7 Variation of the smallest eigenvalue of the Hessian of the cost function with the number of iterations. 50

2.8 Variation of the largest eigenvalue of the Hessian of the cost function with the number of iterations. 50

2.9 Distribution of the randomly perturbed geopotential field. 51

2.10 Reconstructed geopotential field after 6 iterations of minimization. . . 51

2.11 Reconstructed geopotential field after 25 iterations of minimization. . 52

2.12 Variation of the condition numbers of the Hessian of the cost function with respect to the number of iterations. 52

2.13 Time variation of the sensitivities of cost function J 53

2.14 Time variation of the sensitivities of cost function J to 1% observational error at points (x_{10}, y_{10}) (solid line) (x_5, y_{15}) (dotted line), and (x_{15}, y_5) (dash line) in the wind and the geopotential fields. 54

2.15 Time variation of the sensitivities of cost function J to 1% observational error on all grid points in the wind and the geopotential fields. 54

2.16 Distribution of (a) the geopotential and (b) the wind fields for the First Global Geophysical Experiment (FGGE) data at 0UTC 05/26, 1979 on the $500mb$. The contour intervals are $200m^2/s^2$ and the magnitude of maximum vector is 0.311E+02m/s. 55

2.17 Distribution of the difference fields of the geopotential (a) and the wind (b) fields at 18UTC and 0UTC on $500mb$ between 18UTC and 0UTC times. The contour intervals are $100m^2/s^2$ and the magnitude of maximum vector is 0.210E+02m/s. 55

2.18 Time variation of the sensitivities of the cost function J to 1% observational error in the wind and the geopotential fields at grid points (x_{10}, y_{10}) (solid line) and (x_7, y_5) (dotted line). 56

3.1 Variations of the *log* of the scaled cost function (J_k/J_0) 75

3.2 The first 50 components of the difference between the Hessian/vector products 78

4.1 Schematic representation of the vertical structure of the model. The solid lines represents the levels σ_n. The dashed lines represents the layers $\tilde{\sigma}_n$. $\sigma = 0$ is at the top of the atmosphere and $\sigma = 1$ is at the earth's surface. 95

5.1 Verifications of the correctness of the tangent 120

5.2 Verifications of the correctness of the gradient 120
5.3 Variations of the log of the scaled 121
5.4 Dotted and long dash lines denote the rms between 122
5.5 Divergence (top left), vorticity (top right) 123
5.6 Divergence (top left), vorticity (top right), temperature 124
5.7 Same as Fig. 5.5 125
5.8 Variations of the log of the scaled 127
5.9 Divergence (top left), vorticity (top right), temperature 128
5.10 Divergence (top left), vorticity (top right) 129
5.11 Same as Fig. 5.5 130
5.12 Same as Fig. 5.5 131
5.13 Same as Fig. 5.3 132
5.14 Dotted and long dash lines denote the rms between 133

6.1 Verification of the correctness of the gradient calculation 158
6.2 Schematic representation of the assimilation-forecast cycle. 159
6.3 Variation of the nudging coefficients, $G_{\ln p}$ (dash line), ... 160
6.4 Variation of the *log* of the scaled cost function (J_k/J_0) 161
6.5 Divergence (top left), vorticity (top right), temperature (bottom left) and natural log of surface pressure (bottom right) difference fields . . 162
6.6 Variation of the *log* of the scaled cost function (J_k/J_0) 163
6.7 Same as Fig. 6.3 164
6.8 Same as Fig. 6.5 165
6.9 Variation of the *log* of the scaled cost function (J_k/J_0) 166
6.10 Same as Fig. 6.3 167
6.11 Same as Fig. 6.5 168
6.12 Variation of the *log* of the scaled cost function (J_k/J_0) 169

6.13 Same as Fig. 6.3 170
6.14 Solid, long dash and dash lines denote the rms errors between the forecast and exact temperature at the end of the forecast period. the rms between 171

ABSTRACT

This thesis is aimed at (a) conducting an in-depth investigation of the feasibility of the 4-D variational data assimilation (VDA) applied to realistic situations; (b) achieving an improvement of the existing large-scale unconstrained minimization algorithms. The basic theory of the VDA was developed in a general mathematical framework related to optimal control of partial differential equations (PDEs) [116]. The VDA attempts to find an initialization (if control variables are the initial conditions) most consistent with the observations over a certain period of time as linked together with a forecast model through a cost functional which is a statement about our knowledge and uncertainties concerning the forecast model and observations distributed in space and time (hence, the phrase 4-D VDA). This goal is accomplished by defining the cost function to be a weighted sum of the squares of the difference between the model solution and the observations; an attempt to minimize this cost function is then sought by using an iterative minimization algorithm in the framework of optimal control of PDEs [116]. The gradient of the cost function with respect of the control variables is obtained by integrating the so-called first order adjoint model backwards in time.

The present thesis first develops the second order adjoint (SOA) theory and applies it to a shallow-water equations (SWE) model on a limited-area domain. One integration of such a model yields a value of the Hessian (the matrix of second partial derivatives of the cost function with respect to control variables) multiplied by a vector. The SOA model was then used to conduct a sensitivity analysis of the cost function with respect to observations distributed in space and time and to study the evolution of the condition number (the ratio of the largest to smallest eigenvalues)

of the Hessian during the course of the minimization since the condition number is strongly related to the convergence rate of the minimization. It is proved that the Hessian is positive definite during the process of the large-scale unconstrained minimization, which in turn proves the uniqueness of the optimal solution for our test problem.

Experiments using data from an European Center for Medium-range Weather Forecasts (ECMWF) analysis of the First Global Geophysical Experiment (FGGE) show that the cost function J is more sensitive to observations at points where meteorologically intensive events occur. The SOA technique shows that most changes in the value of the condition number of the Hessian occur during the first few iterations of the minimization and are strongly correlated to major large-scale changes in the retrieved initial conditions fields.

It is also demonstrated that the Hessian/vector product thus obtained is more accurate than that obtained by applying a finite-difference approximation to the gradient of the cost function with respect to the initial conditions. The Hessian/vector product obtained by the SOA approach is applied to one of the most efficient minimization algorithms, namely the truncated-Newton (TN) algorithm of Nash ([129], [130], [134]). This modified version of the TN algorithm of Nash differs from it only in the use of a more accurate Hessian/vector product for carrying out the large-scale unconstrained optimization required in VDA. The newly obtained algorithm is referred to as the adjoint truncated-Newton algorithm (ATN). The ATN is applied here to a limited-area SWE model with model generated data where the initial conditions serve as control variables. We then compare the performance of the ATN algorithm with that of the original TN Nash [134] method and the LBFGS method of Liu and Nocedal [117]. Our numerical tests yield results which are twice as fast as these obtained by the TN algorithm both in terms of number of iterations as well

as in terms of CPU time. Further, the ATN algorithm turns out to be faster than the LBFGS method in terms of CPU time required for the problem tested.

Next, the thesis applies the VDA to a realistic 3-D numerical weather prediction model, requiring the derivation of the adjoint model of the adiabatic version of the Florida State University Global Spectral Model (FSUGSM). The experiments with FSUGSM are designed to demonstrate the numerical feasibility of 4-D VDA. The impact of observations distributed over the assimilation period is investigated. The efficiency of the 4-D VDA is demonstrated with different sets of observations. The results obtained from a forecast starting from the new initial conditions obtained after performing VDA of model generated data sets are meteorologically realistic while both the cost function and root-mean-square (rms) errors of all fields were reduced by several orders of magnitude irrespective if whether the initial conditions are shifted or randomly perturbed. It is also demonstrated that the presence of horizontal diffusion in the model yields a more accurate solution to the VDA problem.

In all of the previous experiments, it is assumed that the model is perfect, and so is the data. The solution of the problem will have a perfect fit to the data, with zero difference. This is of course an unrealistic assumption and the lack of inclusion of model errors constitutes a serious deficiency of the assimilation procedure.

The nudging data assimilation (NDA) technique introduced by Anthes [4] consists in achieving a compromise between the model and observations by relaxing the model state towards the observations during the assimilation period by adding a non-physical diffusion-type term to the model equations. Variational nudging data assimilation (VNDA) combines the VDA and NDA schemes in the most efficient way to determine optimally the best initial conditions and optimal nudging coefficients simultaneously. The humidity and other different parameterized physical processes are not included in the adjoint model integration. Thus the calculation of the gradient by the adjoint model is approximate since the forecast model is used in its

full-physics (diabatic) operational form. It is shown that it is possible to perform 4-D VDA with realistic forecast models even prior to more complex adjoint models being developed, such as models including the adjoint of all physical processes [215]. The resulting optimal nudging coefficients are then applied in NDA (or physical initialization) (thus the term optimal nudging data assimilation (ONDA)) [229].

CHAPTER 1

INTRODUCTION

1.1 Overview

We wish to know and understand not only the climatological or current state of the atmosphere, but also to predict its future state (the aim of numerical weather prediction). Beyond the qualitative understanding of the atmosphere, a quantitative estimate of its state in the past and present, as well as quantitative prediction of future states is required. The estimate of the present state is prerequisite for future prediction, and the accuracy of past prediction is essential for an accurate estimate of the present.

How does the estimation of the present proceed in meteorology? A good starting point is Wiener's article [223] on prediction and dynamics. At the time of his writing, meteorology, like economics, could still be considered a semi-exact science, as opposed to the allegedly exact sciences of celestial mechanics [89]. Dynamical processes in the atmosphere were poorly known, while observations were sparse in space and time as well as inaccurate. Relying theoretically on the hope of the system's ergodicity and stationarity, Wiener argued that the best approach to atmospheric estimation and prediction was statistical. In practice, this meant ignoring any quantitative dynamical knowledge of system behaviour, requiring instead a complete knowledge of the system's past history, and using the Wiener-Hopf filter to process this infinite but inaccurate information into yielding an estimate of the present and future of weather dynamics [222].

During roughly the same period, synoptic meteorologists were actually producing charts of atmospheric fields at present and future times guided by tacit principles

similar to those explicitly formulated by Wiener. The main tool was smooth interpolation and extrapolation of observations in space and time. Still, rudimentary and quantitative dynamical knowledge was interpolated into these estimates of atmospheric states, such as the geostrophic relation between winds and heights, and the advection of large-scale features of the prevailing winds.

The first step leading to the present state of art of estimation in meteorology was objective analysis, which replaced manual, graphic interpolation of observations by automated mathematical methods, such as for instance two-dimensional (2-D) polynomial interpolation [160]. Not surprisingly, this step was largely motivated by the rapidly improving knowledge of atmospheric dynamics to produce numerical weather forecasts [27]. The main ideas underlying objective analysis were statistical [57, 63, 165]. Observations are considered to sample a random field, with a given spatial covariance structure which is preconditioned and stationary in time. This generalized in fact the idea of Wiener [223] from a finite-dimensional system governed by Ordinary Differential Equations (ODEs) to an infinite-dimensional one governed by Partial Differential Equations (PDEs) of Geophysical Fluid Dynamics (GFD). In practice, these statistical ideas appeared too complicated and computationally expensive at the time to be adopted as they stood into the fledgling Numerical Weather Prediction (NWP) process. Instead, various shortcuts, such as the successive-correction method were implemented in the operational routine of weather bureaus [39].

Two related developments led to the next step, in which a connection between statistical interpolation on one hand, and dynamics, on the other, became apparent and started to be used systematically. One development was the increasingly accurate nature of short-term numerical weather forecasts; the other was the advent of time continuous, space-borne observing systems. Together, they produced the concept of four-dimensional (4-D) space-time continuous data assimilation in which a model

forecast of atmospheric fields is sequentially updated with incoming observations [28, 192, 178]. Here the model carries forward in time the knowledge of a finite number of past observations, subject to the appropriate dynamics, to be blended with the latest observations.

At this point, we note merely that noisy, inaccurate data should not be fitted by exact interpolation, but rather by a procedure which has to achieve simultaneously two goals (a) to extract the valuable information contained in the data, and (b) to filter out the spurious information, i.e. the noise. Thus the analyzed data should be close to the data but not too close. The statistical approach to this problem is linear regression. The variational approach consists in minimizing the distance, e.g., in a quadratic norm, between the analyzed field and the data, subject to constraints which yield smoother results. The merger of these two approaches into a stationary, ergodic context is intuitively obvious, and is reflected in the fact that root-mean-square (rms) minimization is used in popular parlance for both approaches.

In summary, during the last decade due to a constant increase in the need for more precise forecasting and now-casting, several important developments have taken place in meteorology directed mainly in two directions [107, 109]:

1. Modeling at either large scale or at smaller scales. Recently, many models have been developed including an ever increasing detail of physical processes and parameterizations of sub-grid phenomena.

2. Data: new sources of data such as satellite data, radar, profiles, and other remote sensing devices have led to an abundance of widely distributed data in space and time. However, a common characteristic of these data is to be heterogeneous either in their space or time density or in their quality.

Therefore, a cardinal problem is how to link together the model and the data. This problem includes several questions:

1. How to retrieve meteorological fields from sparse and/or noisy data in such a way that the retrieved fields are in agreement with the general behaviour of the atmosphere? (Data Analysis)

2. How to insert pointwise data in a numerical forecasting model? This information is continuous in time, but localized in space (satellite data for instance)? (Data Assimilation)

3. How to validate or calibrate a model from observational data? The dual question in this case being how to validate observed data when the behaviour of the atmosphere is predicted by a numerical model.

For these questions a global approach can be defined by using a variational formalism. Variational data assimilation (VDA) consists of finding the assimilating model solution which minimizes a properly chosen objective function measuring the distance between model solution and available observations distributed in space and time. The assimilating model solution is obtained by integrating a dynamic system of partial differential equations from a set of initial conditions (and/or boundary conditions for limited area problems). Therefore, the complete description of the initial atmospheric state in a numerical weather prediction method constitutes an important issue. The four dimensional VDA method offers a promising way to achieve such a description of the atmosphere using a methodology originating in the theory of optimal control of distributed parameters.

Several techniques of assimilation have been used so far, e.g. optimal interpolation and inverse methods, blending and nudging methods as well as Kalman filtering applications [41, 68]. The optimal interpolation scheme (OI) is a particular form of statistical interpolation and has been widely used amongst most operational centers [118]. However several weaknesses inherent in the method and its practical implementation are now identified. For instance, the OI analysis extracts information

poorly from observations nonlinearly related to the model variables [2]. The data selection algorithm is also generally a source of noise which persists in the final analysis [161].

The variational approach circumvents some of the practical OI weakness, since it allows the analysis to use all the observations at every model grid point, and can easily handle a non trivial link between the observations and the model state. In its 4-D version, the method consistently uses the information coming from the observations and the dynamics of the model. Over a period of time it produces the same results as the full extended Kalman-Bucy filter approach [207] at a much lower cost. But the Kalman-Bucy filter approach does have its own advantages. For instance, since it is a sequential estimation method it is capable of providing explicit error estimates, such as the error covariance matrix of the obtained solution [68].

2-D VDA was implemented at ECMWF [59]. 3-D VDA has been successfully implemented operationally at the National Meteorological Center (NMC), USA [53], yielding consistently better analyses and forecasts when compared with their classical OI system.

The cost of the 4-D VDA is still prohibitive with current computer power and there are remaining problems (such as, for instance, the treatment of non-differentiable processes in the model) to be solved prior to its being implemented operationally. However, the theoretical advantages of 4-D VDA as compared with the current operational data assimilation systems make it a good candidate for a possible near future operational assimilation scheme. As a matter of fact 4-D VDA has recently been used with real data in a semi-operational set-up at ECMWF.

1.2 Four dimensional variational data assimilation in meteorology

The first application of variational methods in meteorology has been pioneered by Sasaki ([179, 180]). Washington and Duquet [220], Stephens ([196, 197]) and

Sasaki ([181, 182, 183, 184]) have given a great impetus towards the development of variational methods in meteorology.

In a series of basic papers Sasaki ([181, 182, 183, 184]) generalized the application of variational methods in meteorology to include time variations and dynamical equations in order to filter high-frequency noise and to obtain dynamically acceptable initial values in data void area. In all these applications, the Euler-Lagrange equations were used to calculate the optimal state. In general Euler-Lagrange equations are a coupled PDE system of mixed type of well-posed initial-boundary value problems, and can be solved numerically with reasonably computational cost [64, 139, 140, 142]. However, in most cases of real interest, this approach has not proved particularly useful or promising. VDA circumvents the Euler-Lagrange equations by directly minimizing a cost function measuring the misfit between the model solution and the observations with respect to the control variables.

VDA was first applied in meteorology by Marchuk [127] and by Penenko and Obrazstov [164]. Kontarev [95] further described how to apply the adjoint method to meteorological problems, while Le Dimet [106] formulated the method in a general mathematical framework related to optimal control of partial differential equations based on the work of Lions [116]. In the following years, a considerable number of experiments has been carried out on different two dimensional (2-D) barotropic models by several authors, such as Courtier [33]; Lewis and Derber [112]; Derber [50]; Hoffmann [87]; Le Dimet and Talagrand [108]; Derber [51]; Talagrand and Courtier [202]; Lorenc [120, 121]; Thacker and Long [204], Zou *et al.* [227], Navon *et al.* [152]. Most of the published scientific papers including all the aforementioned papers related to the use of adjoint methods in 4-D VDA have employed simplified models. Only has recently the method been applied to more complex models (Thépaut and Courtier [207], Navon *et al.* [149], and Thépaut *et al.* [210]). The main conclusion was that the method may be applicable to realistic situations. However in most of

these papers, observational data to be assimilated is idealized, in the sense that the observations are generated by the model itself. In this case, it is assumed that the model is perfect, and so is the data. The solution of the problem will have a perfect fit to the data, with zero difference. This is an unrealistic assumption and the lack of inclusion of model errors constitutes a serious deficiency of assimilation procedure.

While major advances have been achieved in the application of the adjoint method, this field of research remains both theoretically and computationally active [208, 209, 221, 213]. Nowadays the typical cost functional includes both model errors and observational errors as well as background errors [42, 234]. Additional research to be carried out includes: (a) applications to complicated models such as multilevel primitive equation models related to distributed real data and the inclusion of physical processes in the VDA process; (b) establishing more efficient large-scale unconstrained minimization algorithms; (c) deriving efficient ways to carry out 4-D VDA by using high performance parallel computers by using for instance domain decomposition methods [150].

In parallel with the introduction of variational methods in meteorology, starting in the 1960's and 1970's, mathematicians in collaboration with researchers from other scientific disciplines have achieved significant advances in optimization theory and optimal control, both from the theoretical viewpoint as well as from the computational one. In particular significant advances have been achieved in the development of large-scale unconstrained and constrained optimization algorithms ([10, 60, 61, 73, 123, 134, 143, 144, 145, 146, 147, 167] to cite but a few).

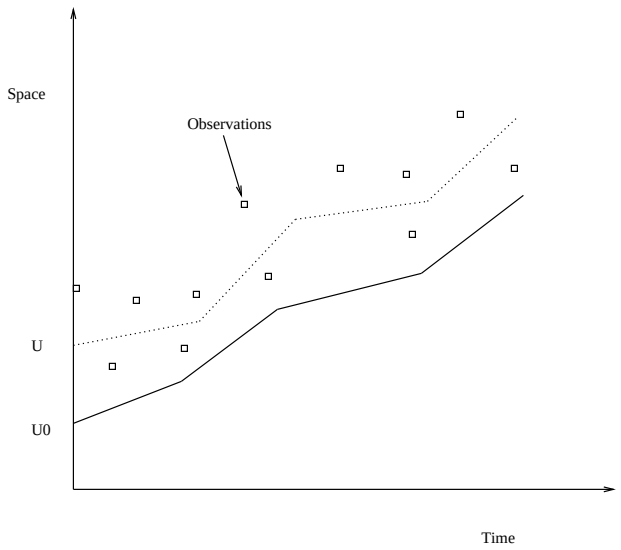

Figure 1.1: Schematic representation of the 4-D VDA where squares represent observations, $U0$ is the initial guess, U is the optimal solution and dotted and solid lines represent the trajectories starting from U and $U0$. Starting from $U0$, the 4-D VDA finds the best initial condition U such that the sum of the weighted squares of the difference between model solution and the observations obtains minimum.

1.3 Preliminaries

1.3.1 The definition of a variational method

The aim of the VDA [148] is to find an initial state for a numerical forecast by processing observations which are distributed in space and time (hence, the phrase 4-D data assimilation). The variational data assimilation attempts to find a set of initial conditions most consistent with the observations over a certain period of time as linked together with a forecast model. This goal is accomplished by defining a cost function consisting of the weighted sum of the squares of the difference between model solution and the observations; an attempt to minimize this cost function is then carried out (see Fig. 1.1). Therefore, the 4-D variational data assimilation

problem may be mathematically defined as

$$\min_{U} \; J(U), \tag{1.1}$$

where the cost function $J(U)$ is defined as

$$J(U) = J_o + J_b + J_p, \tag{1.2}$$

where J_o, J_b and J_p are the distance to the observations, the distance to the background and the penalty term which contains physical constraints applied to the model state $\vec{X}$, respectively [3, 83, 158, 214]. To perform a 4-D VDA over a time period, one needs to find the model trajectory which minimizes the misfit between model solution and the observations over the period, as well as the misfit to a guess at the initial time. This initial guess $\vec{X}_g$ represents the best estimate of the model state at the initial time, prior to any collection of observations at and after the initial time. As in all the operational assimilation schemes, the initial guess $\vec{X}_g$ represents a summary of all the information on $\vec{X}$ accumulated before the initial time. A natural distance for observational errors is the sum of weighted squares of the differences between the model solution and the observations. For simplicity, we assume the cost function is defined by

$$J(U) = J_o = \frac{1}{2}\sum_{i=0}^{M} < W(C\vec{X}_i - \vec{X}_i^o), (C\vec{X}_i - \vec{X}_i^o) >, \tag{1.3}$$

where $\vec{X}_i$ is the state variable vector at the i-th time level in a Hilbert space $\mathcal{X}$ whose inner product is denoted by $< \cdot, \cdot >$ using either an Euclidean norm or other suitably defined norms such as an energy norm. $M + 1$ is the total number of time levels in the time assimilation window $[t_0, t_f]$ where t_0 and t_f are the initial and final times and W is a weighting function usually taken to be the inverse of the covariance matrix of observation errors. The objective function $J(U)$ is the weighted sum of the squares of the distance between the model solution and available observations

distributed in space and time. $\vec{X}_i^o$ is the observation vector at the i-th time level, while the operator C represents a projection operator from the space of the model solution $\vec{X}$ to the space of observations. The control variable U and the state vector $\vec{X}$ satisfy model equations

$$\frac{\partial \vec{X}}{\partial t} = F(\vec{X}), \tag{1.4}$$

$$\vec{X}(t_0) = U, \tag{1.5}$$

where t is the time, and $F \in C^2$ is a function of $\vec{X}$, where C^2 denotes a set of twice continuously differentiable functions.

It is worth noting that the control variable U may consist of initial conditions and/or boundary conditions or model parameters to be estimated. For mesoscale numerical weather prediction (NWP) models and for steady state problems the model equations will be given by $F(\vec{X}) = 0$. Once U is defined, Eq (1.4) has an unique solution, $\vec{X}$.

In order to determine or at least approximate the optimal solution of Eq. (1.1) and therefore the optimal associate state of the atmosphere, we first have to set up an optimality condition. A general optimality condition is given by the variational inequality [116]

$$(\nabla J(U^*), V - U^*) \geq 0, \tag{1.6}$$

for all V belonging to a set of admissible control space U_{ad} where ∇J is the gradient of the cost functional J with respect to the control variable U.

In the case where U_{ad} has the structure of a linear space, the variational inequality (1.6) is reduced to an equality

$$\nabla J(U^*) = 0. \tag{1.7}$$

The aforementioned 4-D VDA problem (1.1) usually can not be solved analytically. Fortunately standard procedures [108, 144] exist allowing us to solve it. Among

the feasible methods for large-scale unconstrained minimization are (a) the limited memory conjugate gradient method ([144, 173, 174, 189, 190]); (b) quasi-Newton type algorithms ([46], [70], [155]); (c) limited-memory quasi-Newton methods such as LBFGS algorithm ([117], [154]) and (d) truncated-Newton algorithms ([129], [130], [131], [138], [185, 186]).

These procedures are iterative procedures, i.e. they start from an initial guess U_0, find a better approximation U_1 to optimal solution U in the sense of $J(U_1) \leq J(U_0)$, and repeat this minimization process until a prescribed convergence condition is met (see Fig. 1.2). A common requirement of all these procedures is the need to explicitly supply the gradient of the cost function with respect the control variables. The question is therefore: how to numerically determine the gradient $\nabla_U J$? One theoretical possibility would be to evaluate the components of $\nabla_U J$ through finite-difference approximations of the form $\Delta J/\Delta u_j$, where ΔJ is the computed variation of J resulting from a given perturbation Δu_j of the j-th component u_j of U. This method has been put forward by Hoffman [85, 86, 87] for performing variational assimilation. But it requires as many model integrations as there are components in U. Its numerical cost would be totally prohibitive in most situations, especially for NWP problems where the number of the components of the control variables is bigger than 10^5.

With present computer power, the only practical way to implement variational assimilation is through an appropriate use of the so-called adjoint of the assimilation model although it is still computational expensive. The adjoint model of a numerical model basically consists of the equations which govern the temporal evolution of a small perturbation imposed on a model equation, written in a form particularly appropriate for the computation of the sensitivities of the output parameters of the model with respect to input parameters [19, 20, 80, 103, 128, 232]. The use of adjoint models is an application of the theory of optimization and optimal control

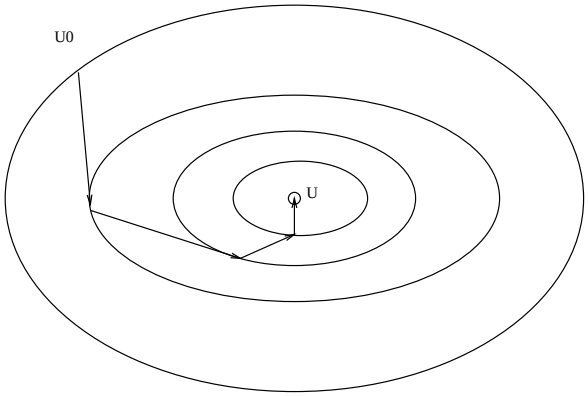

Figure 1.2: Schematic representation of a minimization process, where ellipses are the isolines of the cost function J, $U0$ is the initial guess, U is the optimal solution and each arrow represents an iteration of the minimization process.

of PDEs, which has been developed in the last twenty years by mathematicians [116], and which is progressively being applied to various research disciplines. The principle for deriving adjoint models is based on a systematic use of the chain rule for differentiation.

1.3.2 Description of the first order adjoint analysis

A cost effective way to obtain the gradient of the cost function with respect to the control variables is to integrate the first order adjoint (FOA) model backwards in time from the final time to the initial time of the assimilation window. The theory and application of the FOA model is discussed by several authors, e. g. Talagrand and Courtier [202] and Navon *et al.* [149]. In order to provide a self-contained comprehensive dissertation and develop the second order adjoint (SOA) theory, we briefly summarize the theory of the FOA model.

Let us consider a perturbation, U', on the initial condition U, Eqs (1.4-1.5) become

$$\begin{aligned} & \frac{\partial(\vec{X}+\hat{X})}{\partial t} = F(\vec{X}+\hat{X}), \\ = \;& F(\vec{X}) + \frac{\partial F}{\partial \vec{X}}\hat{X} + \frac{1}{2}\hat{X}^* \frac{\partial F^2}{\partial \vec{X}^2}\hat{X} + O(\|\hat{X}\|^3), \end{aligned} \tag{1.8}$$

$$\vec{X}(t_0) + \hat{X}(t_0) = U + U', \tag{1.9}$$

where $(\cdot)^*$ denotes the complex conjugate transpose, $\|\cdot\|$ denotes the Euclidean norm or other suitably defined norm of "$\cdot$" and $\hat{X}$ is the resulting perturbation of the variable $\vec{X}$. Retaining only the first order term of Eq (1.8), one obtains

$$\frac{\partial \hat{X}}{\partial t} = \frac{\partial F}{\partial \vec{X}}\hat{X}, \tag{1.10}$$

$$\hat{X}(t_0) = U'. \tag{1.11}$$

Eqs. (1.10)-(1.11) define the tangent linear model equations of Eqs. (1.4-1.5) and are linear with respect to the perturbation U' on the initial condition U. Eq. (1.10) is called a tangent linear model because the linearization is around a time-evolving solution, and therefore, the coefficients of the linear model are defined by slopes of the tangent to the nonlinear model trajectory in phase space.

The Gateaux derivative δJ of the cost function in the direction of U' is defined by

$$\delta J = \lim_{\alpha \to 0} \frac{J(U+\alpha U') - J(U)}{\alpha}, \tag{1.12}$$

which can be expressed as

$$\delta J = \nabla_U J \cdot U' \tag{1.13}$$

Eq. (1.13) is, to the first order accuracy, the variation of the distance function J due to the perturbation U' on the initial condition U, i.e.

$$\delta J = \sum_{i=0}^{M} < W(\vec{X}_i - \vec{X}_i^o), \hat{X}_i > . \tag{1.14}$$

Using for example the Euler time differencing scheme one obtains from (1.10)

$$\begin{aligned}\hat{X}_{i+1} &= \hat{X}_i + \Delta t[\frac{\partial F}{\partial \vec{X}}]_i \hat{X}_i \\ &= [I + \Delta t(\frac{\partial F}{\partial \vec{X}})_i]\hat{X}_i \\ &= \prod_{j=0}^{i}[I + \Delta t(\frac{\partial F}{\partial \vec{X}})_j]\hat{X}_0 \end{aligned} \tag{1.15}$$

where Δt is the constant time step, I is the unit matrix operator, $(\partial F/\partial \vec{X})_j$ represents the matrix $\partial F/\partial \vec{X}$ at the j-th time level, and $\prod_{j=0}^{i}$ denotes the product of $i+1$ factors.

Substituting (1.15) into (1.14) and using basic concepts of adjoint operators, we obtain the following expression

$$\begin{aligned}\delta J &= \sum_{i=1}^{M} < \{\prod_{j=0}^{i-1}[I + \Delta t(\frac{\partial F}{\partial \vec{X}})_j]\}^* W(\vec{X}_i - \vec{X}_i^o), \hat{X}_0 > \\ &\quad + < W(\vec{X}_0 - \vec{X}_0^o), \hat{X}_0 > \end{aligned} \tag{1.16}$$

where $(\ )^*$ denotes the adjoint of $(\ )$. On the other hand, we have

$$\delta J =< \nabla_U J, \hat{X}_0 > . \tag{1.17}$$

Equating Eqs. (1.16) and (1.17), one obtains the gradient of the cost function with respect to the initial conditions as

$$\nabla_U J = \sum_{i=0}^{M}\{\prod_{j=0}^{i-1}[I + \Delta t(\frac{\partial F}{\partial \vec{X}})_j]\}^* W(\vec{X}_i - \vec{X}_i^o). \tag{1.18}$$

The i-th term in (1.18) can be obtained by a backwards integration of the following adjoint equation

$$-\frac{\partial P}{\partial t} = (\frac{\partial F}{\partial \vec{X}})^* P, \tag{1.19}$$

from the i-th time step to the initial step, starting from

$$P_i = W(\vec{X}_i - \vec{X}_i^o), \tag{1.20}$$

where P represents the adjoint variables corresponding to $\hat{X}$. It appears that M integrations of the adjoint model, starting from different time steps t_M, t_{M-1} , ..., t_1, are required to obtain the gradient $\nabla_U J$. However, since the adjoint model (1.19) is linear, only one integration from t_M to t_0 of the adjoint equation is required to calculate the gradient of the cost function with respect to the initial conditions.

In summary, the gradient of the cost function with respect to the initial condition U can be obtained by the following procedure:

1. Integrate the forecast model from t_0 to t_M from initial condition (1.5) and store in memory the corresponding sequence of the model states $\vec{X}_i$ (i =0, 1, ..., M);

2. Starting from $P_M = W(\vec{X}_M - \vec{X}^o_M)$, integrate the "$forced$" adjoint equation (1.19) backwards in time from t_M to t_0 where a forcing term $W(\vec{X}_i - \vec{X}^o_i)$ is added to the right-hand-side of (1.19) at the i-th time step whenever an observation is encountered. The final result P_0 is the value of gradient of the cost function with respect to the initial condition,

$$\nabla_U J = P_0. \tag{1.21}$$

It is worth noting that

1. When the observations do not coincide with model grid points, the model solution should be interpolated to the observations, i.e., $C\vec{X} - \vec{X}^o$ should be used instead of $\vec{X} - \vec{X}^o$ in the cost function definition, where the operator C represents the process of interpolating the model solution to space and time locations where observations are available (i.e. C is a projection operator from the space of the model solution $\vec{X}$ to the space of observations).

2. We note that the numerical cost of the adjoint model computation is about the same as the cost of one integration of the tangent linear model, the latter involving a computational cost of at most one integration of the nonlinear model.

A continuous derivation of the gradient of the cost function is provided Chapter 3. In order to provide the readers with some practical examples and some insight about the adjoint theory and the accuracies of the TLM, FOA and SOA models, two examples, one linear and one nonlinear, are presented in Appendices A and B, respectively. Both examples have exact solutions and thus exactly illustrate the adjoint process.

1.4 The theme of my dissertation

This thesis is aimed at (a) carrying out a in-depth investigation of the feasibility of the 4-D VDA applied to realistic situations; (b) achieving an improvement of the existing large-scale unconstrained minimization algorithms. Toward this end, the thesis first develops the SOA theory and a modified version of the truncated Newton (TN) algorithm of Nash [134], secondly develops the SOA model of the shallow-water equations model and finally derives the FOA model of the Florida State University global spectral model (FSUGSM), thirdly conducts a series 4-D VDA experiments and variational nudging data assimilation, and fourthly compares results of the 4-D VNDA and optimal nudging assimilation to assess their ability to retrieve high quality model initial conditions (i.e. physical initialization).

The first part of my thesis (Chapters 2, 3) focuses on the development of the SOA theory and its application. The SOA model serves to study the evolution of the condition number of the Hessian during the course of the minimization since one forward integration of the nonlinear model and the tangent linear model and

one backwards integration in time of the first order adjoint (FOA) model and the SOA system are required to provide the value of Hessian/vector product. This Hessian/vector product may then be used with the power method to obtain the largest and smallest eigenvalues of the Hessian whose dimension is 1083×1083 for our test problem. The dimension of the Hessian will be more than $10^5 \times 10^5$ for 3-D primitive equations models. If the smallest eigenvalues of the Hessian of the cost function with respect to the control variables are positive at each iteration of the VDA minimization process, then the optimal solution of the VDA is unique. This statement is proved to be true for the shallow water equations model (Subsection 2.4.2). The variation of the condition number of the Hessian of the cost function with respect to number of iterations during the minimization process reflects the convergence rate of the minimization. It has been observed [149] that large scale changes occur in the process of minimization during the first 30 iterations, while during the ensuing iterations only small scale features are assimilated. This entails that the condition number of the Hessian of the cost function with respect to the initial conditions experiences a fast change at the beginning of the minimization and then remains almost unchanged during the latter iterations. The condition number can also provide information concerning the error covariance matrix. The rate at which algorithms for computing the best fit to data converge depends on the size of the condition number as well as the distribution of eigenvalues in the spectrum of the Hessian. The inverse of the Hessian can be identified as the covariance matrix that establishes the accuracy to which the model state is determined by the data; the reciprocals of the Hessian's eigenvalues represent the variance of linear combinations of variables determined by the eigenvectors [205]. The above mentioned Hessian/vector product calculation strategy can be efficiently used in the truncated-Newton algorithm of Nash [133], which is referred to as adjoint truncated-Newton algorithm (ATN). The ATN differs from the truncated-Newton algorithm of Nash [134] only in the use of a

more accurate Hessian/vector product for carrying out the large-scale unconstrained optimization required in variational data assimilation. The Hessian/vector product is obtained by solving an optimal control problem of distributed parameters and its accuracy is analyzed and compared with its finite-difference counterpart. The ATN is based on the first and second order adjoint (SOA) techniques allowing to obtain a better approximation to the Newton line search direction for the problem tested here. The ATN is applied here to a limited-area shallow water equations model with model generated data where the initial conditions serve as control variables. We compare the performance of the ATN algorithm with that of the original TN Nash [134] method and the LBFGS method of Liu and Nocedal [117]. Our numerical tests yield results which are twice as fast as these obtained by the TN algorithm both in terms of number of iterations as well as in terms of CPU time. Further the ATN algorithm turns out to be faster than the LBFGS method in terms of CPU time for the problem tested.

The second part of my thesis (Chapters 3, 4) addresses the challenging issues related to 4-D VDA application in NWP problems, i.e. the numerical feasibility of 4-D variational data assimilation applied to complex NWP models. In order to address these questions, I choose the FSUGSM as an experimental model, since it is an operational model developed from the Canadian spectral model of the Recherche en Prevision Numerique [40] by a research group led by Krishnamurti at Florida State University. In the development of the FSUGSM from the Canadian model, the primary effort has been directed at improving the physical effect parameterizations, adapting the model code to run efficiently on the NCAR Cray-1 computer and developing post-forecast diagnostics [162]. The FSUGSM is a very complex model. Developing its first order adjoint model and tuning the minimization algorithm to perform adequately with a control variable vector of dimension 303104 constitute a challenging task.

In this part of the thesis the tangent linear model (TLM) is first developed and then the the first order adjoint model (FOA) of the FSUGSM is developed by writing the TLM backwards in time subroutine by subroutine. The accuracy of the TLM is investigated by comparing its results with those produced by identical perturbations introduced in nonlinear forecasts of the original model. The accuracy of the FOA is verified by using a Taylor expansion.

Through different initial conditions and different scenarios of sets of observations distributed in time, we demonstrate the efficiency of the 4-D VDA in extracting the information contained in the dynamics of the model together with the information contained in the observations. The conditioning is an important factor for an operational implementation. The dynamics of the model may lead to different reductions of rms errors at different times and the loss of the conditioning of the problem, i.e. the shape of the cost function can become strongly elliptic with respect to the non optimal metric used and the gradient can even become almost orthogonal to the direction of the minimum. The solution to this is a knowledge of the structure of the Hessian of the cost function. It seems that the SOA application may constitute an efficient way to to obtain additional information about the Hessian of the cost function.

As mentioned before, an implicit assumption made in 4-D VDA is that the model exactly represents the state of the atmosphere. However, this is clearly not true since any model only approximately represents the state of the atmosphere. The nudging data assimilation (NDA) addresses the imperfectness of the model by adding a non-physical diffusion-type term to the model equations and thus relaxes the model state towards the observations during the assimilation period (hence an equivalent name for it is Newtonian relaxation [98, 99]). The NDA method is a flexible assimilation technique which is computationally much more economical than the VDA method. However, results from NDA are quite sensitive to the *ad hoc* specification of the

nudging relaxation coefficient, and it is not at all clear how to choose a nudging coefficient so as to obtain an optimal solution.

Variational nudging data assimilation (VNDA) combines the aforementioned data assimilation schemes in the most efficient way. It takes both the initial conditions and the nudging coefficients as control variables and finds the best initial conditions and optimal nudging coefficients iteratively. In Chapter 6, the VNDA is applied, in the framework of 4-dimensional variational data assimilation (VDA), to a T42 version of the FSU global spectral model (FSUGSM) in its full-physics operational form with 12 vertical layers. The technique is tested for its ability to obtain the best initial conditions and optimal nudging coefficients. Results from the VNDA indicate significantly better results and a faster convergence rate compared with the VDA for our test problem when a nudging term is added to the model equations and the nudging coefficients are optimally estimated using VDA in a parameter estimation mode. The humidity and all the different parameterizations of the physical processes are not included in the adjoint model integration. Thus the calculation of the gradients by the adjoint model is approximate since the forecast model is used in its full-physics operational form. It is important to note that the approximate gradient obtained from a simplified adjoint model is used in the experiments. It is shown that it is possible to perform variational data assimilation of realistic forecast models even before more complex adjoint models including diabatic processes are developed. The resulting optimal nudging coefficients are then applied in nudging data assimilation (thus the term optimal nudging data assimilation (ONDA)) [229]. Results of data-assimilation experiments involving estimated nudging assimilation, ONDA and VDA are compared for their ability to retrieve high quality model initial conditions (physical initialization).

Conclusions summarizing the main results of this thesis, their implications, limitation and outstanding problems are presented in the last Chapter.

CHAPTER 2

SECOND ORDER ADJOINT ANALYSIS: THEORY AND APPLICATIONS

2.1 Introduction

The complete description of the initial atmospheric state in a numerical weather prediction method constitutes an important issue. The four dimensional variational data assimilation (VDA) method offers a promising way to obtain such a description of the atmosphere. It consists in finding the assimilating model solution which minimizes a properly chosen objective function measuring the distance between model solution and available observations distributed in space and time. The control variables are either the initial conditions or the initial conditions plus the boundary conditions. The boundary conditions have to be specified so that the problem is well posed in the sense of Hadamard. In most unconstrained minimization algorithms associated with the VDA approach, the gradient of the objective function with respect to the control variables plays an essential role. This gradient is obtained through one direct integration of the nonlinear model equations followed by a backwards integration in time of the linear adjoint system of the direct model.

The knowledge of the structure of the Hessian of the cost function with respect to the control variables is useful in many ways. For instance, the positive definiteness of the Hessian implies the uniqueness of the solution, the Hessian/vector production can be used implement truncated-Newton algorithms [134] and the eigenvalue distribution and eigenvalue structure of the Hessian can provide useful information for preconditioning unconstrained minimization algorithms [187]. In this chapter I

will develop the SOA theory to study the evolution of the condition number of the Hessian during the course of the minimization and to conduct the sensitivity analysis of the cost function with respect to distributed observations.

The context of this chapter is outlined as follows: the theory of the SOA is introduced in Section 2.2. In Section 2.3, we present a detailed derivation of the SOA model of a two dimensional shallow water equations model. Quality control methods for the verification of the correctness of the SOA model are discussed in Subsection 2.3.1. Issues concerning uniqueness of the solution and the evolution of the condition number of the Hessian during the course of the minimization as well as issues related to the structure of the retrieved initial conditions are addressed in Section 2.4. Section 2.5 is devoted to a sensitivity study of the solution with respect to distributed inaccurate observations. Finally conclusions are presented in Section 2.6.

2.2 The second order adjoint model

2.2.1 Theory of the second order adjoint model

The forward and backwards integrations of the nonlinear model and its adjoint model respectively, provide the value of the cost function J and its gradient. The following question may then be posed: can we obtain any information about the Hessian (second order derivative matrix) of the cost function with respect to the initial conditions by integrating the adjoint model equations? Once the Hessian/vector product is available, the condition number of the Hessian may be obtained. This condition number may then be used to study the convergence rate of the minimization algorithms used in of VDA. We will show in this section that one integration of the SOA model yields a Hessian/vector product or a column of the Hessian of the cost function with respect to the initial conditions. Therefore, the SOA model

provides an efficient way to compute the Hessian of the cost function by performing M integrations of the SOA model where M is the number of the components in the control variables. For a large dimensional model, obtaining the full Hessian matrix proves to be a computationally prohibitive task beyond the capability of present day computers. The SOA approach will be used to conduct a sensitivity analysis of the observations in Section 2.5. We will also study the relative importance of observations distributed at different space and time locations.

The original idea of the SOA theory was proposed by Le Dimet (personal communication). Here we developed the SOA theory in a clearer and simple manner.

Assume that the model equations are given by Eqs. (1.4) and (1.5). Let us define the cost function as

$$J(U) = \frac{1}{2}\int_{t_0}^{T} < W(C\vec{X} - \vec{X}^o), C\vec{X} - \vec{X}^o > dt, \tag{2.1}$$

where W is a weighting matrix often taken to be the inverse of the estimate of the covariance matrix of the observation errors, T is the final time of the assimilation window, the objective function $J(U)$ is the weighted sum of squares of the distance between model solution and available observations distributed in space and time, $\vec{X}^o$ is observation vector, the projection operator C represents the process of interpolating the model solution $\vec{X}$ to space and time locations where observations are available. The purpose is to find the initial conditions such that the solution of the Eq. (1.4) minimizes the cost function $J(U)$ in a least-squares sense. The FOA model as defined by Eqs. (1.19)-(1.20) may then be rewritten as

$$-\frac{\partial P}{\partial t} = (\frac{\partial F}{\partial \vec{X}})^* P + C^* W(C\vec{X} - \vec{X}^o), \tag{2.2}$$

$$P(T) = 0. \tag{2.3}$$

where P represents the FOA variables vector. The gradient of the cost function with respect to the initial conditions is given by (see Subsection 3.2.1)

$$\nabla_U J = P(t_0), \tag{2.4}$$

Let us now consider a perturbation, U', on the initial condition U. The resulting perturbations for the variables $\vec{X}$, P, $\hat{X}$ and $\hat{P}$ may be obtained from Eqs. (1.4)-(1.5) and (2.2)-(2.3) as

$$\frac{\partial \hat{X}}{\partial t} = \frac{\partial F}{\partial \vec{X}} \hat{X}, \tag{2.5}$$

$$\hat{X}(0) = U', \tag{2.6}$$

$$-\frac{\partial \hat{P}}{\partial t} = (\frac{\partial F}{\partial \vec{X}})^* \hat{P} + [\frac{\partial^2 F}{\partial \vec{X}^2} \hat{X}]^* P + C^* W C \hat{X}, \tag{2.7}$$

$$\hat{P}(T) = 0, \tag{2.8}$$

Eqs. (2.5)-(2.6) and Eqs. (2.7) and (2.8) are called the tangent linear and SOA models, respectively.

Let us denote the FOA variable after a perturbation U' on the initial condition U by $P_{U+U'}$, then according to definition

$$P_{U+U'}(t_0) = P(t_0) + \hat{P}(t_0). \tag{2.9}$$

Expanding $\nabla_{U+U'} J$ at U in a Taylor series and only retaining the first order term, results in

$$\nabla_{U+U'} J = \nabla_U J + \nabla^2 J \cdot U' + O(\|U'\|^2) \tag{2.10}$$

From Eq. (2.4), we know that

$$\nabla_{U+U'} J = P_{U+U'}(t_0), \tag{2.11}$$

Combining Eqs. (2.4), (2.9), (2.10) and (2.11), one obtains

$$\hat{P}(t_0) = \nabla^2 J \cdot U' = HU', \tag{2.12}$$

where $H = \nabla^2 J$ is the second derivative of the cost function with respect to initial conditions.

If we set $U' = e_j$, where e_j is the unit vector with the j-th element equal to 1, the j-th column of the Hessian may be obtained by

$$He_j = \hat{P}(t_0) \tag{2.13}$$

Therefore theoretically speaking, the full Hessian H can be obtained by M integrations of Eqs. (2.7)-(2.8) with $U' = e_i, i = 1, ..., M$ where M is the number of the components of the control variables (the initial conditions $u(t_0)$, $v(t_0)$ and $\phi(t_0)$ in our case).

In summary, the j-th column of the Hessian of the cost function can be obtained by the following procedure (also see Fig. 2.1):

1. Integrate the model (1.4)-(1.5) and the tangent linear model (2.5)-(2.6) forward and store in memory the corresponding sequences of the states $\vec{X}_i$ and $\hat{X}_i$ (i =0, ..., M);

2. Integrate the FOA equations (2.2)-(2.3) backwards in time and store in memory the sequence of P_i (i =0, ..., M);

3. Integrate the SOA model (2.7)-(2.8) backwards in time. The final value $\hat{P}(t_0)$, yields the j-th column of the Hessian of the cost function with respect to the control variables.

2.2.2 The estimate of the condition number of the Hessian

Let us denote the largest and the smallest eigenvalues of the Hessian of the cost function with respect to the control variables and their corresponding eigenvectors by $\lambda_{\max}, \lambda_{\min}, V_{\max}$ and $V_{\min}$, respectively. Then the condition number of the Hessian

is given by

$$\kappa(H) = \frac{\lambda_{\max}}{\lambda_{\min}} \tag{2.14}$$

Considering the eigenvalue problem $HU = \lambda U$ and assuming that the eigenvalues are ordered in decreasing order with $|\lambda_1| \geq |\lambda_2| \geq \ldots\ldots \geq |\lambda_n|$, an arbitrary initial vector X_0 may be expressed as a linear combination of the eigenvectors $\{U_i\}$

$$X_0 = \sum_{i=1}^{n} c_i U_i \tag{2.15}$$

If λ_i is an eigenvalue corresponding to the i-th eigenvector U_i, m times multiplications of the Hessian H to (2.15) result in,

$$X_m = \sum_{i=1}^{n} c_i \lambda_i^m U_i, \tag{2.16}$$

where

$$X_m = H^m X_0. \tag{2.17}$$

Factoring λ_1^m out, we obtain

$$X_m = \lambda_1^m \sum_{i=1}^{n} c_i (\frac{\lambda_i}{\lambda_1})^m U_i. \tag{2.18}$$

Since λ_1 is the largest eigenvalue, the ratio $(\lambda_i/\lambda_1)^m$ approaches zero as m increases (suppose $\lambda_1 \neq \lambda_2$). Therefore we may write

$$X_m = \lambda_1^m c_1 U_1, \tag{2.19}$$

From (2.19) we observe that the largest eigenvalue may be calculated by

$$\lambda_1 = \frac{j\text{th component of } X_{m+1}}{j\text{th component of } X_m} \tag{2.20}$$

This technique is called the power method [199]. We can normalize the vector X_m by its largest component in absolute value. If we denote the new scaled iterate by X'_m, then

$$X_{m+1} = H X'_m, \tag{2.21}$$

and the method is called the power method with scaling. It yields an eigenvector whose largest component is 1.

The main steps in the power method algorithm with scaling are:

1. Generate a starting vector X_0.

2. Form a matrix power sequence $X_m = HX_{m-1}$.

3. Normalize X_m so that its largest component is unity.

4. Return to step (b) until convergence

$$|X_m - X_{m-1}| \leq 10^{-6}, \tag{2.22}$$

is satisfied or a prescribed upper limit of the number of iterations has been attained.

The smallest eigenvalue of H may also be computed by applying the shifted iterated power method to the matrix $Z = z \cdot I - H$, where z is the majorant of the spectral radius of H and I the identity matrix.

2.3 The derivation of the second order adjoint model for the shallow water equations model

In this section, we consider the application of the SOA model to a two dimensional limited-area shallow water equations model. Our purpose is to illustrate how to derive the SOA model explicitly.

The shallow water equations model may be written as

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} + fv - \frac{\partial \phi}{\partial x}, \tag{2.23}$$

$$\frac{\partial v}{\partial t} = -u\frac{\partial v}{\partial x} - v\frac{\partial v}{\partial y} - fu - \frac{\partial \phi}{\partial y}, \tag{2.24}$$

$$\frac{\partial \phi}{\partial t} = -\frac{\partial (u\phi)}{\partial x} - \frac{\partial (v\phi)}{\partial y}, \tag{2.25}$$

where u, v, ϕ and f are the two components of the horizontal velocity and geopotential fields and the Coriolis factor, respectively.

We shall use initial conditions due to Grammeltvedt [75]

$$h = H_0 + H_1 tanh\frac{9(y-y_0)}{2D} + H_2 sech\frac{9(y-y_0)}{D}\sin\frac{2\pi x}{L}, \tag{2.26}$$

where $H_0 = 2000m$, $H_1 = -220m$, $H_2 = 133m$, $g = 10msec^{-2}$, $L = 6000km$, $D = 4400km$, $f = 10^{-4}sec^{-1}$, $\beta = 1.5 \times 10^{-11}sec^{-1}m^{-1}$. Here L is the length of the channel on the β plane, D is the width of the channel and $y_0 = D/2$ is the middle of the channel. The initial velocity fields were derived from the initial height field via the geostrophic relationship, and are given by

$$u = -\frac{g}{f}\frac{\partial h}{\partial y}, \tag{2.27}$$

$$v = \frac{g}{f}\frac{\partial h}{\partial x}. \tag{2.28}$$

The time and space increments used in the model are

$$\Delta x = 300km,\ \Delta y = 220km,\ \Delta t = 600s, \tag{2.29}$$

which mean that there are 21×21 grid point locations in the channel and the number of the components of initial condition vector $(u, v, \phi)^t$ is 1083. Therefore the Hessian of the cost function in our test problem has a dimension of 1083×1083.

The southern and north boundaries are rigid walls where the normal velocity components vanish, and it is assumed that the flow is periodic in the west-east direction with a wave length equal to the length of the channel.

Let us define

$$\vec{X} = (u, v, \phi)^T, \tag{2.30}$$

$$F = -\begin{pmatrix} u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + \frac{\partial \phi}{\partial x} \\ u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + \frac{\partial \phi}{\partial y} \\ \frac{\partial (u\phi)}{\partial x} + \frac{\partial (v\phi)}{\partial y} \end{pmatrix}. \tag{2.31}$$

Then Eqs. (2.23)-(2.25) assume the form of Eq. (1.4). It is easy to verify that

$$\frac{\partial F}{\partial \vec{X}} = -\begin{pmatrix} \frac{\partial (u(\cdot))}{\partial x} + v\frac{\partial (\cdot)}{\partial y} & (\cdot)\frac{\partial u}{\partial y} - f(\cdot) & \frac{\partial (\cdot)}{\partial x} \\ (\cdot)\frac{\partial v}{\partial x} + f(\cdot) & u\frac{\partial (\cdot)}{\partial x} + \frac{\partial (v(\cdot))}{\partial y} & \frac{\partial (\cdot)}{\partial y} \\ \frac{\partial (\phi(\cdot))}{\partial x} & \frac{\partial (\phi(\cdot))}{\partial y} & \frac{\partial (u(\cdot))}{\partial x} + \frac{\partial (v(\cdot))}{\partial y} \end{pmatrix}. \tag{2.32}$$

The adjoint of an operator L, L^* is defined by the relationship

$$< L\vec{X}, \vec{Y} >=< \vec{X}, L^*\vec{Y} >, \tag{2.33}$$

where $< \cdot, \cdot >$ denotes the inner product

$$< \cdot, \cdot >= \int\int_D \cdot \cdot \, dD, \tag{2.34}$$

where D is the spatial domain. Using the definition (2.34), the adjoint of (2.33) can be derived as

$$[\frac{\partial F}{\partial \vec{X}}]^* = -\begin{pmatrix} -u\frac{\partial (\cdot)}{\partial x} - \frac{\partial (v(\cdot))}{\partial y} & (\cdot)\frac{\partial v}{\partial x} + f(\cdot) & -\phi\frac{\partial (\cdot)}{\partial x} \\ (\cdot)\frac{\partial u}{\partial y} - f(\cdot) & -v\frac{\partial (\cdot)}{\partial y} - \frac{\partial (u(\cdot))}{\partial x} & -\phi\frac{\partial (\cdot)}{\partial y} \\ -\frac{\partial (\cdot)}{\partial x} & -\frac{\partial (\cdot)}{\partial y} & -u\frac{\partial (\cdot)}{\partial x} - v\frac{\partial (\cdot)}{\partial y} \end{pmatrix}. \tag{2.35}$$

Therefore the first order adjoint model with the forcing terms may be written as

$$-\frac{\partial u^*}{\partial t} = -(-u\frac{\partial u^*}{\partial x} - \frac{\partial (vu^*)}{\partial y} + v^*\frac{\partial v}{\partial x} + fv^* - \phi\frac{\partial \phi^*}{\partial x}) + W_u(u - u^o), \tag{2.36}$$

$$-\frac{\partial v^*}{\partial t} = -(u^*\frac{\partial u}{\partial y} - fu^* - v\frac{\partial v^*}{\partial y} - \frac{\partial (uv^*)}{\partial x} - \phi\frac{\partial \phi^*}{\partial y}) + W_v(v - v^o), \tag{2.37}$$

$$-\frac{\partial \phi^*}{\partial t} = -(-\frac{\partial u^*}{\partial x} - \frac{\partial v^*}{\partial y} - u\frac{\partial \phi^*}{\partial x} - v\frac{\partial \phi^*}{\partial y}) + W_\phi(\phi - \phi^o), \tag{2.38}$$

with final conditions

$$u(T) = 0,\ v(T) = 0,\ \phi(T) = 0, \tag{2.39}$$

where $P = (u^*, v^*, \phi^*)^t$ is the first order adjoint variable. W_u, W_v, W_ϕ are weighting factors which are taken to be the inverse of estimates of the statistical root-mean-square observational errors on geopotential and wind components respectively. In our test problem, values of $W_\phi = 10^{-4}m^{-4}s^4$ and $W_u = W_v = 10^{-2}m^{-2}s^2$ are used.

Now let us consider a perturbation, U', on the initial condition for $\vec{X}$, $\vec{X}(t_0)$. The resulting corresponding perturbations for variables $\vec{X}$ and P, $\hat{X} = (\hat{u}, \hat{v}, \hat{\phi})^t$ and $\bar{P} = (\bar{u}, \bar{v}, \bar{\phi})^t$, are obtained from Eqs. (2.23)-(2.25) and (2.36)-(2.39) as

$$\frac{\partial \hat{u}}{\partial t} = -(\frac{\partial (u\hat{u})}{\partial x} + v\frac{\partial \hat{u}}{\partial y} + \hat{v}\frac{\partial u}{\partial y} - f\hat{v} + \frac{\partial \hat{\phi}}{\partial x}), \tag{2.40}$$

$$\frac{\partial \hat{v}}{\partial t} = -(\hat{u}\frac{\partial v}{\partial x} + f\hat{u} + u\frac{\partial \hat{v}}{\partial x} + \frac{\partial (v\hat{v})}{\partial y} + \frac{\partial \hat{\phi}}{\partial y}), \tag{2.41}$$

$$\frac{\partial \hat{\phi}}{\partial t} = -(\frac{\partial (\phi\hat{u})}{\partial x} + \frac{\partial (\phi\hat{v})}{\partial y} + \frac{\partial (u\hat{\phi})}{\partial x} + \frac{\partial (v\hat{\phi})}{\partial y}), \tag{2.42}$$

with zero initial conditions, and

$$\begin{aligned} -\frac{\partial \bar{u}}{\partial t} &= -(-u\frac{\partial \bar{u}}{\partial x} - \frac{\partial (v\bar{u})}{\partial y} + \bar{v}\frac{\partial v}{\partial x} - \phi\frac{\partial \bar{\phi}}{\partial x} - \\ &\quad \hat{u}\frac{\partial u^*}{\partial x} - \frac{\partial (\hat{v}u^*)}{\partial y} + v^*\frac{\partial \hat{v}}{\partial x} - \hat{\phi}\frac{\partial \phi^*}{\partial x}) + W_u\hat{u}, \end{aligned} \tag{2.43}$$

$$\begin{aligned} -\frac{\partial \bar{v}}{\partial t} &= -(\bar{u}\frac{\partial u}{\partial y} - v\frac{\partial \bar{v}}{\partial y} - \frac{\partial (u\bar{v})}{\partial x} - \phi\frac{\partial \bar{\phi}}{\partial y} + \\ &\quad u^*\frac{\partial \hat{u}}{\partial y} - \hat{v}\frac{\partial v^*}{\partial y} - \frac{\partial (\hat{u}v^*)}{\partial x} - \hat{\phi}\frac{\partial \phi^*}{\partial y}) + W_v\hat{v}, \end{aligned} \tag{2.44}$$

$$\begin{aligned} -\frac{\partial \bar{\phi}}{\partial t} &= -(-\frac{\partial \bar{u}}{\partial x} - \frac{\partial \bar{v}}{\partial y} - u\frac{\partial \bar{\phi}}{\partial x} - v\frac{\partial \bar{\phi}}{\partial y} - \\ &\quad \hat{u}\frac{\partial \phi^*}{\partial x} - \hat{v}\frac{\partial \phi^*}{\partial y}) + W_\phi\hat{\phi}, \end{aligned} \tag{2.45}$$

with final condition

$$\bar{u}(T) = 0,\ \bar{v}(T) = 0,\ \bar{\phi}(T) = 0. \tag{2.46}$$

Therefore

$$\bar{P}(t_0) = (\bar{u}(t_0), \bar{v}(t_0), \bar{\phi}(t_0))^t = HU', \tag{2.47}$$

where H is the Hessian of the cost function with respect to the initial conditions (the control variables). Equation (2.47) gives the Hessian/vector product. If we choose

U' to be unit vector e_j where the j-th component is unity and all other components are zeros, then the corresponding column H_j of the Hessian H will be obtained after one integration of the SOA backwards in time.

2.3.1 The verification of the correctness of the first and the second order adjoint models

It is very important to verify the correctness of the FOA and SOA codes. A Taylor expansion in the $\vec{Y}$ direction leads to

$$J(\vec{X}+\alpha\vec{Y}) = J(\vec{X}) + \alpha\frac{\partial J(\vec{X})}{\partial \vec{X}}\vec{Y} + \frac{1}{2}\alpha^2\vec{Y}^t\frac{\partial^2 J(\vec{X})}{\partial \vec{X}^2}\vec{Y} + O(\alpha^3), \tag{2.48}$$

where α is a small scalar, $\vec{Y}$ is a random perturbation vector which can be generated by using the randomizer on the Cray-YMP computer and $\vec{Y}^t$ denotes the transpose of the vector $\vec{Y}$. Equation (B.1) can be used to define two functions of α

$$\psi(\alpha) = \frac{J(\vec{X}+\alpha\vec{Y}) - J(\vec{X})}{\alpha\frac{\partial J(\vec{X})}{\partial \vec{X}}\vec{Y}}, \tag{2.49}$$

and

$$\phi(\alpha) = \frac{J(\vec{X}+\alpha\vec{Y}) - J(\vec{X}) - \alpha\frac{\partial J(\vec{X})}{\partial \vec{X}}\vec{Y}}{\frac{1}{2}\alpha^2\vec{Y}^t\frac{\partial^2 J(\vec{X})}{\partial \vec{X}^2}\vec{Y}}, \tag{2.50}$$

then for small α we have

$$\psi(\alpha) = 1 + O(\alpha), \tag{2.51}$$

$$\phi(\alpha) = 1 + O(\alpha), \tag{2.52}$$

For values of α which are small but not very close to the machine zero, one should expect a value of $\psi(\alpha)$ or $\phi(\alpha)$ approaching 1 linearly for a wide range of magnitudes of α.

The experiment was performed using a limited area 2-D shallow water equation model. The results are shown in Figure 2.2. It is clearly seen that for values of α between 10^0-10^{-11} unit values for $\psi(\alpha)$ and $\phi(\alpha)$ are obtained. The correctness of

the gradient of the cost function and the correctness of the Hessian/vector product have therefore been verified.

2.4 Second order adjoint information

2.4.1 Calculation of the Hessian/vector product

There are two practical ways to calculate the Hessian/vector product at a point $\vec{X}$ associated with VDA. One way consists in using a finite difference method while the other way is the SOA method. The finite-difference approach assumes the following form (namely a difference between two adjacent gradient values)

$$f(\alpha) = \nabla J(\vec{X} + \alpha\vec{Y}) - \nabla J(\vec{X}) = \alpha H\vec{Y} + O(\alpha^2), \tag{2.53}$$

where $\vec{Y}$ is a random perturbation vector and H is the Hessian of the cost function. A second way to obtain Hessian/vector product is to integrate the SOA equations model backwards in time. According to Eq. (2.12), we also have

$$f(\alpha) = \alpha H\vec{Y}, \tag{2.54}$$

The computational cost required to obtain the Hessian/vector product is approximately the same for both methods. The SOA approach requires us to integrate the original nonlinear model and its tangent linear model forward in time once and integrate the FOA model and the SOA model backwards in time once. The finite difference approach requires the integration of the original nonlinear model forward in time and the FOA model backwards in time twice. The computational costs for integrating the tangent linear model forward in time, or the FOA model backwards in time and the SOA model backwards in time once are comparable. However the SOA model method gives an accurate value of the Hessian/vector product while the finite-difference method yields only an approximate value, which can be a very poor

estimate when the value α is not properly chosen. Figures 2.3, 2.4, 2.5 presents a comparison between the first 50 components of Hessian/vector products at the *optimal point* obtained by using both the SOA and finite-difference approaches for various scalars α varying from 10, 3 to 0.01. It is clearly seen that the Hessian/vector product obtained by using finite-difference approach converges to that obtained by SOA as the scalar α decreases. With the SOA approach a more accurate result can be obtained with a relatively large perturbation ($\alpha = 10$), while the finite-difference approach is much more sensitive to the magnitude of perturbations. When the perturbations are large, say for $\alpha = 10$, the finite-differencing approach yields no meaningful results (Figure 2.3). When the perturbations are small, the finite-difference approach might involve a subtraction of nearly equal numbers which results in the cancellation of significant digits and the results thus obtained are an inaccurate estimate of the Hessian/vector product. This is the case when the Hessian/vector product is estimated at the *initial guess point* with $\alpha = 0.01$ (Figure 2.6). Therefore it is much more advantageous to use the SOA approach than to use the finite-difference approach.

The calculation of a Hessian/vector product is required in many occurrences. For instance, Nash's Truncated Newton method [130] requires the values of Hessian/vector products. It may also be used to carry out eigenvalue calculations and sensitivity analyses.

2.4.2 The uniqueness of the solution

An important issue related to VDA application is to determine whether the solution obtained is unique. If there is more than one local minimum, then the solution obtained by the minimization process may possibly change depending on different initial guesses.

There are two different but complementary ways to characterize the solution to unconstrained optimization problems. In the local approach, one examines the rela-

tion of a given point to its neighbour. The conclusion is that at an unconstrained relative minimum point of a smooth cost function, the gradient of the cost function vanishes and the Hessian is positive semidefinite; and conversely, if at a point the gradient vanishes and the Hessian is positive definite, that point is a relative minimum point. This characterization has a natural extension to the global approach where convexity ensures that if the gradient vanishes at a point, that point is global minimum point.

The Hessian (the second order derivative of cost function with respect to the control variables) is the generalization to E^n of the concept of the curvature of function, and correspondingly, positive definiteness of the Hessian is the generalization of positive curvature. We sometimes refer to a function as being locally strictly convex if the Hessian is positive definite in the region. In these terms we see that the second order sufficiency result requires that the function be locally strictly convex at the point X^*.

A simple experiment was conducted to find out about the uniqueness of the cost function with respect to the initial conditions using the shallow-water equations model. The experiment is devised as follows: the model-generated values starting from the initial condition of Grammeltvedt (Eq. (2.26)) are used as observations, the initial guess is a randomly perturbed Grammeltvedt initial condition, and the length of the assimilation is 10 hours. we know exactly what the solution is, and the value of the cost function at the minimum must be zero. All the random perturbations used in this chapter are from a uniform distribution. The limited memory quasi-Newton method of of Liu and Nocedal [117] is used for all experiments in this chapter.

The symmetric version of the power and shifted power methods are used to obtain the largest and smallest eigenvalues of the Hessian at each iteration. The results are shown in Figures 2.7 and 2.8. The smallest eigenvalues at each iteration of the minimization process are small positive numbers. The positiveness of the smallest

eigenvalues implies positive definiteness of the Hessian, which in turn proves the uniqueness of the optimal solution.

2.4.3 Convergence analysis

The largest and smallest eigenvalues and the condition numbers are considered here. The purpose of this study is to provide an in-depth diagnosis of the convergence of the VDA applied to a meteorological problem. The various scale changes of different field retrievals with the number of minimization iterations of VDA have attracted the attention of several researchers [149]. In this research work we will attempt to provide an explanation of this phenomenon based on the evolution of the condition number of the Hessian of the cost function with respect to control variables [205]. It has been observed that in VDA large scale changes occur in the first few iterations and small scale changes occur during the latter iterations in the process of the minimization of the cost function.

The same experiment as described in section 4.2 was conducted again this time to follow the quality of the retrievals initial conditions at different stages of the minimization process. Figures 2.9-2.11 show the perturbed geopotential field and the retrieved geopotential fields after 6 and 25 iterations, respectively. It can be clearly seen that most of the large scale retrievals occur within the first 25 iterations of the minimization process. The geopotential field retrieved after 25 iterations is very similar to the one retrieved after 54 iterations at which stage the prescribed convergence criterion

$$\|\nabla J(\vec{X}_k)\| \leq 10^{-14} \times max\{1, \|\vec{X}_k\|\}, \tag{2.55}$$

is satisfied. This clearly indicates that the VDA achieves most of the large scale retrievals during the first 25 iterations and that in the latter part of the minimization process only small scale features are being assimilated. In this case by stopping the minimization process prior to the cost function satisfying the preset convergence

criteria, the expensive computational cost of the VDA process could be cut by more than a half, while satisfactory results may still be obtained.

This in turn is related to the evolution of the largest and smallest eigenvalues of the Hessian spectrum and thus to the change in the condition number of the Hessian with the number of iterations (Figures 2.7, 2.8 and 2.12). From these figures we observe

1. The smallest eigenvalues are positive at each iteration and remain approximately the same except for rather small changes during the first a few iterations (Figure 2.7).

2. The largest eigenvalues decrease quickly during the first a few iterations of the minimization process, then change only slightly for the next 15 iterations and finally remain approximately the same until the convergence criteria are attained (Figure 2.8).

3. The condition numbers of the Hessian/vector product at different steps of the minimization vary in a way similar to the evolution of the largest eigenvalues during the minimization process and they are around 83,000 in magnitude which is very large (Figure 2.12).

We see now that most changes in the condition numbers occur during the early stage of the VDA minimization process. This explains why large scale retrievals occur during the first few (25-30) iterations of the minimization process.

The large condition numbers in the initial stage of the minimization imply that the contour lines of the cost function $J = constant$ are strongly elongated in the parameter space [62, 73, 116, 123], which explains the slow convergence rate of the VDA process. The above experiment was carried out without adding either a penalty or a smoothing term. The addition of such a penalty term, which is positive definite

and quadratic with respect to the initial conditions, will definitely increase the convexity of the cost function. Thus the addition of an adequate quadratic penalty term adding additional information to the cost function changes the condition number of the Hessian and speeds up the convergence of the VDA process.

2.5 Sensitivity analysis for observations

2.5.1 Sensitivity theory for observation

The cost function is also a function of the observations. Different observations will result in different solutions. Due to the errors inherent in the heterogeneous observations, it is important to obtain the sensitivities of the cost function to changes in the observations which quantify to what extent the perturbations in the observations correspond to the perturbation in the solution. If the sensitivities are large, then the model will possess a large uncertainty with respect to changes in the observations.

Conventional evaluation of the sensitivities with respect to model parameters is carried out by changing the values of model parameters and recalculating each model solution for every parameter. Such a calculation is prohibitive for models with a large number of parameters since it requires an exceedingly large amount of computing time. The adjoint sensitivity method [19, 20, 78, 79, 200] has proved to be an efficient way for carrying out sensitivity analysis. The objective of the sensitivity analysis considered here is to estimate changes in the cost function J arising from changes in observations which are distributed in space and time. This will illustrate the relative importance of observations at different time and space locations.

Assume that the operator C in Eq. (2.1) is an unit operator. Due to the equivalent position of the state vector and the observation vector in Eq. (2.1), the cost function can be viewed as depending on both of them, namely

$$J = J(\vec{X} - \vec{X}^o), \tag{2.56}$$

As such, the following identities can be proved using the chain rule:

$$\frac{\partial J}{\partial \vec{X}} = -\frac{\partial J}{\partial \vec{X}^o}, \tag{2.57}$$

$$\frac{\partial^2 J}{\partial \vec{X}^2} = \frac{\partial^2 J}{\partial \vec{X}^{o\,2}}, \tag{2.58}$$

These two equations are used in the following sensitivity analysis.

Let us denote a change in the observations by $\delta\vec{X}^o$. If this change is small, then we may expand the cost function J around $\vec{X}^o$ in a Taylor series as

$$\begin{aligned} J(\vec{X}^o(t_n) \quad + \quad \delta\vec{X}^o(t_n)) = J(\vec{X}^o(t_n)) + \frac{\partial J(\vec{X}^o(t_n))}{\partial \vec{X}^o(t_n)}\delta\vec{X}^o(t_n) \\ + \frac{1}{2}\delta\vec{X}^o(t_n)^t \frac{\partial^2 J(\vec{X}^o(t_n))}{\partial \vec{X}^o(t_n)^2}\delta\vec{X}^o(t_n) + O(\|\delta\vec{X}^o(t_n)\|^3), \end{aligned} \tag{2.59}$$

According to the identities given by Eqs. (2.57) and (2.58), Eq. (2.59) can be written as

$$\begin{aligned} J(\vec{X}^o(t_n) \quad + \quad \delta\vec{X}^o(t_n)) = J(\vec{X}^o(t_n)) - \frac{\partial J(\vec{X}^o(t_n))}{\partial \vec{X}(t_n)}\delta\vec{X}^o \\ + \frac{1}{2}\delta\vec{X}^o(t_n)^t \frac{\partial^2 J(\vec{X}^o(t_n))}{\partial \vec{X}(t_n)^2}\delta\vec{X}^o + O(\|\delta\vec{X}^o(t_n)\|^3), \end{aligned} \tag{2.60}$$

where t_n denotes the time, $t_n = t_0 + n\Delta t$ and Δt is given by Eq. (2.29). Since the first order term in Eq. (2.60) dominates, we obtain from Eq. (2.60)

$$\begin{aligned} J' &= J(\vec{X}^o(t_n) + \delta\vec{X}^o(t_n)) - J(\vec{X}^o(t_n)) \\ &= -\frac{\partial J(\vec{X}^o(t_n))}{\partial \vec{X}(t_n)}\delta\vec{X}^o + O(\|\delta\vec{X}^o(t_n)\|^2). \end{aligned} \tag{2.61}$$

This equation describes changes in the cost function resulting from changes in the observation at time t_n.

If the gradient of the cost function with respect to the state vector $X(t_n)$ is zero, then we obtain

$$\begin{aligned} J' &= J(\vec{X}^o(t_n) + \delta\vec{X}^o(t_n)) - J(\vec{X}^o(t_n)) \\ &= \frac{1}{2}\delta\vec{X}^o(t_n)^t \frac{\partial^2 J(\vec{X}^o(t_n))}{\partial \vec{X}(t_n)^2}\delta\vec{X}^o + O(\|\delta\vec{X}^o(t_n)\|^3), \end{aligned} \tag{2.62}$$

where the second derivative of J with respect to the observations is the Hessian of J with respect to the state variable at time t_n. Equation (2.62) describes the changes in the cost function resulting from a change in the observation at time t_n.

The gradient $\partial J(\vec{X}^o(t_n))/\partial \vec{X}(t_n)$ of the cost function with respect to the state variable $\vec{X}(t_n)$ is the value of the first order adjoint variable at time t_n. The Hessian $\partial^2 J(\vec{X}^o(t_n))/\partial \vec{X}(t_n)^2$ of the cost function with respect to the state variable $\vec{X}(t_n)$ is not required. Only the Hessian/vector product $[\partial^2 J(\vec{X}^o(t_n))/\partial \vec{X}(t_n)^2]\vec{X}^o(t_n)$ is required and it may be obtained using an integration of the second order adjoint model Eqs. (2.7) and (2.8).

In summary, the perturbation in the cost function resulting from a perturbation in the observations at the time t_n may be obtained by performing the following operations provided the gradient of the cost function with respect to the state vector $X(t_n)$ does not vanish.

1. Generate a perturbation on the observation at time t_n;

2. Calculate the gradient of the cost function with respect to state variable $X(t_n)$, i.e. starting from $P_M = W(\vec{X}_M - \vec{X}^o_M)$, integrate the "$forced$" adjoint equation

$$P_i = [I + \Delta t(\frac{\partial F}{\partial \vec{X}})^*_i]P_{i+1} + W(\vec{X}_i - \vec{X}^o_i), \tag{2.63}$$

   backwards in time from t_M to t_n. The final result P_n is the gradient of the cost function;

3. Use equation (2.61) to obtain the corresponding perturbation in the cost function resulting from the perturbation in the observations at time t_n.

If the gradient of the cost function with respect to the state vector $X(t_n)$ is close to zero, then we modify the last two steps as

1. Calculate the Hessian/vector product $[\partial^2 J(\vec{X}^o(t_n))/\partial \vec{X}(t_n)^2]\vec{X}^o(t_n)$. That is to integrate the second order adjoint model Eqs. (2.7) and (2.8) backwards in time from t_M to t_n. The final result $\hat{P}_n$ is the required Hessian/vector product;

2. Use equation (2.62) to obtain the corresponding perturbation in the cost function resulting from the perturbation in the observations at time t_n.

It is worthwhile noting that we need not integrate the FOA equations repeatedly to obtain the gradient of the cost function with respect to $\vec{X}_{t_n}$. We need only to integrate the FOA equations backwards in time starting from t_M to t_n and store the FOA variable at each iteration in memory. The final value is the gradient of the cost function with respect to the state variables at time t_n. Once these gradients are calculated, they need not be recalculated. They can be used repeatedly to calculate the perturbations in the cost function for different perturbations in the observations.

2.5.2 Numerical results from model generated data

A sensitivity study was conducted by using the same model as described in Section 3 and the sensitivity calculations were carried out using formula (5.7) at the end of the variational data assimilation. First we choose a point $(x_{15}\,,\, y_{10})$ in the assimilation window where $x_{15} = x_0 + 15\Delta x$ and $y_{10} = y_0 + 10\Delta y$. Suppose a 1% perturbation in the observations occurs only at this point for the two components of the wind velocity field and the geopotential field. The variation of the resulting perturbations in the cost function as a function of the number of time steps is displayed in Figure 2.13 (solid line). The results indicate that the changes in the cost function with respect to the changes in the observations at a fixed point are different for different times in the window of assimilation. If perturbations are imposed first only on the $u-$wind component, then only on the $v-$wind component and then only on the geopotential field ϕ, the corresponding perturbations in the cost function exhibit

different variations with time as shown in Figure 2.13 by the dotted line, dashed line and dash-dot line, respectively. This figure indicates also that perturbations in the observed geopotential field have more impact on the cost function than these in the observed velocity field. The changes in the cost function arising from changes in the u-wind component and v-wind component observations are close to zero at all times. Similar experiments conducted at different grid points yielded similar results.

In order to study the importance of observations at different space locations, three different points are chosen. They are located at (x_5, y_{15}), (x_{10}, y_{10}), (x_{15}, y_5), respectively, representing low, middle and high points in the isoline values of the geopotential field. From Figure 2.14 we observe that changes in the magnitudes of the cost function at the three points behave in a similar pattern and the sensitivities at the three points differ dramatically at the beginning of the assimilation window.

Finally, we study the impact of the perturbations on all the observational data. The results are displayed in Figure 2.15. This figure clearly indicates that perturbations of the observations at the beginning of the window of assimilation have a larger impact on the sensitivity of the cost function with respect to the observations.

Presented in Table 1 are sensitivities (the column labeled "Predicted change") of the cost function due to 1% changes in the observations at the beginning of the assimilation window and the corresponding actual changes (the column labeled "Actual change") obtained by rerunning the model after actually varying the observations by 1%. The good agreement obtained between the actual and predicted changes as shown in Table 1 provides evidence as to the adequacy of the numerical method used to obtain these sensitivities.

2.5.3 Numerical results using real analysis of First Global Geophysical Experiment (FGGE) data

In order to examine the sensitivity of the cost function with respect to real analyses, we employed a set of FGGE data of height and horizontal wind fields at 500mb level at 0 and 18UTC, May 26, 1979. The data are equally spaced with $\Delta\lambda = \Delta\phi = 1.875°$. Using the formula

$$\phi(J) = \frac{y}{\alpha} + \phi_0 = \frac{-2200 + (J-1) * 220}{\alpha} + \phi_0, \tag{2.64}$$

$$\lambda(I) = \frac{x}{\alpha \cos\phi(J)} + \lambda_0, \tag{2.65}$$

we obtain a correspondence between points on the sphere and grid points located on a limited area on a $\beta-$plane approximation at $(32°, 130°)$, which approximately represents the center of the zonal jet. Using a cubic interpolation we obtained the height and horizontal wind data on the grid points. Then we carried out another cubic interpolation near the left boundary in order to impose a periodic boundary condition in the x-direction. Near the top and bottom boundaries we used a linear interpolation to impose solid boundary conditions. The fields thus obtained are shown in Figure 2.16.

The geopotential and wind fields at time 0UTC were used to produce the model-generated observations. The minimization started from geopotential and wind fields distribution at time 18UTC. The difference between these two fields is shown in Figure 2.17. Having the model generated observations, the minimization should be able to reduce the value of the cost function as well as the norm of its gradient, and the retrieved differences should be zero. This turns out to be the case.

The sensitivity calculations were carried out using formulas (2.61) plus (2.62) at the beginning of the variational data assimilation. From Figure 2.16, we note that meteorologically intensive events occur at the center area of the limited-area domain while fewer such events occur at the corners of the limited-area domain. We choose

two points (x_{10}, y_{10}) and (x_7, y_5), which are located in the center and in the left bottom corner of the limited-area domain, respectively. We then introduced a 1% perturbation in the geopotential and wind fields at these two points. The variations of sensitivities of the cost function with the number of time steps in the window of the assimilation are displayed in Figure 2.18, the solid line corresponding to sensitivity at the point (x_{10}, y_{10}) and the dotted line to sensitivity at the point (x_7, y_5). Clearly the sensitivity of the cost function with respect to observations at point (x_{10}, y_{10}) is larger than that at the point (x_7, y_5). This confirms that the cost function is more sensitive to observations at points where intensive events occur. It also means that the accuracy of observations at locations where intensive events occur has more impact on the quality of the VDA retrieval.

2.6 Conclusions

In this chapter, a SOA model was developed, providing second order information. The coding work involving in obtaining the SOA model is rather modest once the FOA model has been developed.

One integration of the SOA model yields an accurate value of a column of the Hessian of the cost function with respect to the control variables provided the perturbation vector is a unit vector with one component being unity and the rest being zeros. Numerical results show that the use of the SOA approach to obtain the Hessian/vector product is much more advantageous than the corresponding finite-difference approach, since the latter yields only an approximated value of the Hessian/vector product which may be a very poor approximation. The numerical cost of using the SOA approach is roughly the same as that of using the finite-difference approach. This application of the SOA model is crucial in the implementation of the large-scale truncated Newton method, which was proved to be a very efficient algorithm for large-scale unconstrained minimization [225].

Another application of the SOA model is the calculation of eigenvalues and eigenvectors of the Hessian. There are several iterative methods such as the power, Rayleigh quotient or the Lanczos method [199], which (combined with deflation) require only information of the Hessian multiplied by a vector to calculate several dominant eigenvalues and their respective eigenvectors. Such a calculation using the power method is presented in this chapter and reveals that most changes in the largest eigenvalue of the Hessian occur during the first few iterations of the minimization procedure, which may explain why most of large-scale features are retrieved earlier than the small scale features in the VDA retrieval solution during minimization [149] while the positiveness of the smallest eigenvalues of the Hessian of the cost function during the minimization process indicate the uniqueness of the optimal solution.

We also examined the sensitivity of the cost function to observational errors using a two dimensional limited-area shallow water equations model. We found that the sensitivity depends on the time when the errors occur, the specific field containing the errors, and the spatial location where these errors occur. The cost function is more sensitive to observational errors occurring at the beginning of the assimilation window, to errors in the geopotential field, and to errors at grid point locations where intensive events occur.

Sensitivity analysis using balanced perturbations will be reported later where we pay special attention to the spatial scale of the perturbations. Further research on the issue of calculating the inverse Hessian multiplied by a vector is currently under consideration, the latter being of crucial importance for developing a new efficient large-scale unconstrained minimization algorithm.

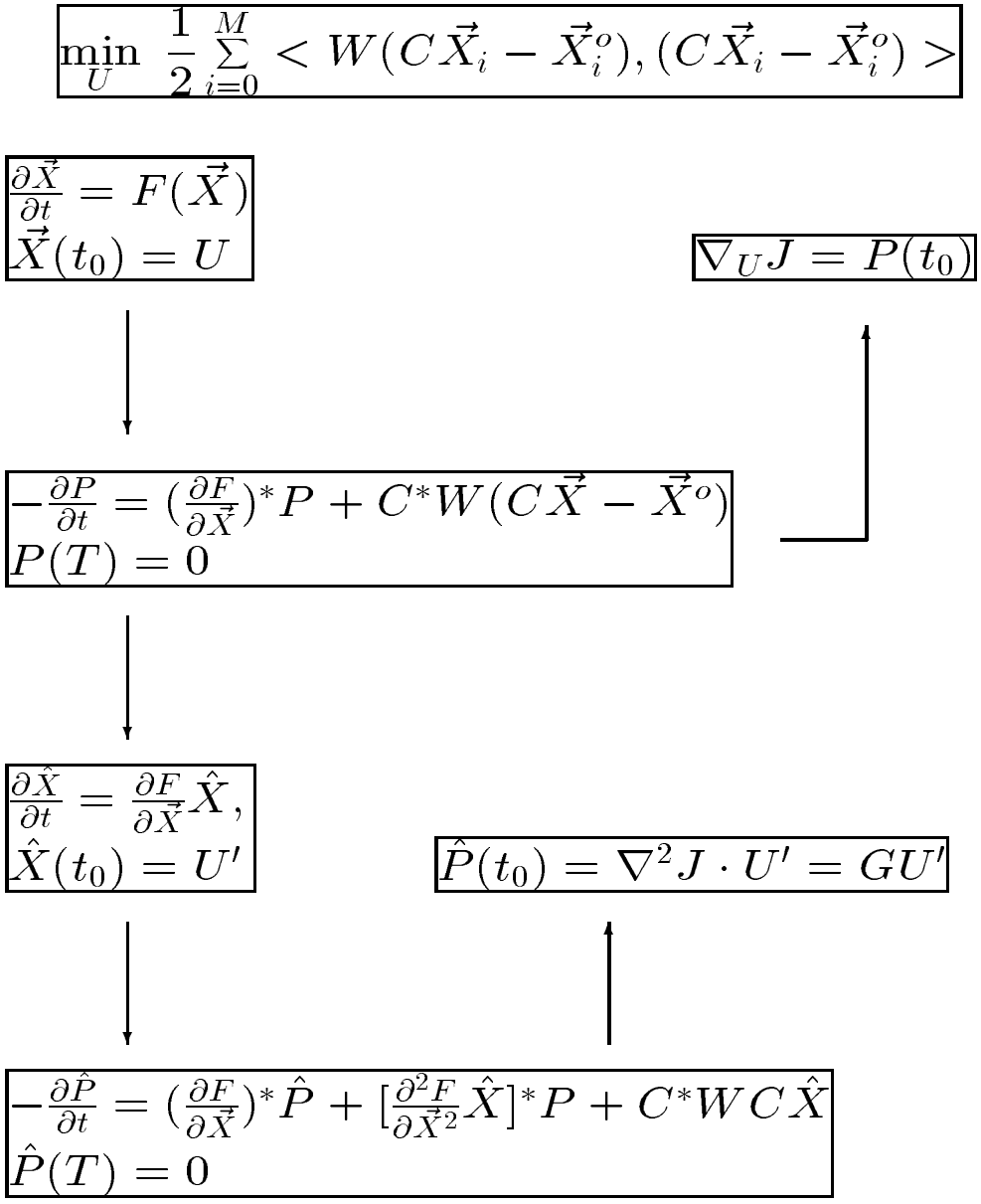

Figure 2.1: Schematic flowchart of the relationships between the model equations and the adjoint equations. The gradient $\nabla_U J$ and the Hessian/vector product GU' are obtained by integrating the FOA and SOA models, respectively.

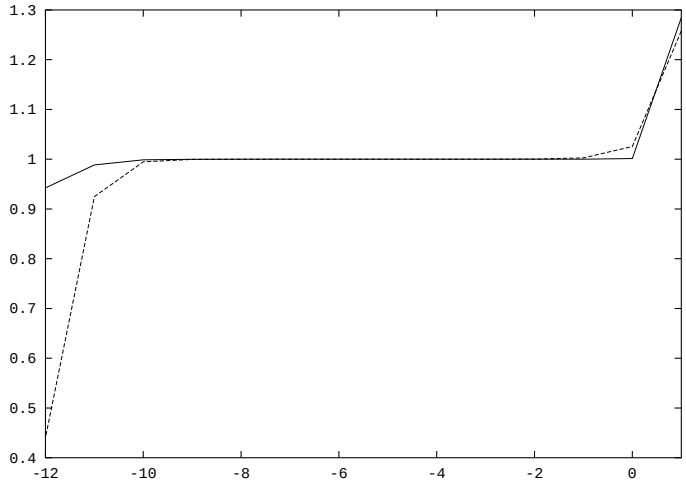

Figure 2.2: Verifications of the correctness of the gradient calculation (dotted line) and Hessian/vector product calculation (solid line) by FOA and SOA models, respectively.

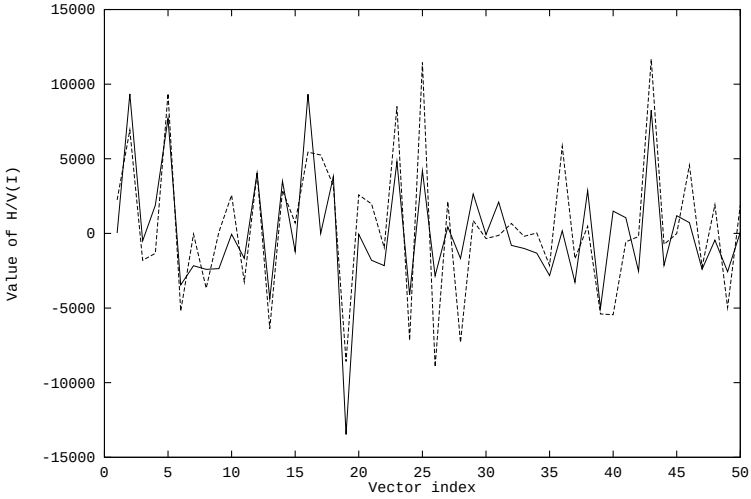

Figure 2.3: The first 50 components of the Hessian/vector products at the optimal solution obtained by the finite-difference approach (dash line), and the second order adjoint method (solid line) when $\alpha = 10$

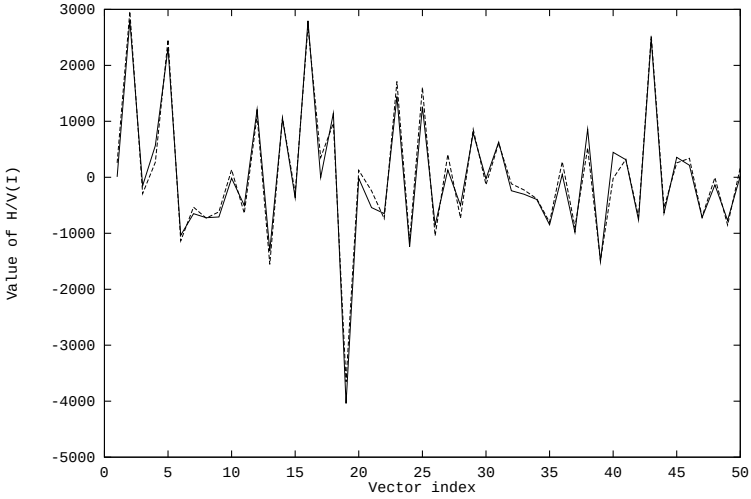

Figure 2.4: Same as Figure 2.3 except $\alpha = 3$.

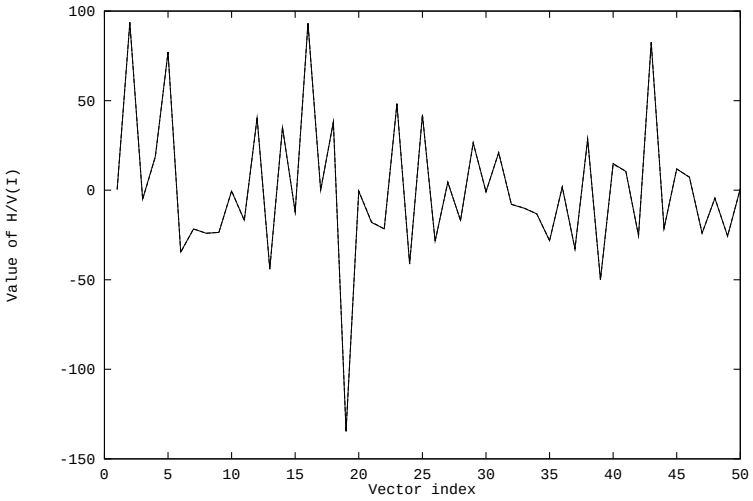

Figure 2.5: Same as Figure 2.3 except $\alpha = 0.01$.

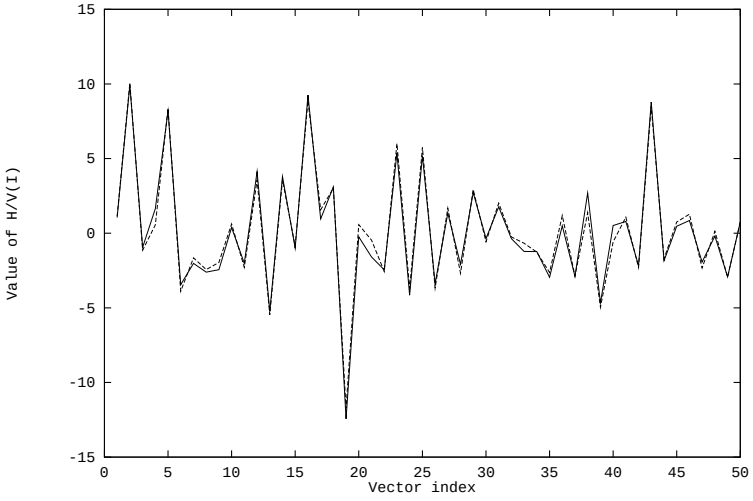

Figure 2.6: Same as Figure 2.5 except at the initial guess point.

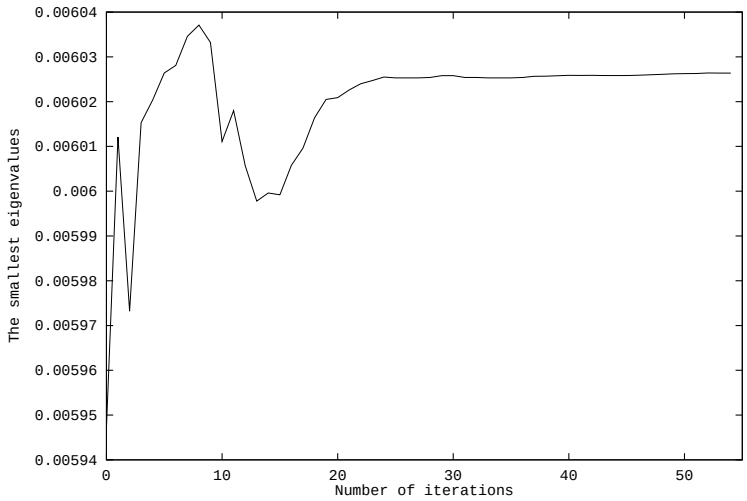

Figure 2.7: Variation of the smallest eigenvalue of the Hessian of the cost function with the number of iterations.

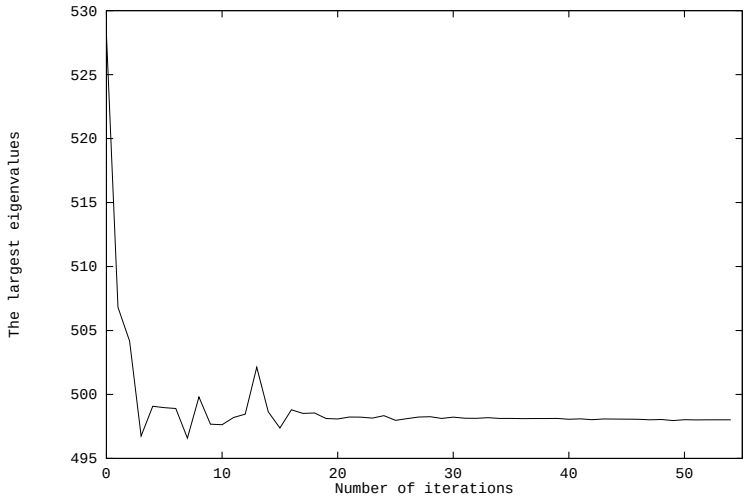

Figure 2.8: Variation of the largest eigenvalue of the Hessian of the cost function with the number of iterations.

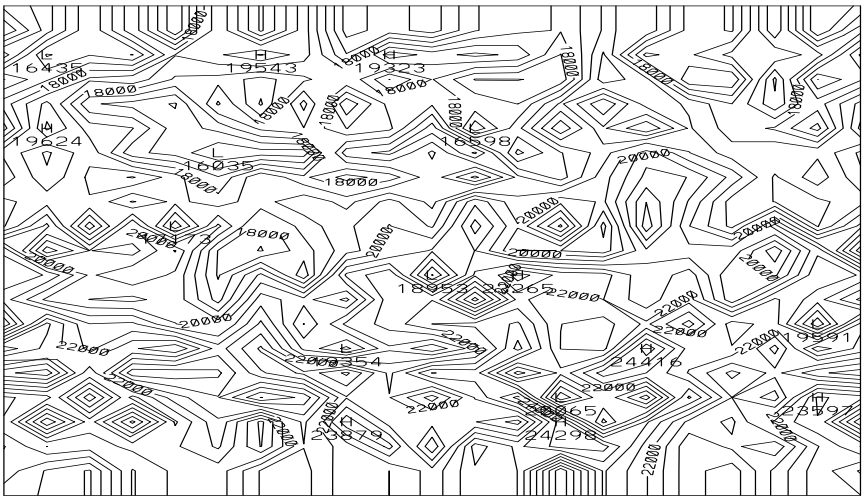

Figure 2.9: Distribution of the randomly perturbed geopotential field.

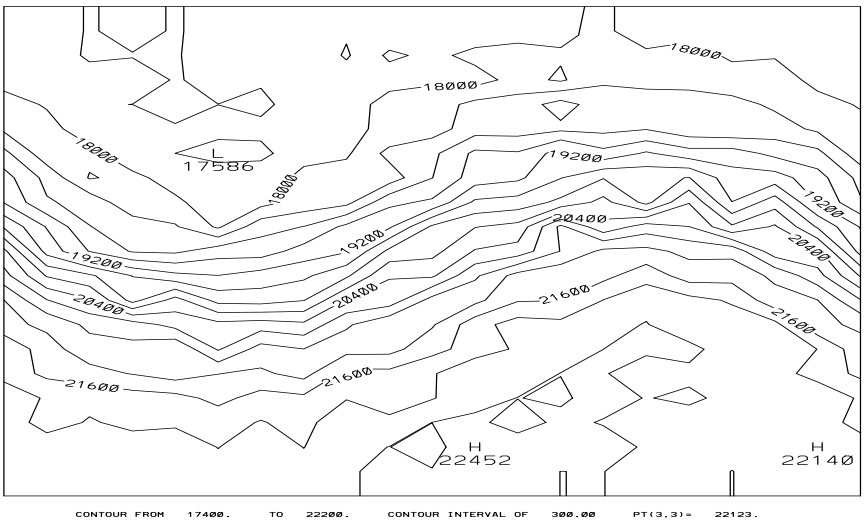

Figure 2.10: Reconstructed geopotential field after 6 iterations of minimization.

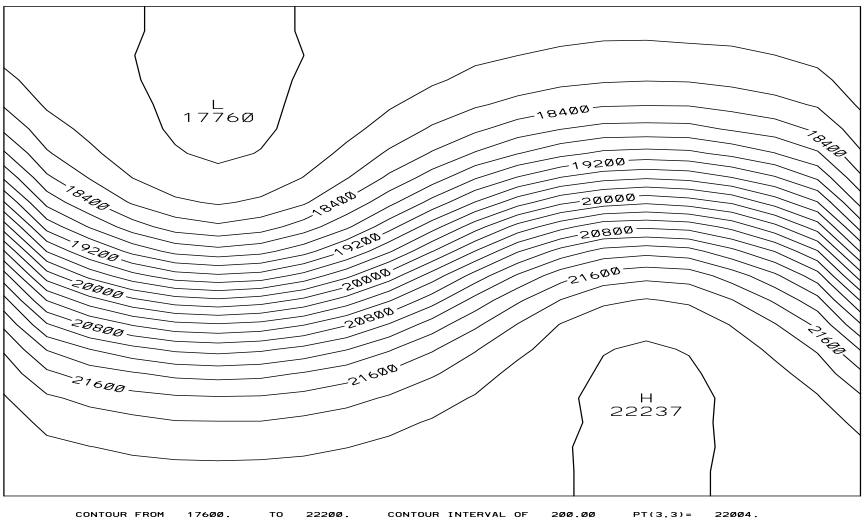

Figure 2.11: Reconstructed geopotential field after 25 iterations of minimization.

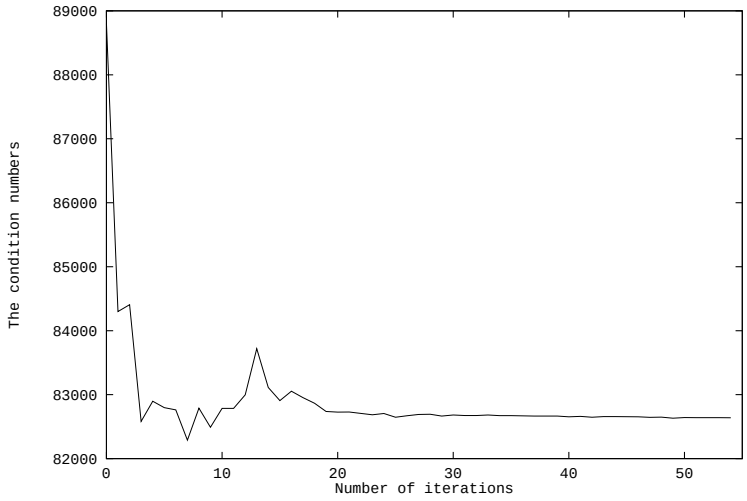

Figure 2.12: Variation of the condition numbers of the Hessian of the cost function with respect to the number of iterations.

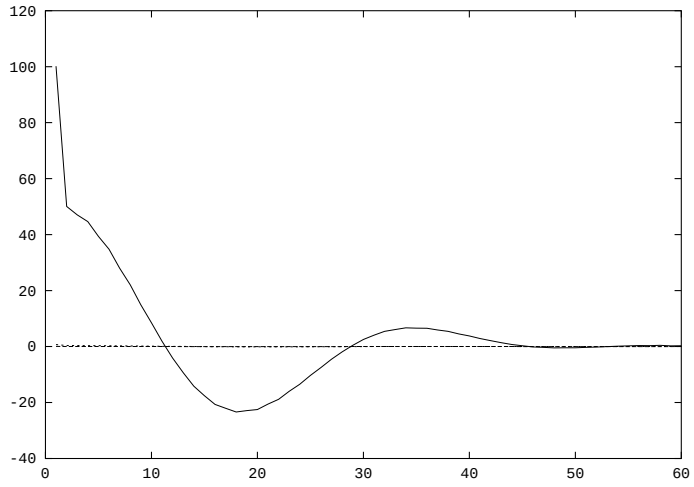

Figure 2.13: Time variation of the sensitivities of cost function J to 1% observational error in the v-wind component (dashed line), the u-wind component (dotted line which coincides with the dashed line), the geopotential field ϕ (dash-dotted line which coincides with the solid line) and in all the three fields (solid line) at point (x_{15}, y_{10}).

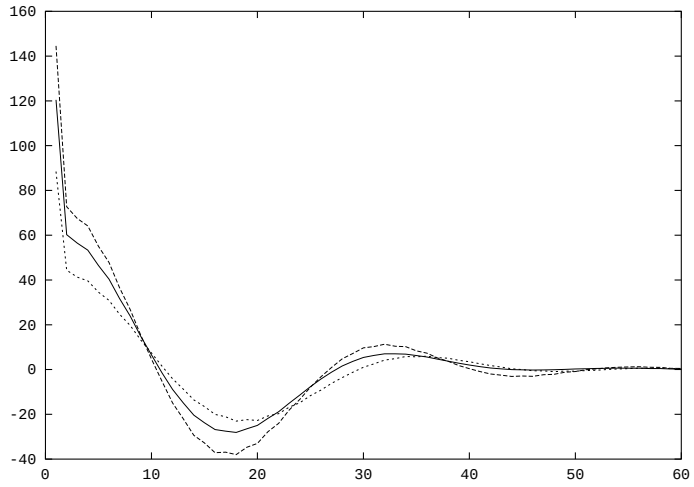

Figure 2.14: Time variation of the sensitivities of cost function J to 1% observational error at points (x_{10}, y_{10}) (solid line) (x_5, y_{15}) (dotted line), and (x_{15}, y_5) (dash line) in the wind and the geopotential fields.

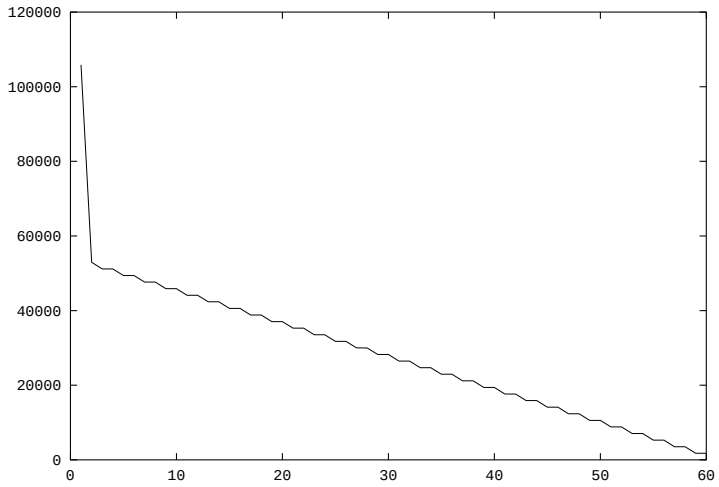

Figure 2.15: Time variation of the sensitivities of cost function J to 1% observational error on all grid points in the wind and the geopotential fields.

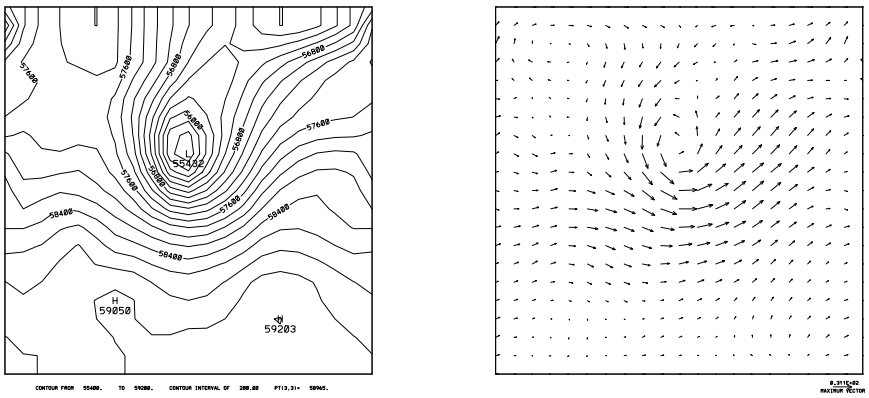

Figure 2.16: Distribution of (a) the geopotential and (b) the wind fields for the First Global Geophysical Experiment (FGGE) data at 0UTC 05/26, 1979 on the $500mb$. The contour intervals are $200m^2/s^2$ and the magnitude of maximum vector is 0.311E+02m/s.

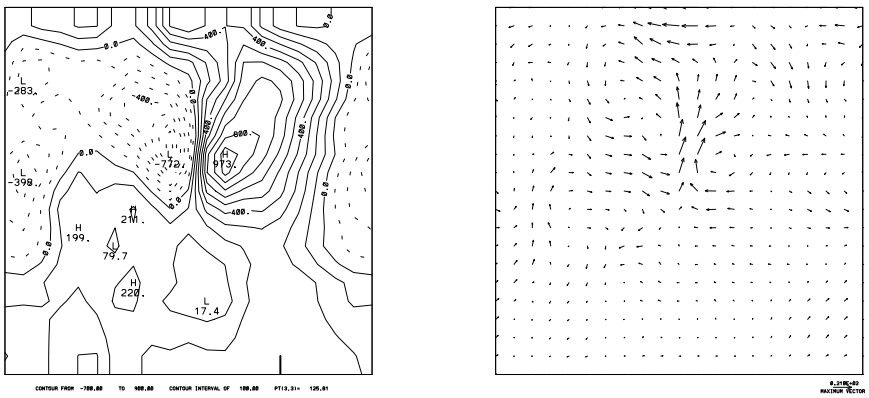

Figure 2.17: Distribution of the difference fields of the geopotential (a) and the wind (b) fields at 18UTC and 0UTC on $500mb$ between 18UTC and 0UTC times. The contour intervals are $100m^2/s^2$ and the magnitude of maximum vector is 0.210E+02m/s.

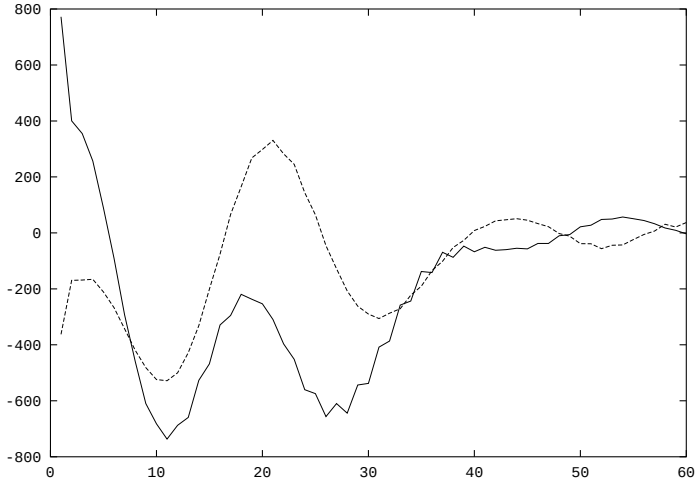

Figure 2.18: Time variation of the sensitivities of the cost function J to 1% observational error in the wind and the geopotential fields at grid points (x_{10}, y_{10}) (solid line) and (x_7, y_5) (dotted line).

Table 2.1: A comparison between the actual and predicted changes due to identical perturbations at specified locations at the beginning of the assimilation window.

Location where perturbations are given	Specific perturbations at the given location	Predicted change	Actual change
(x_{15}, y_{10})	1% on u, v and ϕ.	100.1707	100.2702
(x_{15}, y_{10})	1% on u only.	0.7124	0.7328
(x_{15}, y_{10})	1% on v only.	0.0633	0.0652
(x_{15}, y_{10})	1% on ϕ only.	99.1629	99.2294
(x_{10}, y_{10})	1% on u, v and ϕ.	120.4539	120.2349
(x_5, y_{15})	1% on u, v and ϕ.	144.5532	144.3873
(x_{15}, y_5)	1% on u, v and ϕ.	88.4600	88.4590
All points.	1% on u, v and ϕ.	105860.5213	105867.8337

CHAPTER 3

THE ADJOINT TRUNCATED NEWTON ALGORITHM FOR LARGE-SCALE UNCONSTRAINED OPTIMIZATION

3.1 Introduction

The issue of variational 4-dimensional data assimilation (VDA) applied to the shallow water equations (SWE) model has been the subject of several recent research papers such as [34], [108], [218], [225], [227], [228] etc. In Chapter 2, we introduced the SOA theory and applied it the SWE model. It was found that the Hessian/vector product obtained by using the SOA approach is more accurate than the finite difference approach and the convergence rate is slow.

Newton descent methods have never been used in a realistic large-scale optimal control of distributed parameters problems due to the fact that computation and storage of the Hessian matrix are too costly to be practical. Among feasible methods for large-scale unconstrained minimization are (a) the limited memory conjugate gradient method ([144], [190]); (b) quasi-Newton type algorithms ([46], [70], [155]); (c) limited-memory quasi-Newton methods such as LBFGS algorithm ([117], [154]) and (d) truncated-Newton algorithms ([49, 131, 133, 134, 137, 185, 186]).

The TN algorithm can be applied to many problems such as to the above unconstrained minimization problem (Eq. (2.1)) and to parallel minimization problems ([135], [136]). The main purpose of the present work is to propose a modified version of the truncated-Newton (TN) algorithm of Nash [134] which uses the SOA technique to obtain a more accurate Hessian/vector product required in calculating the

Newton line search direction and to compare the numerical results obtained by the ATN algorithm with these obtained by the TN algorithm of Nash [134] as well as by the LBFGS algorithm [117]. The ATN algorithm is only useful for optimal control problems where adjoint model codes exist or could be easily derived.

The truncated-Newton algorithms attempt to blend the rapid (quadratic) convergence rate of classic Newton method with feasible storage [134] and computational requirements [137] for large-scale unconstrained minimization problems. When these algorithms are used for the large-scale unconstrained minimization [227], the Hessian/vector product is usually obtained by applying a finite-difference (FD) approximation technique to the gradient of the cost function with respect to the initial conditions, while the gradient is calculated by using the first order adjoint (FOA) technique. Other alternatives for obtaining the Hessian/vector product such as automatic differentiation ([55], [69]), analytical evaluation and higher order FD approximations exist. However, they are either not feasible for this optimal control problem applied to VDA or are too costly to apply.

We will introduce in Section 3.2 the FOA and the SOA techniques used in the VDA to obtain a more accurate Hessian/vector product. In Section 3.3 we present numerical results obtained by using the ATN algorithm , and compare them with the results obtained by using the limited memory quasi-Newton algorithm of Liu and Nocedal [117] and the truncated-Newton algorithm of Nash [134]. A detailed discussion concerning the accuracy of the FD approximation of the Hessian/vector product and the sources of error related to this approximation is provided. Conclusions as well as topics for further research related to the ATN approach are presented in Section 3.4.

3.2 The adjoint Newton algorithm

3.2.1 Description of the first and second order adjoint theories

For a perturbation U' on the initial condition U in Eq. (1.5), the exact evolution of the resulting solution due to the initial perturbation U' is given by the difference, $\vec{X}_2 - \vec{X}_1$, between the two solutions of the model Eq. (1.4) with initial conditions $\vec{X}_1(t_0) = U$ and $\vec{X}_2(t_0) = U + U'$, respectively. The tangent linear variable defined by Eqs. (1.10)-(1.11) only yields a second order approximation, with respect to U', to the exact evolution of the perturbation due to this perturbation U'. The Gateaux derivative, $< \nabla_U J, U' >$, of the cost function J is given by

$$\delta J =< \nabla_U J, U' >= \int_{t_0}^{t_f} < W(C\vec{X} - \vec{X}^o), C\hat{X} > dt, \tag{3.1}$$

To exhibit the linear dependence of δJ with respect to U' and consequently in order to compute the gradient of J, we introduce the adjoint variable P. Taking the inner product of Eqs. (1.10) with $-P$ and integrating between times t_0 and t_f gives

$$\int_{t_0}^{f_t} < -P, \frac{\partial \hat{X}}{\partial t} > dt = \int_{t_0}^{f_t} < -P, \frac{\partial F}{\partial \vec{X}} \hat{X} > dt, \tag{3.2}$$

Integrating by parts, one obtains

$$\begin{aligned} & < -P(t_f), \hat{X}(t_f) > + < P(t_0), \hat{X}(t_0) > + \int_{t_0}^{t_f} < \hat{X}, \frac{\partial P}{\partial t} > \\ = & -\int_{t_0}^{t_f} < [\frac{\partial F}{\partial \vec{X}}]^* P, \hat{X} > dt, \end{aligned} \tag{3.3}$$

or

$$\begin{aligned} & < -P(t_f), \hat{X}(t_f) > + < P(t_0), \hat{X}(t_0) > \\ = & \int_{t_0}^{t_f} < \hat{X}, -\frac{\partial P}{\partial t} - [\frac{\partial F}{\partial \vec{X}}]^* P > dt, \end{aligned} \tag{3.4}$$

Now let us define P as being the solution of

$$-\frac{\partial P}{\partial t} = \Big(\frac{\partial F}{\partial \vec{X}}\Big)^* P + C^* W(C\vec{X} - \vec{X}^o), \tag{3.5}$$

$$P(t_f) = 0, \tag{3.6}$$

Then from Eq. (3.1) one obtains

$$\delta J(U, U') =< \nabla_U J, U' >=< U', P(t_0) > \tag{3.7}$$

and

$$\nabla_U J = P(t_0). \tag{3.8}$$

Therefore the gradient of the cost function is obtained by a backwards integration of the adjoint system (3.5-3.6). The dependence of J on the initial condition U is implicit via X.

It is important to realize that the TLM defined by Eqs. (1.10)-(1.11) is only second order accurate except when the original model equations Eq. (1.4) are linear. Although the TLM is used to derive Eq. (3.8), Eq. (3.8) is exact since there is no TLM variable in the FOA model defined by Eqs. (3.5 and 3.6). This point is illustrated in Appendix B by using a simple nonlinear model.

For the perturbation U' on the initial condition U, the resulting perturbation on the variable P, $\hat{P}$, may be obtained from Eqs. (3.5)–(3.6) as

$$-\frac{\partial \hat{P}}{\partial t} = \Big(\frac{\partial F}{\partial \vec{X}}\Big)^* \hat{P} + \Big[\frac{\partial^2 F}{\partial \vec{X}^2}\hat{X}\Big]^* P + C^* W C\hat{X}, \tag{3.9}$$

$$\hat{P}(t_f) = 0. \tag{3.10}$$

Eqs. (3.9) and (3.10) define the SOA model of Eqs. (1.4) and (1.5).

Let us denote the FOA variable after a perturbation U' on the initial condition U by $P_{(U+U')}$, then according to the definition

$$P_{(U+U')}(t_0) = P(t_0) + \hat{P}(t_0). \tag{3.11}$$

Expanding $\nabla_{U+U'}J$ around U in Taylor series and retaining only the first order terms, results in

$$\nabla_{U+U'}J = \nabla_U J + \nabla_U^2 J \cdot U' + \frac{1}{2}(U')^* \frac{\partial^3 J(\vec{\xi}_0)}{\partial U^3} U', \tag{3.12}$$

where $\nabla_U^2 J$ is the Hessian matrix of second derivatives of the cost function with respect to the initial condition and $\vec{\xi}_0$ is a point in the interval $[U, U+U']$. From Eq. (3.8), we know that

$$\nabla_{U+U'}J = P_{(U+U')}(t_0). \tag{3.13}$$

Using Eq. (3.8) and Eqs. (3.11)-(3.12), one obtains

$$\nabla_U^2 J \cdot U' = \hat{P}(t_0) - \frac{1}{2}(U')^* \frac{\partial^3 J(\vec{\xi}_0)}{\partial U^3} U', \tag{3.14}$$

where $G = \nabla_U^2 J$. Therefore the error in a SOA approximation is proportional to $\|U'\|^2$.

3.2.2 The adjoint truncated-Newton method.

We provide a brief description of the ATN algorithm which differs from the TN algorithm ([134], [137]) only in the Hessian/vector calculation required for solving the Newton equations at the k-th iteration

$$G_k \vec{d}_k = -\vec{g}_k, \tag{3.15}$$

where G_k is the Hessian of the cost function, $\vec{d}_k$ is the linear search direction and $\vec{g}_k = \nabla J(\vec{U}_k)$ is the gradient of the cost function with respect to the initial conditions. For a complete description of the TN algorithm, see Nash [134] and Appendix C.

The main steps of the ATN algorithm are as follows:

(1) Choose $\vec{U}_0$, an initial guess to the minimizer $\vec{U}^*$ and set the iteration counter to $k = 0$.

(2) Test $\vec{U}_k$ for convergence. If the following convergence criterion is satisfied

$$\|\vec{g}_k\| < 10^{-5} \cdot \|\vec{g}_0\|, \tag{3.16}$$

then stop. Otherwise continue.

(3) Solve approximately the Newton Eqs. (3.15) using a preconditioned modified-Lanczos algorithm where the Hessian/vector product is obtained using a backwards integration of of the SOA model given by Eqs. (3.9) and (3.10).

(4) Set $k = k + 1$ and update

$$\vec{U}_{k+1} = \vec{U}_k + \alpha_k \vec{d}_k, \tag{3.17}$$

where α_k is the step-size obtained by conducting a line search using Davidon's cubic interpolation method [46]. Go to step 2.

The computational cost required to obtain the Hessian/vector product is similar for both the FD approach and the SOA approach. The SOA approach requires us to integrate the original nonlinear model and its tangent linear model forward in time once and integrate the FOA model and the SOA model backwards in time once. The FD approach requires the integration of the original nonlinear model forward in time and the FOA model backwards in time twice. The computational costs for each model integration are comparable.

3.3 Numerical results obtained using the adjoint truncated-Newton algorithm.

In this section we display numerical results obtained by applying the ATN algorithm for the large-scale unconstrained minimization of the functional in the VDA

and compare the results obtained by the ATN algorithm with these obtained by both the TN and the LBFGS algorithms.

3.3.1 Application of the adjoint truncated-Newton algorithm to variational data assimilation

A simple experiment was conducted applying the ATN algorithm to minimize the cost functional J given by Eq. (2.1) in the VDA using the SWE model. The experiment is devised as follows: the model generated values starting from the initial condition of Grammeltvedt [75] are used as observations, the initial guess is a randomly perturbed Grammeltvedt initial condition, and the length of the time assimilation window is 10 hours. We know the exact solution, i.e. that the value of the cost function at the minimum must be zero since we must retrieve the original initial conditions. The ATN algorithm described in Section 3.2.2 is used here for large-scale unconstrained minimization in the VDA experiment. The maximum number of conjugate gradient inner-iterations (MCGI) allowed for each ATN iteration is chosen as 50. The number of BFGS corrections that is to be stored in memory is denoted by M.

Computations were performed on the CRAY-YMP supercomputer at the Supercomputer Computations Research Institute in Florida State University. All the routines are coded in single precision FORTRAN. The runs were made on the CRAY-YMP supercomputer, for which the relative machine precision ϵ is approximately 10^{-14}. The variation of the objective function scaled by its initial value (J/J_0) as well as that of the norm of the gradient also scaled by its initial value ($\|\vec{g}\|/\|\vec{g}_0\|$) as a function of the number of iterations are displayed in Figure 3.1, respectively. Figure 3.1 shows that after 16 minimization iterations the value of the cost function and the norm of the gradient were reduced by 10 and 6 orders of magnitude,

respectively. At this stage the prescribed convergence criterion given by Eq. (3.16) is satisfied. The CPU time used by the ATN algorithm is 10.817s. The rms error, $\sqrt{\|\phi_r - \phi_u\|^2/N}$, between retrieved geopotential field ϕ_r after 16 iterations and the unperturbed geopotential field ϕ_u is $0.9069 m^2/s^2$ which is 3 orders of magnitude smaller than that of the perturbations, where N is the number of components in ϕ_r. We conclude therefore that the ATN algorithm performs well both in terms of CPU time and the number of iterations.

3.3.2 Numerical results

In this section we compare the numerical behaviour of the ATN unconstrained minimization algorithm with those of other robust large-scale unconstrained minimization methods. The methods tested are:

(1) TN–the truncated-Newton method of Nash ([132], [134]).

(2) LBFGS–the limited memory quasi-Newton method of Liu and Nocedal [117].

The test problem is the same as that described in section 3.3.1. Computational efficiency and accuracy were used as leading criteria. For the ATN, TN and LBFGS algorithms, the same convergence criterion as set by Eq. (3.16) is applied. The numerical results obtained are displayed in Table 3.1.

The first column displays the unconstrained minimization algorithms tested. The second column displays the parameters used for each algorithm. The third, fourth and eighth columns record the number of iterations (Iter.), the number of function calls (NFC), and the CPU time in seconds required to satisfy prescribed convergence criteria, respectively. The fifth column records the total number of conjugate gradient iterations (NCG) used to determine the Newton descent directions. The next two columns display the scaled cost function (J_k/J_0) and the scaled gradient norm ($\|\vec{g}_k\|/\|\vec{g}_0\|$) at the end of the assimilation process, respectively.

Both the ATN and TN algorithms obtain the Newton descent direction by solving approximately the Newton equations using a truncated conjugate gradient algorithm. They only differ in the calculation of the Hessian/vector product. However their relative performances turn out to be dramatically different. Table 3.1 indicates that for different MCGI allowed for each inner iteration, the ATN algorithm outperforms the TN algorithm. When MCGI is 5, the TN algorithm stopped without satisfying the prescribed convergence criterion. When MCGI is 4, the TN algorithm required three times the CPU time used by the ATN algorithm to satisfy the same convergence criteria. In terms of CPU times the ATN, TN and LBFGS algorithms yield optimal results when MCGI is 4 and 3 and M is 5, respectively. In this case the CPU time required by ATN algorithm is about half of that required by either the TN or the LBFGS algorithms.

If we relax the convergence criterion given by Eq. (3.16) by two orders of magnitude, then the ATN, TN and LBFGS algorithms require 8, 16 and 55 iterations and take 5s, 8s and 6s of CPU time to converge where we used MCGI=50 for the ATN and TN and M=5 for the LBFGS, respectively. Therefore even for the relaxed accuracy requirement the ATN algorithm performed slightly better than the LBFGS algorithm in terms of CPU time required to satisfy the convergence criteria.

Let us define the degree of nonlinearity of the cost function at $\vec{U}$ as

$$DN(\vec{U}) = (J(\vec{U}) - [J(\vec{U}^*) + p^t \nabla J(\vec{U}^*) + 0.5 p^t \nabla^2 J(\vec{U}^*) p])/\|p\|^3, \tag{3.18}$$

where $\vec{U}$ is a point between the starting point U_0 and the solution $\vec{U}^*$, $p = \vec{U} - \vec{U}^*$ and p^t denotes the transpose of p. $DN(\vec{U})$ gives a measure of the size of the third derivative or a deviation from quadratic behaviour (see Nash and Nocedal [137]).

For our test problem we noticed that $DN(\vec{U})$ increases from $1.7E-7$ to $4.7E+12$ as $\vec{U}$ approaches $\vec{U}^*$ from the starting point. Therefore we may classify our test problem as a highly nonlinear problem near the solution. Then it is not surprising that

the LBFGS algorithm outperforms the TN algorithm in terms of CPU time (Table 3.1) if we take into account Nash's observation that for most of highly nonlinear problems, the LBFGS algorithm performs better than the TN algorithm [134]. Our point is that even in this case, if we use a more accurate line search direction in the TN algorithm, the TN can outperform the LBFGS as the ATN does for the particular optimal control problem tested here.

Therefore we conclude that the ATN algorithm turns out to be the most accurate and robust amongst the large-scale unconstrained minimization algorithms tested in terms of both CPU time and number of iterations for the specific problem tested here.

3.3.3 An accuracy analysis of the Hessian/vector product.

The Hessian/vector product $G_k U'$ for a given U' required by the inner conjugate algorithm of the ATN algorithm is obtained by the SOA technique where the vector U' serves as the initial condition for the TLM model. The Hessian/vector product $G_k U'$ of the TN algorithm is obtained by the following FD approximation $G_k U'|_{FD}$

$$G_k U'|_{FD} = \frac{\hat{g}(\vec{U}_k + hU') - \hat{g}(\vec{U}_k)}{h}, \tag{3.19}$$

where G_k is the Hessian matrix at the k-th outer iteration, $h = \sqrt{\epsilon \times (1 + \|\vec{U}_k\|)}$ is the differencing parameter, where ϵ is taken to be the machine precision [134], and the computed gradients $\hat{g}(\vec{U}_k + hU')$ and $\hat{g}(\vec{U}_k)$ are obtained by using Eq. (3.8).

In order to consider the accuracy of the FD approximation with respect to the differencing parameter h, we assume $h \in [h_{min}, h_{max}]$ where h_{min} and h_{max} are taken to be the machine accuracy ϵ and 10^3, respectively, such that the interval $[h_{min}, h_{max}]$ contains all reasonable sequences of differencing parameters.

Although a FOA integration yields an exact gradient of the cost function with respect to the control variables, when a computer is used to calculate the gradient

there is always computational errors involved. Let a positive quantity ϵ_0 denote an error bound on the absolute error in the computed gradients of the cost function using Eq. (3.8) at $\vec{U}_k$ and $\vec{U}_k + hU'$. It will be assumed throughout this chapter that the value of ϵ at the given point is available; an effective technique for computing ϵ is given by Hamming [82].

So one has

$$\hat{g}_t(\vec{U}_k) = \vec{g}(\vec{U}_k) + \theta_0\epsilon_0, \tag{3.20}$$

$$\hat{g}_t(\vec{U}_k + hU') = \vec{g}(\vec{U}_k + hU') + \theta_1\epsilon_0, \tag{3.21}$$

where $|\theta_0| \leq 1$ and $|\theta_1| \leq 1$. Using a Taylor series expansion, one obtains

$$\begin{aligned} \vec{g}(\vec{U}_k + hU') &= \vec{g}(\vec{U}_k) + h\frac{\partial \vec{g}(\vec{U}_k)}{\partial \vec{U}_k}U' + \frac{h^2}{2}(U')^*\frac{\partial^2 \vec{g}(\vec{U}_k)}{\partial \vec{U}_k^2}U' + \cdots \\ &= \vec{g}(\vec{U}_k) + hG_kU'|_t + \frac{h^2}{2}(U')^*\frac{\partial^2 \vec{g}(\vec{\xi})}{\partial \vec{U}_k^2}U', \end{aligned} \tag{3.22}$$

where $G_kU'|_t = [\partial \vec{g}(\vec{U}_k)/\partial \vec{U}_k]U'$ denotes the true value of the Hessian/vector product and $\vec{\xi}$ is a point in the interval $[\vec{U}_k, \vec{U}_k + hU']$.

Solving for $G_kU'|_t$ from Eq. (3.22) and using Eqs. (3.21) and (3.20), one obtains

$$\begin{aligned} G_kU'|_t &= \frac{\vec{g}(\vec{U}_k + hU') - \vec{g}(\vec{U}_k)}{h} - \frac{h}{2}(U')^*\frac{\partial^2 \vec{g}(\vec{\xi})}{\partial \vec{U}_k^2}U' \\ &= \frac{\hat{g}(\vec{U}_k + hU') - \hat{g}(\vec{U}_k)}{h} + \frac{(\theta_0 - \theta_1)\epsilon_0}{h} - \frac{h}{2}(U')^*\frac{\partial^2 \vec{g}(\vec{\xi})}{\partial \vec{U}_k^2}U', \end{aligned} \tag{3.23}$$

where $-h(U')^*[\partial^2 \vec{g}(\vec{\xi})/\partial \vec{U}_k^2]U'/2$ is the truncation error and $(\theta_1 - \theta_0)\epsilon_0/h$ is the condition error [70]. Combining Eqs. (3.19) and (3.23), one obtains

$$G_kU'|_t = G_kU'|_{FD} + \frac{(\theta_0 - \theta_1)\epsilon_0}{h} - \frac{h}{2}(U')^*\frac{\partial^2 \vec{g}(\vec{\xi})}{\partial \vec{U}_k^2}U'. \tag{3.24}$$

According to Eq. (3.14)

$$G_kU'|_t = G_kU'|_{SOA} - \frac{1}{2}(U')^*\frac{\partial^3 J(\vec{\xi_0})}{\partial U_k^3}U', \tag{3.25}$$

where $G_k U'|_{SOA}$ denotes the Hessian/vector product obtained by using the SOA technique. Since $G_k U'|_t$ and $\partial^3 J(\vec{\xi}_0)/\partial U_k^3$ are not available, one can not calculate the over all errors in both SOA and finite-difference approximations. However one may investigate the truncation and condition errors by looking at the difference between the $G_k U'|_{SOA}$ and $G_k U'|_{FD}$,

$$\begin{aligned} & G_k U'|_{SOA} - G_k U'|_{FD} = \\ + \quad & \frac{(\theta_0 - \theta_1)\epsilon_0}{h} - \frac{h}{2}(U')^* \frac{\partial^2 \vec{g}(\vec{\xi})}{\partial \vec{U}_k^2} U' \frac{1}{2}(U')^* \frac{\partial^3 J(\vec{\xi}_0)}{\partial U_k^3} U', \end{aligned} \tag{3.26}$$

where the error $U'(U')^*[\partial^3 J(\vec{\xi}_0)/\partial U_k^3]U'/2$ resulting from the SOA approximation is fixed for fixed U' while the truncation and condition errors resulting from the finite-difference approximation change with differencing parameter.

The rms errors between the the Hessian/vector products obtained by using the SOA and the FD techniques with various differencing parameters are displayed in Table 3.2. This Table indicates that (a) the Hessian/vector products obtained by the FD and the SOA techniques follow a relationship given by Eq. (3.26) where the error term $O(h)$ dominates for large differencing parameter h, (b) for a small differencing parameter, the error term $O(1/h)$ in Eq. (3.26) dominates. The FD technique involves subtractions of nearly equal numbers which result in cancellations of significant digits and which are the reason for the increase in the rms errors and (c) the differencing parameter h in the TN Nash [134] may become too small near the end of the minimization process. Thus in practice, the SOA approximation is more accurate than the finite-difference approximation in calculating the Hessian/vector product.

Figure 3.2 shows the first 50 components of the scaled difference between the Hessian/vector products obtained by using the SOA and the FD techniques respectively after 15 iterations using the ATN algorithm. The differencing parameter is chosen as $h_1 = h$ (solid line), $h_2 = h \times 10^{-1}$ (dotted line) and $h_3 = h \times 10^3$ (dashed line)

where h is the differencing parameter used in the original TN Nash [134] minimization algorithm, respectively. The results with $h_3 = h \times 10^3$ clearly illustrate that toward the end of the minimization the differencing parameter in the TN Nash [134] is too small. This result agrees with the rms error evolution in Table 3.2. It can be seen that the FD technique can yield an approximation of similar accuracy to that obtained by SOA technique if the differencing parameter is properly chosen. However simply increasing or decreasing the differencing parameter h at every iteration will not improve the performance of the TN minimization algorithm since the differencing parameter depends on the vector U' at each iteration and the vector U' is not known prior to performing the minimization iteration.

The differencing parameter h should be chosen such as (a) it balances the truncation error of the order h with the condition error ([70], [72], [74]) of order $1/h$, (b) h must be adjusted to the size of vector U' [185, 186] and (c) h should not become so small as to cause cancellations of significant digits [198]. It is difficult to choose an h which satisfies all these requirements. Some good choices of h in addition to that used in the TN Nash minimization algorithm [134] are $h = 2(1 + \|\vec{U}_k\|)\sqrt{\epsilon}/\|U'\|$ [185, 186], $h = \sqrt{\epsilon}/\|U'\|$ [49] and $h = 2(1 + \|\vec{U}_k\|)\sqrt{\epsilon}/\|U'\|^2$ [155] etc. But results (not shown here) obtained using these choices of h are no better than those obtained when $h = \sqrt{\epsilon \times (1 + \|\vec{U}_k\|)}$ as used in the TN Nash [134]. All these choices of h cause cancellations of significant digits at some stage of the minimization process.

The convergence rate of the ATN minimization algorithm is best understood intuitively by splitting the error into three terms [155]. Let $\vec{U}^*$ be the true solution to the problem, $G(\vec{U}_k)$ the exact Hessian at $\vec{U}_k$, G_k the approximate Hessian obtained by the FD or by the SOA technique, and H_k the approximate matrix to the inverse Hessian matrix that the inner conjugate gradient algorithm actually used, i.e. $\vec{p}_k = -H_k\vec{g}_k$. Then, if the step size chosen was 1 and if $G(\vec{U}_k)$ is positive definite,

$$\vec{U}_{k+1} \quad - \quad \vec{U}^* = \vec{U}_k - H_k\vec{g}_k - \vec{U}^*$$

$$= \quad (\vec{U}_k - G(\vec{U}_k)^{-1}\vec{g}_k - \vec{U}^*) + (G(\vec{U}_k)^{-1} - G_k^{-1})\vec{g}_k + (G_k^{-1} - H_k)\vec{g}_k \quad (3.27)$$

Therefore

$$\|\vec{U}_{k+1} - \vec{U}^*\| \leq \|(\vec{U}_k - G(\vec{U}_k)^{-1}\vec{g}_k - \vec{U}^*)\| + \|(G(\vec{U}_k)^{-1} - G_k^{-1})\vec{g}_k\| + \|(G_k^{-1} - H_k)\vec{g}_k\| \quad (3.28)$$

The first error term is the Newton error at the $(k+1)$-th step. The second error term is due to the error in the SOA technique or due to discrete differencing and depends on the choice of h in the conjugate gradient algorithm. The third error term is an error due to round-off and early termination (truncation) in the conjugate gradient inner iteration. The second error induced by the SOA technique is smaller than or as big as that caused by the FD technique. Therefore the ATN minimization algorithm yields a speed-up for our test problem. It is also clear that both the ATN and TN algorithms have the same convergence rate. We know that if cost function J is sufficiently smooth, G_k is a strongly consistent approximation to $G(\vec{U}_k)$, and $G(\vec{U}^*)$ is nonsingular, then local quadratic convergence can be obtained if the differencing parameter h decreases sufficiently rapidly ([16], [71]). Under the same conditions we expect the ATN algorithm obtains the same convergence. However in practice both the ATN and TN algorithms have only a super linear convergence rate.

If we choose the vector U' as one of the coordinate directions e_1, ... e_n where e_i is the i-th unit vector and n is the number of the components in U', respectively, either the SOA or the FD technique will generate an approximation to the Hessian. Since, for smooth functions, the Hessian is symmetric, the approximate Hessian is often symmetrized by averaging corresponding elements in the upper and lower triangles. Again the SOA technique will obtain a more accurate approximation to the Hessian than (or an approximation to the Hessian as accurate as) that obtained by using the FD technique. Burger [15] pointed out that the SOA technique requires less

computing time than direct differentiation in obtaining the approximate Hessian and thus the results from the former method are more accurate than these from the latter method. Our results are consistent with those of Burger [15] in as far as the aspect of accuracy is concerned. Similar results to ours were obtained by Symes [201] showing the advantage of using the SOA technique to obtain the Hessian/vector product in the TN algorithm.

In summary, we conclude that the SOA technique yields a more accurate value of the Hessian/vector product compared to the FD technique for optimal control problems tested here. When the differencing parameter h is too small or too big, cancellations of significant digits or truncation errors dominate the FD approximation. It is hard to avoid the occurrence of these two types of error in the minimization process when the FD technique is used [211]. Use of more accurate Hessian/vector products in the inner conjugate gradient iteration of the TN algorithm results in a better line search direction as measured by the amount of decrease in the cost function, but not in the residuals [138].

3.4 Conclusions

In this chapter, we proposed a modified version of the Nash truncated-Newton algorithm [138] by using the FOA and the SOA techniques, i.e. we proposed a new method to obtain a Hessian/vector product to be used in the process of solving the Newton equations for the TN minimization algorithm for an optimal control problem. The costs of the SOA approach and the FD approach are computationally comparable, each of them requiring four different model integrations when applied to the optimal control problem tested here. But the former approach yields a more accurate Hessian/vector product, while the latter provides only a less accurate approximation to the Hessian/vector product if the finite-differencing parameter is taken to be

the square root of the machine accuracy. The numerical results indicate that the new Hessian/vector calculation strategy employed in the modified-Lanczos algorithm of Nash [138] allows the ATN algorithm to perform better than either the TN or the LBFGS algorithms both in terms of CPU time as well as in term of number of iterations required to satisfy the prescribed convergence criterion in our test problems. This result may be very useful, since truncated-Newton type methods and conjugate-gradient methods have comparable storage requirements and constitute the only practical methods for solving many large-scale unconstrained minimization problems arising in 4-D VDA.

In our test example, the eigenvalues of the Hessian at each iteration are positive [218], which implies the Hessians are positive definite and the cost function is strictly convex. Therefore the existence of a local minimum is assured. However the condition numbers at each iteration are large [218], which explains why all three algorithms require more than 16 iterations to satisfy the prescribed convergence criteria.

Theoretically, the TN and the ATN minimization algorithms have the same convergence rate. However the ATN algorithm results in a speed-up due to the use of a more accurate Hessian/vector product to obtain the line search direction.

The ATN algorithm, like its TN counterpart, is a close approximation to the Newton method at reasonable storage and computational cost. We expect the ATN algorithm to yield a similar speed-up for other large-scale unconstrained minimization problems related to optimal control and VDA. The idea of obtaining the Hessian/vector product using the SOA technique can be applied in other settings requiring large-scale minimization, in cases where an adjoint model formulation is possible, or via automatic differentiation techniques [77] - pointing to a more general applicability of this idea. An application of the ATN minimization algorithm for minimizing a cost functional in optimal control of distributed parameters in a primitive equations

3-D spectral model will be reported separately, once the SOA model of this model is derived.

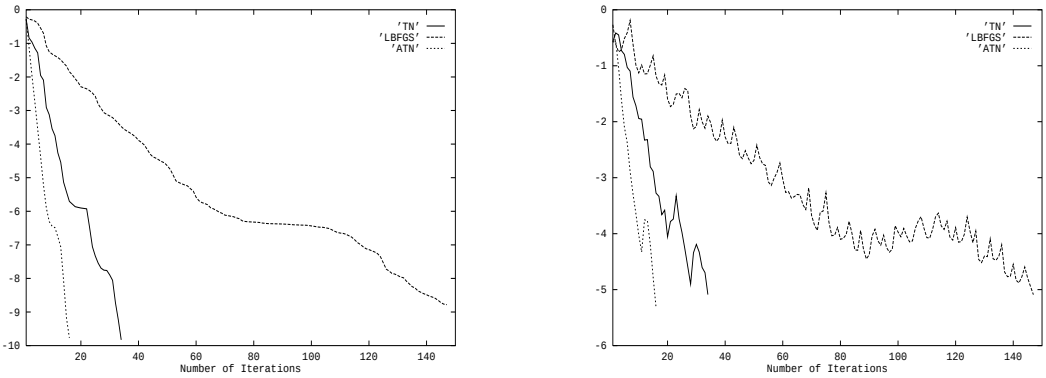

Figure 3.1: Variations of the *log* of the scaled cost function (J_k/J_0) (a) and the scaled gradient norm ($\|\vec{g}_k\|/\|\vec{g}_0\|$) (b) with the number of iterations using algorithms: ATN (dotted line), TN (solid line) and LBFGS (dashed line), respectively.

Table 3.1: Numerical results of minimization algorithms: adjoint truncated Newton, truncated Newton and LBFGS for the minimization of the cost function in the variational data assimilation problem when observations are model generated and control variables are the initial conditions.

Algorithms		Iter.	NFC	NCG	J_k/J_0	$\|\vec{g}_k\|/\|\vec{g}_0\|$	CPU
ATN	MCGI=1	96	139	93	3.622×10^{-9}	8.552×10^{-6}	23.615s
	MCGI=2	42	43	84	8.597×10^{-11}	6.922×10^{-6}	13.407s
	MCGI=3	30	46	89	1.716×10^{-10}	9.347×10^{-6}	14.116s
	MCGI=4	15	16	58	6.540×10^{-10}	8.467×10^{-6}	7.866s
	MCGI=5	20	43	91	2.590×10^{-10}	8.467×10^{-6}	13.992s
	MCGI=50	16	17	85	1.485×10^{-10}	4.822×10^{-6}	10.817s
TN	MCGI=1	104	164	104	3.211×10^{-10}	6.455×10^{-6}	35.862s
	MCGI=2	50	51	100	1.291×10^{-10}	8.577×10^{-6}	20.663s
	MCGI=3	32	38	88	6.440×10^{-10}	8.785×10^{-6}	17.125s
	MCGI=4	38	87	106	1.114×10^{-10}	5.929×10^{-6}	25.202s
	MCGI=5	25	74	75	1.008×10^{-6}	1.113×10^{-3}	Failure
	MCGI=50	34	69	116	1.485×10^{-10}	8.156×10^{-6}	24.491s
LBFGS	M=3	163	167		7.892×10^{-9}	7.804×10^{-6}	16.511s
	M=4	157	167		2.211×10^{-9}	9.655×10^{-6}	16.724s
	M=5	147	153		1.658×10^{-9}	8.138×10^{-6}	15.585s
	M=6	153	159		1.585×10^{-9}	7.368×10^{-6}	16.290s
	M=7	148	158		2.080×10^{-9}	8.136×10^{-6}	16.333s

Table 3.2: Rms errors between the Hessian/vector products obtained by using the second order adjoint and the finite-difference techniques for various differencing parameters at the end of 15 iterations of the adjoint truncated Newton minimization algorithm, respectively. The maximum number of conjugate gradient inner-iterations is 4 and h is the differencing parameter used in the original truncated Newton algorithm of Nash minimization algorithm.

Differencing parameters	rms errors
$h \times 10^{9}$	$7.3674493981314 \times 10^{-7}$
$h \times 10^{8}$	$6.9428037222996 \times 10^{-8}$
$h \times 10^{7}$	$6.8992950502148 \times 10^{-9}$
$h \times 10^{6}$	$6.8262063502701 \times 10^{-10}$
$h \times 10^{5}$	$6.2331779844824 \times 10^{-11}$
$h \times 10^{4}$	$1.1954425938165 \times 10^{-11}$
$h \times 10^{3}$	$1.4501562065627 \times 10^{-11}$
$h \times 10^{2}$	$2.9876262946020 \times 10^{-11}$
$h \times 10^{1}$	$2.6527501712652 \times 10^{-10}$
$h \times 10^{0}$	$2.6074180374738 \times 10^{-9}$
$h \times 10^{-1}$	$2.5350296945351 \times 10^{-8}$

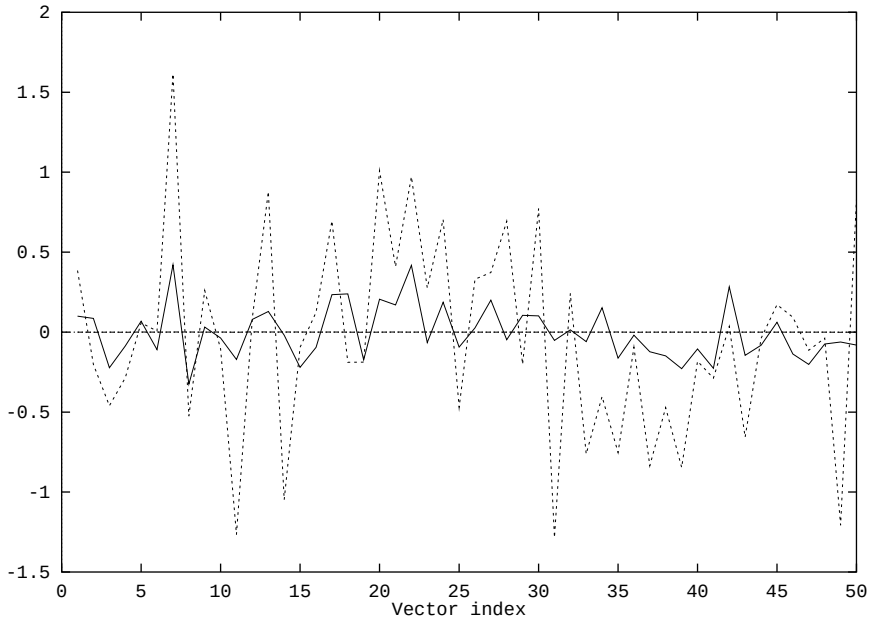

Figure 3.2: The first 50 components of the difference between the Hessian/vector products obtained by using the SOA and the FD techniques scaled by a factor $1.E09$ after 15 iterations using the ATN algorithm where the differencing parameter is chosen as $h_1 = h$ (solid line), $h_2 = h \times 1.E-01$ (dotted line) and $h_3 = h \times 1.E+03$ (dashed line) instead of h in the original TN Nash, respectively.

CHAPTER 4

THE FSU GLOBAL SPECTRAL MODEL

4.1 Introduction

The spectral method is a popular technique for hemispheric or global atmospheric modeling, in both operational and research applications. This has largely been a consequence of the discovery of the transform technique, developed by Orszag [156], and later refined by Bourke [12]. There are a number of advantages which bespeak the superiority of the spectral method over the grid-point method for large-scale prediction models. In the first place, the spectral coefficient representation of the dependent atmospheric variables carries more information (i.e. smaller scales) than the grid-point depiction, for an equal number of degree of freedom. Moreover, the Galerkin procedure (used in most of spectral models), by definition, assures that when nonlinear terms are evaluated there is no aliasing of unresolved scales into those scales which are resolved by the model's truncation limit. Aliasing is a common cause of nonlinear instability and phase errors in grid-point models, even for those using fairly sophisticated finite difference approximations for the nonlinear advective terms. With the aid of transform technique, the spectral model can efficiently compute the nonlinear terms by forming products in the real (grid) space and transforming them back to the spectral (coefficients) space. The essence of the Galerkin procedure will be preserved if exact numerical quadrature is employed for the transforms. Prior to the introduction of the transform technique, it was thought necessary to remain in the spectral space and to compute and store a tremendous number of interaction coefficients (arising from the nonlinearities in the governing equations) in the course

of a numerical forecast. The removal of this encumbrance has made spectral models competitive, in terms of computational overhead, with respect to their grid-point counterparts.

An added advantage which arises from the transform procedure is seen when physical effects are to be incorporated into the prediction scheme. The availability of the dependent variables at the transform grid points enables the calculations of parameterized physical effects in the real space. These may then be added to the dynamic tendencies prior to transforming back to the spectral domain. Drawbacks of the spectral method are related to computational issues which may become a problem for high resolution spectral methods, specially related to the unavailability of fast Legendre transforms. Moreover transporting spectral methods to high performance parallel computers of multiple instruction multiple data type (MIMD) may be less efficient than the corresponding procedure for finite difference or finite element methods.

The FSUGSM has been developed from the Canadian spectral model of the Recherche en Prevision Numerique [40]. In the development of the FSUGSM from the Canadian model, the primary effort has been directed at improving the physical effect parameterizations, adapting the model code to run efficiently on the NCAR Cray-1 computer and developing post-forecast diagnostics [162].

The FSUGSM is a multi-level primitive equations model with λ (longitude), θ (latitude) as horizontal coordinates in the real space and σ

$$\sigma = \frac{p}{p_s}, \tag{4.1}$$

where p and p_s are pressure and surface pressure, respectively, as vertical coordinate. In the horizontal direction, all dependent variables are expanded in a truncated series of spherical harmonic functions $Y_l^m(\lambda, \theta)$. Here $Y_l^m(\lambda, \theta) = X_l^m(\sin\theta)e^{im\lambda}$, where $X_l^m(\theta)$ are the associated Legendre functions of the first kind of order m and degree l (m is the zonal wave number and $m - l$ is the number of meridional nodes from

pole to pole). For real to spectral space transform, any variable, say $f(\lambda, \theta, \sigma, t)$, is assumed to be expanded in a truncated series of spherical harmonics

$$f(\lambda, \theta, \sigma, t) = \sum_{m=-J}^{J} \sum_{l=|m|}^{|m|+J} f_l^m(\sigma, t) Y_l^m(\lambda, \theta), \tag{4.2}$$

if rhomboidal truncation is used or

$$f(\lambda, \theta, \sigma, t) = \sum_{m=-J}^{J} \sum_{l=|m|}^{J} f_l^m(\sigma, t) Y_l^m(\lambda, \theta), \tag{4.3}$$

if triangular truncation is used (all experiments in the following chapters are carried out using a triangular truncation), where $f_l^m(\sigma, t)$ is the spectral coefficients and J denotes the rhomboidal or triangular truncation limit.

The rhomboidal type of truncation is so named because it allows an equal number of meridional degrees of freedom, J, for each of the J zonal (wavenumber) degrees of freedom. For triangular truncation all mode would have to satisfy $-J \leq m \leq J$, $l \leq J$. There are several advantages to the triangular truncation, particularly in the calculation of energetics and energy spectra. However, the superiority of triangular truncation with respect to the integration of a complicated numerical model has never been demonstrated. Moreover, the rhomboidal truncation is much simpler to code, particularly if it is desired to take full advantage of symmetry in integrating a hemispheric model.

Spectral harmonics are orthogonal, with the normalization constant being incorporated into the Legendre functions such that (for arbitrary zonal wave numbers m, n and meridional indices l, j):

$$\frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{0}^{2\pi} Y_l^m(\lambda, \theta) (Y_j^n(\lambda, \theta))^t \cos\theta d\lambda d\theta = \delta_l^j \delta_m^n, \tag{4.4}$$

where $(Y_j^n(\lambda, \theta))^t = (-1)^n Y_j^{-n}$ is the complex conjugate of $Y_j^n(\lambda, \theta)$ and δ_l^j, δ_m^n are Kronecker deltas. By choosing our spectral expansion functions to be orthogonal, we greatly simplify the application of the Galerkin weighted residual procedure. In

principle, one could use any set of expansion functions; the Galerkin procedure only requires that the residue incurred by the spectral approximation of governing equations be orthogonal to those of the functions [125].

Given the condition imposed by Eq. (4.4) we can determine the spectral expansion coefficients $f_l^m(\sigma, t)$ in (4.2) via

$$f_l^m(\sigma, t) = \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\pi} f(\lambda, \theta, \sigma, t)(Y_j^n(\lambda, \theta))^t \cos\theta d\lambda d\theta. \tag{4.5}$$

It should be noted that if $f(\lambda, \theta, \sigma, t)$ is an explicit function of the dependent variables in the governing equations, then the notation f_l^m is used for the left-hand side of Eq. (4.5).

4.2 Model equations

It is important to notice that $\sigma = 0$ at the top of the atmosphere and $\sigma = 1$ is at the earth's surface. We will employ the usual boundary conditions $\dot{\sigma} = 0$ at $\sigma = 1$ and $\sigma = 0$.

It is convenient to define the following operators before developing the governing equations:

$$\hat{F}^\sigma = \int_\sigma^1 F d\sigma; \hat{F} = \int_0^1 F d\sigma. \tag{4.6}$$

The real space forms of the governing equations of the FSUGSM are as follows. This model uses the differentiated forms of the unfiltered primitive equations of motion and other model equations are the thermodynamic equation, continuity equation, hydrostatic equation, and a moisture equation, as follows:

Vorticity equation

$$\frac{\partial \zeta}{\partial t} = -\nabla \cdot (\zeta + f)\vec{V} - \vec{k} \cdot \nabla \times (RT\nabla q + \dot{\sigma}\frac{\partial \vec{V}}{\partial \sigma} - F), \tag{4.7}$$

Divergence equation

$$\begin{aligned}\frac{\partial D}{\partial t} &= \vec{k}\cdot\nabla\times(\zeta+f)\vec{V} \\ &- \nabla\cdot(RT\nabla q+\dot{\sigma}\frac{\partial\vec{V}}{\partial\sigma}-F)-\nabla^2(\Phi+\frac{\vec{V}\cdot\vec{V}}{2}),\end{aligned} \tag{4.8}$$

Thermodynamic equation

$$\frac{\partial T}{\partial t}=-\nabla\cdot\vec{V}T+TD+\dot{\sigma}\gamma-\frac{RT}{C_p}(D+\frac{\partial\dot{\sigma}}{\partial\sigma})+H_T, \tag{4.9}$$

Surface pressure tendency equation

$$\frac{\partial q}{\partial t}=-D-\frac{\partial\dot{\sigma}}{\partial\sigma}-\vec{V}\cdot\nabla q, \tag{4.10}$$

Hydrostatic equation

$$\sigma\frac{\partial\Phi}{\partial\sigma}=-RT, \tag{4.11}$$

and Moisture equation

$$\begin{aligned}\frac{\partial S}{\partial t} &= -\nabla\cdot\vec{V}S+SD-\dot{\sigma}\frac{\partial S}{\partial\sigma}+H_T-H_M \\ &- [\frac{RT}{C_p}-\frac{RT_d^2}{\epsilon L(T_d)}][D+\frac{\partial\dot{\sigma}}{\partial\sigma}-\frac{\dot{\sigma}}{\sigma}],\end{aligned} \tag{4.12}$$

where

ζ = vertical component of vorticity = $\vec{k}\cdot\nabla\times\vec{V}$,

f =Coriolis parameter,

$\vec{V}$= horizontal vector wind,

D = horizontal divergence =$\nabla\cdot\vec{V}$,

T = the absolute temperature,

$q=\ln p_s$,

γ = static stability =$\frac{RT}{C_p\sigma}-\frac{\partial T}{\partial\sigma}$,

$\dot{\sigma}$= vertical motion in sigma coordinates

$=(\sigma-1)(\hat{D}+\hat{V}\cdot\nabla q)+\hat{D}^{\sigma}+\hat{V}^{\sigma}\cdot\nabla q$,

Φ= geopotential height,

F= the horizontal frictional force per unit mass,

H_T =the diabatic heating,

R= the gas constant for dry air,

C_p =the specific heat of dry air at constant pressure,

T_d =the dew point temperature,

$S=T-T_d$ is the dew point depression,

ϵ=the ratio of the molecular weight of water vapor

to effective molecular weight of dry air (0.622),

$L(T_d)$ =the latent heat of vaporization of water or ice,

H_M =represents moisture sources or sinks.

The moisture Eq. (4.12) results from the combination of the thermodynamic Eq. (4.9) with an expression for the time rate of change of the dewpoint temperature [40].

To facilitate the application of the spectral transforms and the implementation of the time differencing scheme, some manipulation of the basic Eqs. (4.7) through (4.12) will be performed. First, let us introduce some notation. For eastward (zonal) and northward (meridional) wind components u and v, respectively, we define

$$\begin{aligned} U &= u\cos\theta/a \\ V &= v\cos\theta/a, \end{aligned} \tag{4.13}$$

where a is the earth's radius. With this transformation, the wind components become true scalars, with no discontinuities at the north or south poles [176]. Also, the

horizontal average over a global σ-surface is defined as,

$$\bar{F}^H(\sigma,t) = \frac{1}{4\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_0^{2\pi} F(\lambda,\theta,\sigma,t)\cos\theta d\lambda d\theta,, \tag{4.14}$$

For implementation of the semi-implicit algorithm (discussed below), the temperature is expressed as an initial horizontal (global) mean plus a deviation; hence,

$$T(\lambda,\theta,\sigma,t) = T^*(\sigma,t) + T'(\lambda,\theta,\sigma,t),, \tag{4.15}$$

where $T^*(\sigma,t) = \bar{T}^H(\sigma,t)$ at $t=0$. Likewise, we also have for static stability factor $\gamma = \gamma^* + \gamma'$. Furthermore, let us define a horizontal differential operator for two functions A and B:

$$\alpha(A,B) = \frac{1}{\cos^2\theta}[\frac{\partial A}{\partial\lambda} + \cos\theta\frac{\partial B}{\partial\theta}]. \tag{4.16}$$

Now let us expand the governing Eqs. (4.7–4.12) from vector into scalar (spherical coordinate) form, make use of Eqs. (4.16) and (4.6) and the definitions of U, V, T^*, T', γ^* and γ'. With these manipulations, the governing Eqs. (4.7–4.12) become:

$$\frac{\partial\zeta}{\partial t} = -\alpha(A,B), \tag{4.17}$$

$$\frac{\partial D}{\partial t} = \nabla^2(\Phi + RT^*q) = \alpha(B,-A) - a^2\nabla^2 E, \tag{4.18}$$

$$\frac{\partial T}{\partial t} - \gamma^*\dot{\sigma} - \frac{RT^*}{C_p}\frac{\partial q}{\partial t} = -\alpha(UT', VT') + B_T, \tag{4.19}$$

$$\frac{\partial q}{\partial t} + \hat{G} + \hat{D} = 0, \tag{4.20}$$

$$\frac{\partial S}{\partial t} = -\alpha(US, VS) + B_S, \tag{4.21}$$

and the hydrostatic Eq. (4.11) is unchanged. Here,

$$A = (\zeta+f)U + \dot{\sigma}\frac{\partial V}{\partial\sigma} + \frac{RT'}{a^2}\cos\theta\frac{\partial q}{\partial\theta} - \cos\theta\frac{F_\theta}{a}, \tag{4.22}$$

$$B = (\zeta+f)V - \dot{\sigma}\frac{\partial U}{\partial\sigma} - \frac{RT'}{a^2}\frac{\partial q}{\partial\lambda} + \cos\theta\frac{F_\lambda}{a}, \tag{4.23}$$

$$G = \frac{1}{\cos^2\theta}[U\frac{\partial q}{\partial\lambda} + V\cos\theta\frac{\partial q}{\partial\theta}], \tag{4.24}$$

$$
\begin{aligned}
E &= \frac{(U^2+V^2)}{2\cos^2\theta}, && (4.25)\\
B_T &= T'D + \gamma'\dot{\sigma} - \frac{RT'}{C_p}(\hat{G}+\hat{D}) + \frac{RT}{C_p}G + H_T, && (4.26)\\
\dot{\sigma} &= (\sigma-1)(\hat{G}+\hat{D}) + \hat{G}^{\sigma} + \hat{D}^{\sigma}, && (4.27)\\
B_S &= SD - \dot{\sigma}\frac{\partial S}{\partial \sigma} \\
&+ [\frac{RT}{C_p} - \frac{RT_d^2}{\epsilon L(T_d)}][\frac{\partial \dot{\sigma}}{\partial \sigma} + G - \hat{G} - \hat{D}] + H_T - H_M, && (4.28)
\end{aligned}
$$

F_θ and F_λ are the northward and eastward components of the frictional force, respectively. ∇^2 is the horizontal Laplacian operator in spherical coordinates.

We have transferred some of the terms to the left-hand side in Eqs. (4.18), (4.19) and (4.20). These terms will be treated implicitly when the semi-implicit algorithm is introduced. Following Robert *et al.* [176] we introduce two new variables, P and W:

$$P = \Phi + RT^*q, \tag{4.29}$$

$$W = \dot{\sigma} - \sigma(\hat{G}+\hat{D}), \tag{4.30}$$

From the hydrostatic law (4.11), we can used Eq. (4.29) to express T in term of P:

$$T = q\sigma\frac{\partial T^*}{\partial \sigma} - \frac{\sigma}{R}\frac{\partial P}{\partial \sigma}. \tag{4.31}$$

With these new variables, we will be able to isolate those terms in the governing equations which must be time averaged in the semi-implicit time differencing scheme while simplifying the implementation of this scheme in the prediction. Substituting these new variables into the left-hand side of the divergence, thermodynamic and continuity equations, one obtains

$$\frac{\partial D}{\partial t} + \nabla^2 P = \alpha(B, -A) - a^2\nabla^2 E, \tag{4.32}$$

$$\sigma\frac{\partial^2 P}{\partial t \partial \sigma} + R\gamma^* W = R\alpha(UT', VT') - RB_T, \tag{4.33}$$

and

$$\frac{\partial q}{\partial t} - W_S = 0. \tag{4.34}$$

We now have one more variable than the number of equations and we therefore generate a diagnostic equation for W from the definitions of $\dot{\sigma}$ and W.

$$\frac{\partial W}{\partial \sigma} + D = B_W, \tag{4.35}$$

where $B_W = -G$.

Thus, the dependent variables of the model are now ζ, S, D, P and q, which are governed by Eqs. (4.17), (4.21), (4.32), (4.33) and (4.34).

4.3 Spectral form of the governing equations

Now we shall cast the governing equations of the model into the spectral space. D, T, Φ, W and S are expanded in the same manner as (4.2). The variable q is similarly expanded except that the q_l^m are functions of time only. In the expansion of U and V there is one extra component for each m, in order to be consistent with the expansions for ζ and D [57].

$$U = \sum_{m=-J}^{J} \sum_{l=|m|}^{|m|+J+1} U_l^m Y_l^m, \tag{4.36}$$

$$V = \sum_{m=-J}^{J} \sum_{l=|m|}^{|m|+J+1} V_l^m Y_l^m, \tag{4.37}$$

The fields of U and V are derived from stream function, ψ, and the velocity potential, χ. Since

$$\zeta = \nabla^2 \psi; \quad D = \nabla^2 \chi, \tag{4.38}$$

we can express these relations spectrally as

$$\psi_l^m = -\frac{a^2}{l(l+1)} \zeta_l^m, \quad \chi_l^m = -\frac{a^2}{l(l+1)} D_l^m. \tag{4.39}$$

U and V are related to ψ and χ through

$$U = \frac{1}{a^2}(\frac{\partial \chi}{\partial \lambda} - \cos\theta \frac{\partial \psi}{\partial \theta}); \quad V = \frac{1}{a^2}(\frac{\partial \psi}{\partial \lambda} + \cos\theta \frac{\partial \chi}{\partial \theta}), \tag{4.40}$$

Using these relations and the properties of spherical harmonics, it can be shown that

$$l(l+1)U_l^m = -(l+1)\epsilon_l^m \zeta_{l-1}^m + l\epsilon_{l+1}^m \zeta_{l+1}^m - imD_l^m, \tag{4.41}$$

$$l(l+1)V_l^m = (l+1)\epsilon_l^m D_{l-1}^m - l\epsilon_{l+1}^m D_{l+1}^m - im\zeta_l^m, \tag{4.42}$$

where

$$\epsilon_l^m = \sqrt{\frac{l^2 - m^2)}{(4l^2-1)}}, \tag{4.43}$$

and $U_0^0 = \epsilon_1^0 \zeta_1^0$, $V_0^0 = -\epsilon_1^0 D_1^0$ are special cases.

The diagnostic relationships involving P, Φ T and q can be obtained from hydrostatic Eq. (4.31):

$$T_l^m = -\frac{\sigma}{R}\frac{\partial P_l^m}{\partial \sigma} + \sigma \frac{\partial T^*}{\partial \sigma}, \tag{4.44}$$

and

$$T_l^m = -\frac{\sigma}{R}\frac{\partial \Phi_l^m}{\partial \sigma}. \tag{4.45}$$

Now the spectral form of the governing equations can be written as, for all m, l:

$$\frac{\partial \zeta_l^m}{\partial t} = -\{\alpha(A,B)\}_l^m, \tag{4.46}$$

$$\frac{\partial D_l^m}{\partial t} - a^{-2}l(l+1)P_l^m = \{\alpha(B,-A) - a^2\nabla^2 E\}_l^m, \tag{4.47}$$

$$\sigma\frac{\partial^2 P_l^m}{\partial t \partial \sigma} + R\gamma^* W_l^m = R\{\alpha(UT', VT') - B_T\}_l^m, \tag{4.48}$$

$$\frac{\partial W_l^m}{\partial \sigma} + D_l^m = \{B_W\}_l^m, \tag{4.49}$$

$$\frac{\partial q_l^m}{\partial t} - \{W_S\}_l^m = 0. \tag{4.50}$$

$$\frac{\partial S_l^m}{\partial t} = \{-\alpha(US, VS) + B_S\}_l^m, \tag{4.51}$$

Integrals on the right-hand side of Eqs. (4.46) to (4.51) all fall into one of three categories, typified by

$$\{B_T\}_l^m = \frac{1}{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_0^{2\pi} B_T (Y_j^m)^t \cos\theta d\lambda d\theta. \tag{4.52}$$

$$\{a^2\nabla^2 E\}_l^m = -\frac{l(l+1)}{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_0^{2\pi} \frac{U^2+V^2}{2\cos\theta}(Y_j^m)^t d\lambda d\theta, \tag{4.53}$$

$$\{\alpha(A,B)\}_l^m = \frac{1}{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_0^{2\pi} \frac{1}{\cos\theta}[\frac{\partial A}{\partial\lambda} + \cos\theta\frac{\partial B}{\partial\theta}](Y_j^m)^t d\lambda d\theta. \tag{4.54}$$

Now B_T, E, A, B and $\alpha(A,B)$ are all nonlinear expressions. The integrals of type (4.52), (4.53) and (4.54) are calculated using the methods of Eliassen *et al.* [57] and Orszag [156]. That is all variables (ζ, D, U, V, T, S, $\partial q/\partial\lambda$, and $\cos\theta\partial q/\partial\theta$) required in the calculation of the right-hand side of Eq. (4.46) to (4.51) are first synthesized onto the real space transform grid. From the real space form of the variables, the nonlinear expressions, A, B, E, $T'U$, $T'V$, SU, SV, B_W, B_T and B_S, can be calculated. The integrals on the right-hand side of Eq. (4.46) to (4.51) are then calculated by exact numerical quadrature. Integrals of type (4.52) and (4.53) are straightforward, but integrals of type (4.54) employ an integration by parts in the manner of Bourke [12, 13]. Thus if

$$A = \sum_{m=-J}^{J} A_m e^{im\lambda}; \quad B = \sum_{m=-J}^{J} B_m e^{im\lambda}, \tag{4.55}$$

then

$$\{\alpha(A,B)\}_l^m = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} [imA_m X_l^m - B_m\cos\theta\frac{\partial X_l^m}{\partial\theta}]\frac{d\theta}{\cos\theta}, \tag{4.56}$$

where A_m and B_m are the Fourier coefficients of A and B, respectively.

4.3.1 Grid to spectral transform

For a given field f, we define its Fourier coefficients at a given latitude as

$$f^m(\theta) = \frac{1}{2\pi}\int_0^{2\pi} f(\theta,\lambda)e^{-im\lambda}d\lambda, \tag{4.57}$$

which can be evaluated using a discrete Fourier transform, provided f is a trigonometric polynomial. In general, if $f(x)$ is a trigonometric polynomial of degree not exceeding $J-1$,

$$\int_0^{2\pi} f(x)dx = \frac{2\pi}{J}\sum_{j=0}^{J-1} f(\frac{2\pi j}{J}), \tag{4.58}$$

is an exact evaluation of the integral [100]. Since the model's variables are assumed to have a spherical harmonic expansion, they are represented in the zonal direction by a trigonometric polynomial of degree J. Quadratic terms will contain powers of at most $2J$, and the integrand in Eq. (4.57) will be of degree not exceeding $3J$. For exact integration in the zonal direction, we therefore require at least $3J+1$ points around a latitude circle.

The integration in the meridional direction is performed by Gaussian quadrature [100]. That is,

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} g(\theta)\cos\theta d\theta = \int_{-1}^{1} g(x)dx = \sum_{j=1}^{N} W_j g(x_j), \tag{4.59}$$

where N and x_j are the number of the roots and the roots of the ordinary Legendre polynomial X_N^0, respectively. Eq. (4.59) is an exact integration of the function g, provided that $g(x)$ is a polynomial of degree not exceeding $2N-1$, and the function g is known at the Gaussian points x_j.

Using Eqs. (4.57) and (4.59), the grid to spectral transform may be written as

$$f_l^m = \int_{-1}^{1} f^m(x)X_l^m(x)dx = \sum_{j=1}^{N} W_j f^m(x_j)X_l^m(x_j). \tag{4.60}$$

For rhomboidal truncation, we must have $N \geq (5J+1)/2$ while for triangular truncation, $N \geq (3J+1)/2$. For an arbitrary resolution, the maximum power of $f^m(x)X_l^m(x)$ must be evaluated, and the condition imposed by the Gaussian quadrature requirement must be applied.

4.3.2 Spectral to grid transform

The computation of grid values from spectral expansion given by Eq. (4.2) or Eq. (4.3) is carried out in the following two steps,

$$f^m(x_j) = \sum_{l=|m|}^{N(m)} f_l^m X_l^m(x_j), \tag{4.61}$$

and

$$f(\theta_j, \lambda) = \sum_{m=-J}^{J} f^m(x_j) e^{im\lambda}, \tag{4.62}$$

where $N(m) = J+|m|$ for rhomboidal truncation and $N = J$ for triangular truncation.

The grid used in the present model was designed following the above principles, however it should be realized that higher than quadratic (i.e. cubic) order terms are involved, which results in some aliasing. For a $T42$ mode, the Gaussian (transform) grid consists of 128 equally spaced (2.8125° increment) longitudes beginning at the Greenwich Meridian and 64 Gaussian latitudes (the roots of X_{64}^0). These latitudes are tabulated in Table (4.1); note that they are symmetric with respect of equator and are nearly equally spaced (roughly a 2.76° increment).

4.4 The semi-implicit algorithm

The semi-implicit formulation follows Robert *et al.* [176]. In Eqs. (4.46) to (4.51), time derivatives are replaced by centered time differences (except at initial time where forward differences are applied), i.e. $\partial f/\partial t$ are approximated by $[f(t+\Delta t) - f(t-\Delta t)]/(2\Delta t)$ where Δt is the time step. The remaining terms on the left hand sides are handled implicitly by the application of the time averaging operator ($\overline{\ }^t$) defined by $\overline{f}^t = [f(t+\Delta t) + f(t-\Delta t)]/2$, while the right-hand-sides are calculated explicitly.

Two equations remain fully explicit. They are the vorticity and moisture equations:

$$\zeta_l^m(t+\Delta t) = -2\Delta t\{\alpha(A,B)\}_l^m + \zeta_l^m(t-\Delta t), \tag{4.63}$$

Table 4.1: Gaussian latitudes of the T-42 model. + indicates north, and − south.

±2.14°	±24.42°	±46.75°	±69.07°
±4.90°	±27.21°	±49.54°	±71.86°
±7.69°	±30.00°	±52.33°	±74.65°
±10.47°	±32.79°	±55.12°	±77.44°
±13.26°	±35.58°	±57.91°	±80.23°
±16.05°	±38.37°	±60.70°	±83.02°
±18.84°	±41.16°	±63.49°	±85.81°
±21.63°	±43.96°	±66.28°	±88.60°

and

$$S_l^m(t+\Delta t) = 2\Delta t\{-\alpha(US, VS) + B_S\}_l^m + S_l^m(t-\Delta t). \tag{4.64}$$

The remaining equations are handled in a semi-implicit manner using the $(\overline{\ }^t)$ operator:

$$\overline{D_l^m}^t - a^{-2}\Delta t l(l+1)\overline{P_l^m}^t = \Delta t\{\alpha(B, -A) - a^2\nabla^2 E\}_l^m + D_l^m(t-\Delta t), \tag{4.65}$$

$$\sigma\overline{\frac{\partial P_l^m}{\partial\sigma}}^t + \Delta t R\gamma^*\overline{W_l^m}^t = \Delta t\{R\alpha(UT', VT') - RB_T\}_l^m + \sigma\frac{\partial P_l^m}{\partial\sigma}(t-\Delta t), \tag{4.66}$$

$$\overline{q_l^m}^t - \Delta t\overline{W_{Sl}}^{m\,t} = q_l^m(t-\Delta t), \tag{4.67}$$

$$\overline{\frac{\partial W_l^m}{\partial\sigma}}^t + \overline{D_l^m}^t = \{B_w\}_l^m. \tag{4.68}$$

The integrals on the right hand side are all evaluated at time t. Some of the terms in the integrals are forcing terms and should really be calculated at time $t-\Delta t$ to avoid the growth of a computational mode. In the case of non-linear forcing terms this would necessitate the transform from spectral to real space of many $t-\Delta t$ variables.

This is undesirable from the point of view of computational efficiency, so we have chosen to calculate certain forcing terms at time t and suppress the computational mode with a very weak time filter [7].

In the manner of Robert *et al.* [176], $\overline{D_l^m}^t$ is eliminated between Eqs. (4.65) and (4.68). Then $\overline{W_l^m}^t$ is eliminated between the resulting equation and Eq. (4.66) giving a diagnostic equation for $\overline{P_l^m}^t$

$$\frac{\partial}{\partial\sigma}\frac{\sigma}{\gamma^*}\frac{\overline{\partial P_l^m}^t}{\partial\sigma} - \frac{R\Delta t^2}{a^2}l(l+1)\overline{P_l^m}^t = \frac{\partial}{\partial\sigma}\Big\{\frac{C_T}{\gamma^*}\Big\}_l^m + \{C_D\}_l^m, \tag{4.69}$$

where

$$\{C_T\}_l^m = R\Delta t\{\alpha(UT', VT') - B_T\}_l^m + \sigma\frac{\partial P_l^m}{\partial\sigma}(t-\Delta t), \tag{4.70}$$

and

$$\begin{aligned}\{C_D\}_l^m &= R\Delta t\{B_W - \Delta t\alpha(B,-A) + a^2\Delta t\nabla^2 E\}_l^m \\ &+ R\Delta t D_l^m(t-\Delta t).\end{aligned} \tag{4.71}$$

Thus, Eqs. (4.65) to (4.68) have been reduced to a two-point boundary value problem for each horizontal mode. If $\{C_T\}_l^m$ and $\{C_D\}_l^m$ are known at time t, the $\overline{P_l^m}^t$ can be calculated for each m, l provided the boundary conditions at $\sigma = 0$ and $\sigma = 1$ are given. The remaining variables $\overline{D_l^m}^t$, $\overline{W_l^m}^t$, $\overline{W_{Sl}^m}^t$, $\overline{q_l^m}^t$ can be calculated by back substitution into Eqs. (4.65) to (4.68) and the same variables at $t+\Delta t$ can then be calculated from the definition of the $(\overline{\ }^t)$ operator. Discussion of the appropriate boundary condition for the solution of Eq. (4.69) will be delayed until the vertical discretization has been introduced in the next section.

It is appropriate at this time to comment briefly upon the semi-implicit algorithm just derived, since it might appear that certain non-linear terms are being handled implicitly. Thus, at first sight, it appears that the non-linear term G is being included on the left hand (implicit) side of Eq. (4.33) because of the definition for W given in Eq (4.30). This is not so, as was pointed out by Asselin [7], because the equation for

W that is actually used in the model is Eq. (4.35), in which the non-linear term G is on the right-hand (explicit) side. Thus, the introduction of the variable W does not allow the implicit treatment of some non-linear terms, but it is convenient, as will be demonstrated subsequently.

4.5 Vertical discretization

The vertical finite-difference scheme is a somewhat more general form of the scheme used by Robert *et al.* [176]. The basic feature of the scheme is that the temperatures are carried at levels intermediate to the levels of the geopotentials. We will dwell principally on the finite-difference analogues to the left-hand-sides of Eqs. (4.65)-(4.69). The vertical finite-differencing involved in the right-hand-sides of these equations is conventional and will not be discussed in length. It should be stated, though, that logarithmic vertical differencing has generally been used, in particular

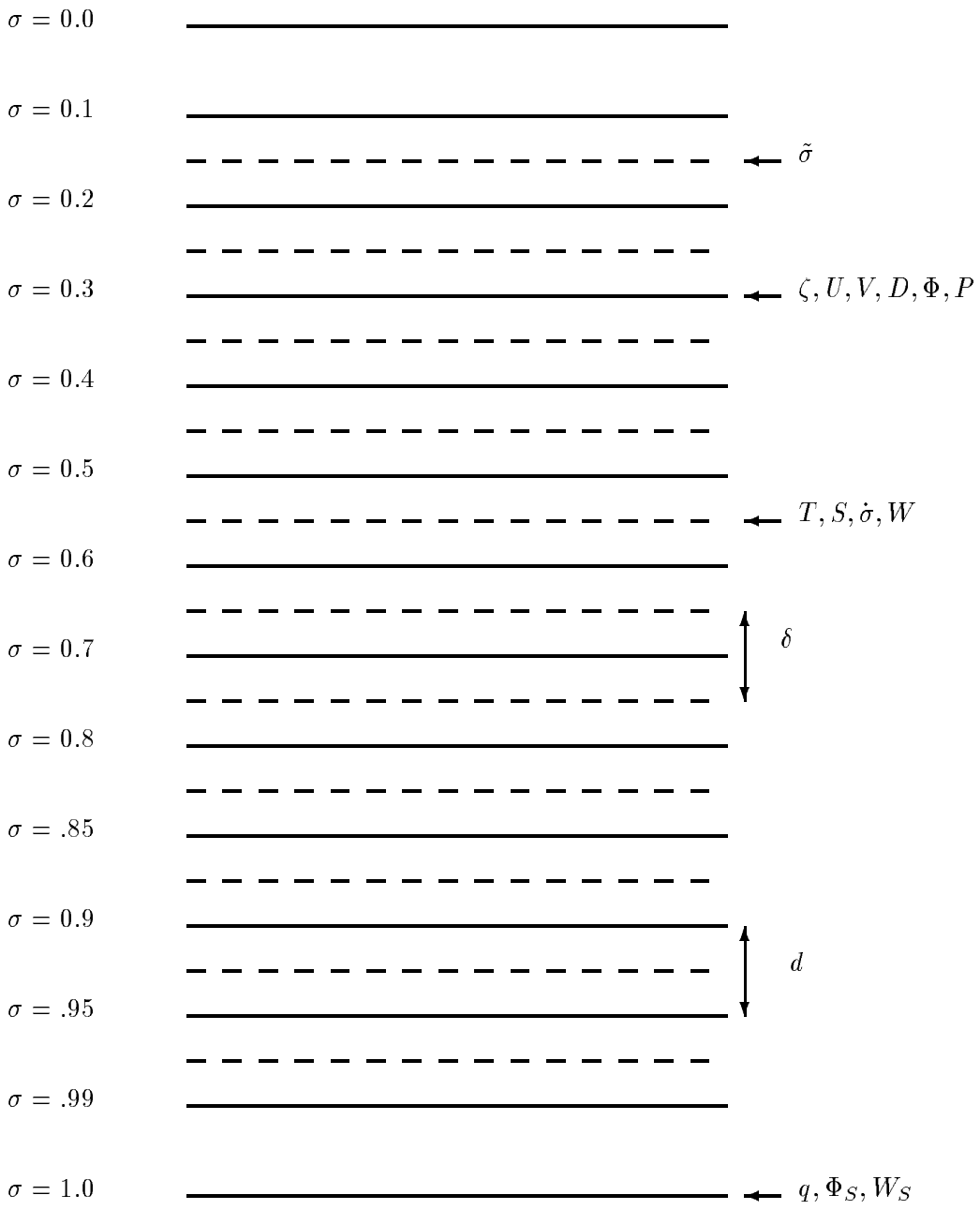

Figure 4.1: Schematic representation of the vertical structure of the model. The solid lines represents the levels σ_n. The dashed lines represents the layers $\tilde{\sigma}_n$. $\sigma = 0$ is at the top of the atmosphere and $\sigma = 1$ is at the earth's surface.

with respect to the thermodynamic variables. To simplify the notation, the subscript l and superscript m will be assumed. Thus, f_l^m is written as f, and $\{f\}_l^m$ as $\{f\}$. The levels (in σ) of the FSUGSM are placed as shown in Fig. 4.1. This is a twelve-level model. Note (in Fig. 4.1) the vertical staggering of the levels, where ζ, D and P (hence, also U , V and Φ) are defined at levels shown by solid lines (here after referred to as the σ-levels) while T, S and W (also $\dot{\sigma}$) are defined at levels indicated by dashed lines (hence after referred to as the $\tilde{\sigma}$-levels). Second order accuracy for the hydrostatics demands that the $\tilde{\sigma}$-levels be located midway between the σ-levels in the $\ln \sigma$ coordinate, thus

$$\ln \tilde{\sigma}_k = \frac{\ln \sigma_k + \ln \sigma_{k+1}}{2}, \tag{4.72}$$

which implies

$$\tilde{\sigma}_k = \sqrt{\sigma_k \sigma_{k+1}}, \qquad k = 1, \ldots, N \tag{4.73}$$

where the subscripts refer to the level in question and N, the index of the bottom model level, is equal to 12. Note also in Fig. 4.1 that the increment of the σ-levels is 0.1 above $\sigma = 0.8$ and becomes smaller below $\sigma = 0.8$, which affords greater vertical resolution in the planetary boundary layer.

The thermodynamic equation is applied at intermediate levels. These levels are obtained from a finite difference form of the hydrostatic equation. The following finite-difference approximation to the hydrostatic equation (4.47) is second-order accurate, provided that the layer temperatures $\tilde{T}_n$ are defined as the geometric mean of the adjacent σ-level values, i.e. by

$$\tilde{T}_n = \frac{\Phi_n - \Phi_{n+1}}{R d_n}, \qquad n = 1, \ldots, N, \tag{4.74}$$

where

$$\Phi_n = \Phi(\sigma = \sigma_n), \qquad d_n = \ln(\sigma_{n+1}/\sigma_n). \tag{4.75}$$

We will hereafter refer to the σ-values, defined by Eq. (4.73) where the thermodynamic equation is applied, as the layer values. The tilde ($\sim$) will be used to indicate which variables are carried in the layers ($\sigma = \tilde{\sigma}_n$); these are $\tilde{T}_n$, $\tilde{S}_n$, and $\tilde{W}_n$. Note that $\tilde{W}_N$ is not carried; instead we use W_s which is applied at a level $\sigma_{N+1} = 1$ and therefore appears without a tilde.

It is convenient to define vertical increments in σ; i.e.

$$\tilde{\delta}_n = \tilde{\sigma}_{n+1} - \tilde{\sigma}_n, \qquad n = 1, \ldots, N-2. \tag{4.76}$$

Both the top and bottom increments are special cases. Thus

$$\tilde{\delta}_0 = \tilde{\sigma}_1 \qquad \tilde{\delta}_{N-1} = 1 - \tilde{\sigma}_{N-1}. \tag{4.77}$$

Now $P_n = \Phi_n + RT_n^* q$ is defined on the σ_n levels, whereas the temperatures $\tilde{T}_n$ and thus $\tilde{T}_n^*$ are carried in the $\tilde{\sigma}_n$ layers. We obtain T_n^* and thus P_n from $\tilde{T}_n^*$ and q by simply taking the logarithmic average of the adjacent layer temperatures $\tilde{T}_n^*$ and $\tilde{T}_{n+1}^*$. To second order accuracy

$$T_n^* = \frac{d_{n-1}\tilde{T}_n^* + d_n\tilde{T}_{n-1}^*}{d_{n-1} + d_n}, \qquad 2 \le n \le N. \tag{4.78}$$

T_1^* can not be calculated in this manner and is instead simply obtained by linear extrapolation.

$$T_1^* = 1.5\tilde{T}_1^* - 0.5T_2^*, \tag{4.79}$$

The T_n' are obtained in a similar manner.

The static stability γ^* can be calculated at the $\tilde{\sigma}_n$ layers simply as

$$\tilde{\gamma}_n^* = \frac{1}{\tilde{\sigma}_n}\Big(\frac{R\tilde{T}_n^*}{C_p} - \frac{T_{n+1}^* - T_n^*}{d_n}\Big), \qquad n = 1, \ldots, N \tag{4.80}$$

where $T_{N+1}^* = T_s^*$. In the absence of an equation for T_s, it has been assumed that $\tilde{\gamma}_N^* = \tilde{\gamma}_{N-1}^*$. This, in effect, formally defines T_s^* in terms of $\tilde{T}_N^*$ and T_N^*. A similar expression and assumption is used to relate $\tilde{\gamma}_n'$ to $\tilde{T}_n'$, T_n' and T_{n+1}'.

Vertically discretized analogues of the vertical integration operators ($\hat{}^{\sigma}$) and ($\hat{}$) defined in Eq. (4.6) are required for the calculation of $\hat{G}$, $\hat{D}$, $\hat{G}^{\sigma}$, $\hat{D}^{\sigma}$ and ultimately $\dot{\sigma}$ from $\dot{\sigma} = (\sigma - 1)(\hat{D} + \hat{V} \cdot \nabla q) + \hat{D}^{\sigma} + \hat{V}^{\sigma} \cdot \nabla q$ and Eq. (4.24) at $\tilde{\sigma}_n$ layers. The vertical integral $\hat{F}^{\sigma}$ is approximated by $\hat{F}^{n}$, a simple quadrature extending from the surface ($\sigma = 1$) to $\tilde{\sigma}_n$. $\hat{F}$ is approximated by $\hat{F}^{0}$, the same quadrature extending from $\sigma = 1$ to $\sigma = 0$. Thus, if F is a level variable, the approximation for $\hat{F}^{\sigma}$ at $\sigma = \tilde{\sigma}_n$ is

$$\hat{F}^{n} = \sum_{k=n+1}^{N} F_k \tilde{\delta}_{k-1}, \qquad n = 1, \ldots, N-1 \tag{4.81}$$

and the approximation for $\hat{F}$ is

$$\hat{F}^{0} = \sum_{k=1}^{N} F_k \tilde{\delta}_{k-1}. \tag{4.82}$$

In this manner it is possible to obtain the quadrature analogues of $\hat{G}$, $\hat{D}$, $\hat{G}^{\sigma}$, $\hat{D}^{\sigma}$ and thus $\dot{\sigma}$ at each of the layers σ_n, $1 \leq n \leq N-1$. At σ_N we have assumed $\tilde{\sigma_N} = h\tilde{\sigma_{N-1}}$ where h is an empirical constant $0 \leq h \leq 1$.

Similar approximations are used for any remaining terms on the right-hand-sides of Eqs. (4.65-4.69).

All that remains to be discussed is the solution of the boundary value problem (4.69). The consistent vertically discretized form of this equation for $2 \leq n \leq N-1$ is

$$\frac{1}{\tilde{\delta}_{n-1}}\left[\left(\frac{\overline{P_{n+1}}^{t} - \overline{P_n}^{t}}{\tilde{\gamma}^{*}_{n} d_n}\right) - \left(\frac{\overline{P_n} - \overline{P_{n-1}}^{t}}{\tilde{\gamma}^{*}_{n-1} d_{n-1}}\right)\right] - \frac{R\Delta t^2}{a^2} l(l+1)\overline{P_n}^{t} = \frac{1}{\tilde{\delta}_{n-1}}\left[\frac{1}{\tilde{\gamma}^{*}_{n}}\{\tilde{C_T}\}_n - \frac{1}{\tilde{\gamma}^{*}_{n-1}}\{\tilde{C_T}\}_{n-1}\right] + \{C_D\}_n, \tag{4.83}$$

where, as mentioned previously, n refers to level or layer, ($\tilde{}$) indicates a layer value, and $\{\ \}$ refers to a horizontal integration (Eq. (4.5)).

Eq. (4.83) yields $N-2$ equations in the N unknowns $\overline{P_n}^{t}$, $1 \leq n \leq N$. To close the system it is necessary to obtain two additional equations involving respectively $\overline{P_1}^{t}$ and $\overline{P_N}^{t}$.

The top boundary condition is straightforward to apply. At $\sigma = 0$, $\dot{\sigma} = 0$, which implies $W = 0$ in the finite difference analogue of Eq. (4.68) applied at σ_1. This leads to the following equation relating $\overline{\tilde{W}_1}^t$ to $\overline{D_1}^t$:

$$\frac{\overline{\tilde{W}_1}^t}{\tilde{\delta}_0} + \overline{D_1}^t = \{B_w\}_1. \tag{4.84}$$

Combining the above equations with finite difference analogues of equations (4.65) and (4.66), applied at σ_1 and $\tilde{\sigma}_1$ respectively, yields an equation involving $\overline{P_1}^t$ and $\overline{P_2}^t$ only:

$$\frac{1}{\tilde{\delta}_0}\left(\frac{\overline{P_2}^t - \overline{P_1}^t}{\tilde{\gamma}_1^* d_1}\right) - \frac{R\Delta t^2}{a^2} l(l+1)\overline{P_1}^t = \frac{\{\tilde{C}_T\}}{\tilde{\delta}_0 \tilde{\gamma}_1^*} + \{C_D\}_1. \tag{4.85}$$

To close the system at the bottom boundary is less straightforward. Recalling the definition of W and assumptions made for $\tilde{\gamma}_N$ and $\tilde{\sigma_N}$ and noting that Φ_S is time invariant, it can be shown that the finite difference analogue of Eq. (4.66) degenerates into the following form when applied at $\tilde{\sigma_N}$.

$$\frac{-\overline{P_N}^t}{\tilde{\lambda}_N^* d_N} + \frac{R\Delta t}{\tilde{\lambda}_N^*} h\tilde{\gamma}_{N-1}^* \overline{\tilde{W}_{N-1}}^t + R\Delta t \overline{\tilde{W}_S}^t = \frac{\{\tilde{C}_T\}_N}{\tilde{\lambda}_N^*}, \tag{4.86}$$

where

$$\tilde{\lambda}_N^* = \frac{R\tilde{T}_N^*}{C_p} + \frac{T_N^*}{d_N} - h\tilde{\gamma}_{N-1}^* \tilde{\sigma}_{N-1}, \tag{4.87}$$

The finite difference analogues of Eqs. (4.65) and (4.68) applied at σ_N yield a relation between $\overline{W_S}^t$, $\overline{\tilde{W}_{N-1}}^t$ and $\overline{P_N}^t$. A third relation between $\overline{P_{N-1}}^t$, $\overline{P_N}^t$ and $\overline{\tilde{W}_{N-1}}^t$ can be obtained by applying the finite difference analogue of Eq. (4.66) at $\tilde{\sigma_{N-1}}$. This yields 3 equations in 4 unknowns $\overline{P_{N-1}}^t$, $\overline{P_N}^t$ and $\overline{\tilde{W}_{N-1}}^t$ and $\overline{W_S}^t$ which can be used to obtain a single equation in $\overline{P_{N-1}}^t$ and $\overline{P_N}^t$; i.e.

$$\begin{aligned}\frac{1}{\tilde{\delta}_{N-1}}\left[\frac{-\overline{P_N}^t}{\tilde{\lambda}_N^* d_N} - \left(\frac{\overline{P_N} - \overline{P_{N-1}}^t}{\tilde{\lambda}_{N-1}^* d_{N-1}}\right)\right] - \frac{R\Delta t^2}{a^2} l(l+1)\overline{P_N}^t = \\ \frac{1}{\tilde{\delta}_{N-1}}\left[\frac{1}{\tilde{\lambda}_N^*}\{\tilde{C}_T\}_N - \frac{1}{\tilde{\lambda}_{N-1}^*}\{\tilde{C}_T\}_{N-1}\right] + \{C_D\}_N,\end{aligned} \tag{4.88}$$

where

$$\tilde{\lambda}^*_{N-1} = \frac{\tilde{\gamma}^*_{N-1}\tilde{\lambda}^*_N}{\tilde{\lambda}^*_N + h\tilde{\gamma}^*_{N-1}}. \tag{4.89}$$

Eqs. (4.83), (4.85) and (4.88) define a $N \times N$ tridiagonal matrix relationship for each horizontal mode (m, l) of the $\overline{P_n}^t$, $1 \leq n \leq N$. These matrix problems are easily solved by an efficient tridiagonal matrix algorithm of the Thomas algorithm variety. Once the $\overline{P_n}^t$ have been calculated, back substitution into the vertically discretized forms of Eqs. (4.65) and (4.66) yields $\overline{D_n}^t$ and $\overline{\tilde{W}_n}^t$, respectively. Substitution of $\overline{P_n}^t$ and $\overline{\tilde{W}_n}^t$ into Eq. (4.86) yields $\overline{W_S}^t$, from which $\overline{q}^t$ can be obtained using Eq. (4.67). The definition of the $(\overline{\ }^t)$ operator then yields $P_n(t+\Delta t)$, $D_n(t+\Delta t)$ and $q(t+\Delta t)$. The remaining variables $\Phi_n(t+\Delta t)$ and $\tilde{T}_n(t+\Delta t)$ are calculated diagnostically using vertically discretized forms of Eqs. (4.44) and (4.45). The explicit prognostic Eqs. (4.63) and (4.64) yield $\zeta_n(t+\Delta t)$ and $S_n(t+\Delta t)$.

This formulation of the semi-implicit method does have the advantage of yielding a tridiagonal matrix in the vertical at the expense of some conceptual complication. However, the extra calculations per time step due to the semi-implicit calculation are completely negligible.

The physical processes included in the FSUGSM are described in Appendix F.

CHAPTER 5

4-D VARIATIONAL DATA ASSIMILATION WITH THE FSU GLOBAL SPECTRAL MODEL

5.1 Introduction

The numerical feasibility of variational data assimilation has been extensively demonstrated for 2-D models, but has only recently been applied to 3-D models [11, 38, 54, 157, 168, 169, 170, 171, 233]. The objective of the current chapter is to demonstrate the feasibility of 4-D variational data assimilation applied to an adiabatic version of the FSUGSM. The number of components of the control variables vector (the initial conditions vector) is 303104. As stated before the numerical cost of the 4-D VDA originates both from the model integration and the gradient calculation of the cost function with respect to the control variables. With present numerical models, numerical algorithms and presently available computer power, the only practical way to carry out the 4-D VDA is through an appropriate use of the adjoint techniques in calculating the gradient of the cost function with respect to the control variables [35]. The adjoint techniques were developed in the fields of optimal control theory of PDEs of distributed parameters [116], and their application to 4-D VDA requires to obtain the gradient of the cost function with respect to the control variables which is obtained by one backwards integration of the adjoint model in time. Therefore, we first develop the TLM and the adjoint model of a dry version (i.e., no physical effects are involved except horizontal diffusion) of the FSUGSM and verify their accuracies in Section 5.2. The dry version of the FSUGSM is used because we do not want to add, at this stage,

the complexity inherent in the inclusion of effects of highly nonlinear and threshold on-off diabatic physical parameterizations, which are still at the research stage [216, 231, 236]. The numerical algorithms employed in 4-D VDA and related scaling issues are discussed in Section 5.3. In Section 5.4, various variational assimilation experiments are conducted. The ability of 4-D VDA to retrieve observational fields is investigated starting from either shifted or randomly perturbed initial conditions. We evaluate the impacts of different observations distributed in space and time as well as the impact of the horizontal diffusions. Through different scenarios of sets of observations, we demonstrate the efficiency of 4-D VDA in extracting the information contained in the dynamics of the model together with information in the observations. Conclusions as well as topics for further research related to the 4-D VDA with FSUGSM are presented in Section 5.5.

5.2 Model developments and numerical validation

5.2.1 The cost function and its gradient with respect to control variables

The cost function will be defined in physical space rather than in spectral space because it is clearer and easier to implement proper scaling in physical space than in spectral space. Therefore the cost function is still defined by Eq. 1.3 where W is an $N \times N$ diagonal weighting matrix with $W_{\vec{\zeta}}$, $W_{\vec{D}}$, $W_{\vec{S}}$, $W_{\vec{q}}$, and $W_{\vec{T}}$ as its diagonal sub-matrices entries of weighting factors for vorticity, divergence, moisture, surface pressure, and temperature fields, respectively. Here $N = L(4K + 1)$ is the number of components of the state vector $\vec{X}$, K is the number of vertical levels and L is the number of Gaussian grid points. $\vec{X} = (\vec{\zeta}, \vec{D}, \vec{S}, \vec{q}, \vec{T})^t$ denotes the state vector, where $\vec{\zeta}, \vec{D}, \vec{S}, \vec{q}, \vec{T}$ contain values of vorticity, divergence, moisture, surface pressure, and temperature over all grid points on all levels at time t. $\vec{X}^o = (\vec{\zeta}^o, \vec{D}^o, \vec{S}^o, \vec{q}^o, \vec{T}^o)^t$, denotes the observation vector where $\vec{\zeta}^o, \vec{D}^o, \vec{S}^o, \vec{q}^o, \vec{T}^o$ are observational values of

vorticity, divergence, moisture, surface pressure, and temperature over all grid points on all levels at time t. F is a vector function with five components being the right hand sides of vorticity, divergence, thermodynamic, surface pressure tendency and moisture equations (4.7 – 4.7 and 4.12).

The 4-D variational data assimilation problem (1.1) is usually solved by using large-scale unconstrained minimization algorithms such as LBFGS of Liu and Nocedal [117], truncated-Newton [130], the adjoint truncated-Newton [219] and quasi-Newton algorithms. All these minimization algorithms require the calculation of the gradient of the cost function with respect to control variables. The current studies will obtain the gradient of the cost function by using the adjoint model techniques. Other alternatives for obtaining the gradient of the cost function such as automatic differentiation ([55], [69]), analytical evaluation and finite-difference (FD) approximations exist. However, they are either not feasible for this optimal control problem applied to VDA or are too costly to apply.

The adjoint model of Eqs. (1.10) and (1.10) is defined as

$$-\frac{\partial \bar{X}}{\partial t} = (\frac{\partial F}{\partial \vec{X}})^* \bar{X}, \tag{5.1}$$

$$\bar{X}(t_f) = U'', \tag{5.2}$$

where $\bar{X}$ is the adjoint variable and $(\cdot)^*$ denotes the complex conjugate transpose.

The FOA model of Eqs. (1.10) and (1.11) is defined by Eqs. (2.2) and (2.3). Clearly the FOA model given by Eq. (2.2) can be obtained by adding a forcing term to the adjoint model given by Eq. (5.1) and setting the final condition (5.2) to zero.

One can prove that the gradient of the cost function with respect to the initial conditions is equal to the value of the FOA model variable at the initial time

$$\nabla_U J = P(t_0), \tag{5.3}$$

(see Section 1.3 and [149, 202, 218]).

5.2.2 The accuracy of the tangent linear model

The TLM model of the adiabatic FSUGSM was developed and its validity was verified. The TLM model is the result of the linearization of the direct adiabatic FSUGSM in the vicinity of a given model trajectory [113]. If we view the direct adiabatic FSUGSM as, M, the result of multiplication of a number of matrices:

$$M = A_1 A_2 ... A_N, \tag{5.4}$$

where each matrix A_i $(i = 1, ..., N)$ represents either a subroutine or a single do loop, then the TLM model can be viewed as Tlm

$$Tlm = B_1 B_2 ... B_N, \tag{5.5}$$

where $B_i = A_i$ if A_i is linear and B_i is a linearization of A_i in the vicinity of a basic state of $\vec{X}$ if A_i is nonlinear for $i = 1, ..., N$.

Suppose that U_1' and U_2' are two initial conditions for the TLM model then

$$B_1 B_2 ... B_N (U_1' + U_2') = B_1 B_2 ... B_N (U_1') + B_1 B_2 ... B_N (U_2'). \tag{5.6}$$

Eq. (5.6) can be used as a preliminary criterion for the correctness-check of the TLM model.

For a given state variable U, and a perturbation, U' on U, the real evolution of the solution due to a perturbation $\alpha U'$ is given by

$$M(U + \alpha U') - M(U) \tag{5.7}$$

where α is a scalar and $M(U)$ represents the numerical integration of the direct adiabatic FSUGSM starting from the initial conditions U. If $Tlm(\alpha U')$ represents the numerical integration of the TLM models starting from the initial conditions $\alpha U'$, then according to the definition of the TLM one obtains

$$\psi(\alpha) = \frac{\|M(U + \alpha U') - M(U)\|}{\|Tlm(\alpha U')\|} = 1. + O(\alpha \|U'\|), \tag{5.8}$$

where $\|\cdot\|$ denotes the Euclidean norm of "$\cdot$".

We take the fields of vorticity, divergence, moisture, surface pressure, and temperature at 0 UTC 1 June 1988 as U, and choose $U' = \alpha U$ where α is small scalar but not too close to the machine accuracy. The direct adiabatic FSUGSM and TLM models are integrated for an hour with various α values. Fig 5.1 indicates that the TLM model produced a good asymptotic approximation to the evolution of perturbations for the forecast period when α varies from 10^{-7} to 1. It is surprising that even for perturbations with magnitudes similar to that of U, the TLM model still yields a relatively good approximation to the evolution of perturbations for the forecast period.

When the forecast period is increased to 6 hours, the results shown in Fig 5.1 are still correct. For instance, when $\alpha = 0.1$, $\psi(\alpha) = 0.9995405246$. Therefore the TLM is correct for reasonable long forecast period and for perturbations with magnitudes similar to that of the initial condition of the nonlinear model. So the correctness of the TLM model has been verified.

5.2.3 The accuracy of the first order adjoint model

The adjoint model of Eq. (5.5) can be obtained by taking its complex conjugate i.e.

$$Tlm^t = B_1^t B_2^t ... B_N^t. \tag{5.9}$$

In this way, the discrete adjoint model can be directly derived from the discrete linear model, which in turn is derived from the FSUGSM by linearization around a basic state. This simplifies not only the complexity of constructing the adjoint model but also avoids the inconsistency generally arising from the derivation from the adjoint equations in analytic form followed by the discrete approximation.

The correctness of the adjoint model can be checked in three ways. First, at any level of the code, the adjoint code can be checked by using the following identity

$$(BQ)^t(BQ) = Q^t[B^t(BQ)], \tag{5.10}$$

where B and Q represent any code and its input in the discrete TLM model, respectively. For example, if $B = Tlm$, the initial condition Q of the TLM model is chosen as the basic state $\vec{X}$ at 0 UTC 1 June 1988, the basic states at later times are obtained by integrating the FSUGSM starting with the same initial condition Q and all models are integrated for six hours, then $B^t(BQ)$ denotes the backwards integration of the adjoint model given by Eq. (5.1) with final condition taken as the output of the TLM model integration, then one obtains

$$(BQ)^t(BQ) = 1.0189576063747 \times 10^{13}, \tag{5.11}$$

$$Q^t[B^t(BQ)] = 1.0189576063753 \times 10^{13}, \tag{5.12}$$

which prove the correctness of the discrete adjoint model given by Eq. (5.9) up to the machine's precision of 14 digits.

Second, the TLM describes the evolutions of perturbations in a forecast model. Eq. (5.10) indicates that the accuracy of the TLM is the same as that of its adjoint model given by Eq. (5.9) or (5.1)-(5.1). This can also be demonstrated intuitively by the following experiment.

Let observations consist of a full-model-state vector at the end of the assimilation period, then the cost function assumes the following form

$$J(U) =< W(\vec{X}(t_f) - \vec{X}^o(t_f)), (\vec{X}(t_f) - \vec{X}^o(t_f)) > . \tag{5.13}$$

If we denote $\Delta J(U)$ as a first order approximation to $J(U + U') - J(U)$ due to the change U' in the initial condition U, and Tlm as the numerical integrations of the

TLM, then

$$\begin{aligned} J(U+U') - J(U) &\approx \Delta J(U) = \frac{\partial J(\vec{X}(t_f))}{\partial \vec{X}(t_f)} \cdot \Delta \vec{X}(t_f) \\ &\approx \frac{\partial J(\vec{X}(t_f))}{\partial \vec{X}(t_f)} \cdot Tlm(U') \\ &= U'^t \cdot Tlm^t(\frac{\partial J(\vec{X}(t_f))}{\partial \vec{X}(t_f)}) = U'^t \cdot \nabla_U J, \end{aligned} \tag{5.14}$$

where the superscript t denotes a transpose and $\Delta\vec{X}(t_f)$ is the change in the state variable $\vec{X}$ at the final time t_f due to the change U' in the initial condition U, i.e. given U and U' then identical values of $\Delta J(U)$ are obtained whether the TLM or its adjoint model is applied, implying both values are equally accurate [58].

Still, it is worthwhile to provide an example of the accuracy of an adjoint calculation, especially since we will claim that the accuracy is much better than one would have expected.

Suppose that the cost function is defined as the difference between the model solution and the observation at the final time (six hours from the initial time) and the observation is taken as a full-state vector solution at a time which is twelve hours apart from the initial time. The weighting factor for divergence field is calculated according to the following formula

$$W_{\vec{D}} = \frac{1}{4\|\vec{D}(t_f) - \vec{D}^o(t_f)\|^2}, \tag{5.15}$$

with similar expressions for the moisture, temperature, surface pressure, and vorticity fields. The term $\partial J(\vec{X}(t_f))/\partial\vec{X}(t_f)$ in Eq. (5.14) is obtained by integrating the forward model and using the above specified weighting factors, i.e.

$$\frac{\partial J(\vec{X}(t_f))}{\partial \vec{X}(t_f)} = 2W(\vec{X}(t_f) - \vec{X}^o(t_f)), \tag{5.16}$$

The initial condition for the TLM model is taken as $\partial J(\vec{X}(t_f))/\partial\vec{X}(t_f)$, i.e.

$$U' = \frac{\partial J(\vec{X}(t_f))}{\partial \vec{X}(t_f)} \tag{5.17}$$

$\partial J(\vec{X}(t_f))/\partial \vec{X}(t_f)$ is also used as the final condition for the adjoint model given by Eqs. (5.1)-(5.1). If the perturbation is applied to an six hour forecast, then

$$\frac{\frac{\partial J(\vec{X}(t_f))}{\partial \vec{X}(t_f)} \cdot Tlm U'}{U'^t \cdot Tlm^t(\frac{\partial J(\vec{X}(t_f))}{\partial \vec{X}(t_f))}} = \frac{0.00040447305742735}{0.00040447305742624} = 1.000000000003, \tag{5.18}$$

which proves numerically that both the TLM and its adjoint model are correct and have the same accuracy.

Third, the verification of the correctness of the gradient can be conducted as described below. The Taylor formula applied in the direction $U' = \partial J(U)/\partial \vec{X}(t_f)$ yields

$$\phi(\alpha) = \frac{J(U + \alpha U') - J(U)}{\alpha U'^t \cdot \nabla_U J} = 1. + O(\alpha \|U'\|). \tag{5.19}$$

For small α but not too close to the machine accuracy, the value of $\phi(\alpha)$ should be close to one if the gradient $\nabla_U J$ is correctly calculated. Fig 5.2 shows that for α in the range between 10^{-3} and 10^{-14} an unit value of $\phi(\alpha)$ is obtained.

5.3 Descent algorithms and scaling

5.3.1 The LBFGS algorithm of Liu and Nocedal

Liu and Nocedal [117] compared the combined conjugate-gradient quasi-Newton method of Buckley and LeNir [14], the limited-memory quasi-Newton method described by Nocedal [154] (called LBFGS method), and the partitioned quasi-Newton method of Griewank and Toint [76]. Zou *et al.* [227] compared four of the state-of-the-art limited-memory quasi-Newton methods on several test problems, including problems in meteorology and oceanography. Their results show that the LBFGS algorithm is the most efficient and particularly robust amongst the algorithms tested for the set of problems considered here. Therefore the LBFGS algorithm is chosen for this study since it deals with the critical issue of storage in large-scale problems.

In the LBFGS algorithm, the user provides a sparse symmetric and positive definite matrix H_0, which approximates the inverse Hessian of the cost function, and specifies the number m of the quasi-Newton up-dates, therefore controls the amount of the storage required. The LBFGS algorithm updates the formula generated matrices which approximate the Hessian by building curvature information using information from the last m quasi-Newton iterations. After having used the m vector storage locations for m quasi-Newton updates of the Hessian, the quasi-Newton approximation of the Hessian matrix is updated by dropping the oldest update vector and replacing it with the newest update vector. A new line search direction, which is an estimate of the relative change to the current variables vector that produces the maximum reduction in the cost function, is then computed. It employs a cubic line search required to satisfy a Wolfe condition [224], and an unit step size is always tried first. For a precise description of the LBFGS algorithm, please see [117].

5.3.2 Weighting and scaling

Weights in the cost function have a serious impact on the minimization algorithm [26]. A poor set of weights may result in elongated contour lines of the cost function, or even in a total failure of the VDA. There are several ways to determine weights such as: (1) trial and error method [110], (2) the statistical method, i.e. taking the reciprocal of the variance of the observation errors [32], and (3) temporal weighting method [34]. The first method may be very costly since it depends on both luck and experience. In the second method, the variances are difficult to specify properly, and much more research work is necessary in this area. In the third method weights are assigned to individual observations and vary linearly with time. The total sum of the weights assumes the same value as the reciprocal of squared estimates of the statistical root-mean-square observational error.

In our experiments with FSUGSM, a diagonal weighting matrix

$$W = diag\{W_{\vec{\zeta}}, W_{\vec{D}}, W_{\vec{S}}, W_{\vec{q}}, W_{\vec{T}}\}, \tag{5.20}$$

is used where $W_{\vec{\zeta}}$, $W_{\vec{D}}$, $W_{\vec{S}}$, $W_{\vec{q}}$, and $W_{\vec{T}}$ are constant diagonal sub-matrices. The diagonal elements of $W_{\vec{D}}$ are calculated according to the following formula

$$(W_{\vec{D}})_{jj} = \frac{1}{4\sum_{i=0}^{M} \|\vec{D}(t_i) - \vec{D}^o(t_i)\|^2}, \tag{5.21}$$

with similar expressions for the moisture, temperature, surface pressure, and vorticity fields. Model solutions are obtained by integrating the model from an initial condition. Weights given by Eq. (5.21) nondimensionalize the cost function, and result in an unit cost function at the beginning of the variational data assimilation.

Scaling is a crucial issue in the success of nonlinear unconstrained optimization problems, and considerable research has been carried out on scaling nonlinear problems. It is well known that a badly scaled nonlinear programming problem can be almost impossible to solve [149, 34]. An effective automatic scaling procedure would ease these difficulties and could also render problems that are well scaled easier to solve by improving the condition number of their Hessian matrix [205]. Thus scaling is a crude way of carrying out preconditioning. For more discussion about preconditioning, please see Appendix C and references [134, 236]

In the FSUGSM, the variables in the control vector have enormously different magnitudes varying over a range of eight orders of magnitude. Typical values of vorticity, divergence, natural log of the surface pressure and temperature have the magnitudes of 10^{-5}, 10^{-6}, 10^{-2} and 100, respectively. Scaling by variable transformation converts variables from units that reflect the physical properties to units that display desirable properties for minimization process. Given a diagonal scaling matrix $S = diag\{S_{\vec{\zeta}}, S_{\vec{D}}, S_{\vec{S}}, S_{\vec{q}}, S_{\vec{T}}\}$ where $S_{\vec{\zeta}}$, $S_{\vec{D}}$, $S_{\vec{S}}$, $S_{\vec{q}}$, and $S_{\vec{T}}$ are constant diagonal sub-matrices, the general scaling procedure may be written as

$$\vec{X} = S\vec{X}^s, \tag{5.22}$$

$$\vec{g}^s = S\vec{g}, \tag{5.23}$$

$$H^s = SHS, \tag{5.24}$$

where H is the Hessian matrix. The constant diagonal elements of sub-matrix $S_{\vec{D}}$ will be calculated by

$$\max_{i,j} |D_j(t_i) - D_j^o(t_i)|, \tag{5.25}$$

and similarly for $S_{\vec{\zeta}}$, $S_{\vec{S}}$, $S_{\vec{q}}$, and $S_{\vec{T}}$ where $D_j(t_i)$ and $D_j^o(t_i)$ are the j-th components of vectors $\vec{D}$ and $\vec{D}^o$ at time t_i, respectively.

For complicated functions, difficulties may be encountered in choosing suitable scaling factors. There is no general rule to determine the best scaling factors for all minimization problems, and good scaling is problem dependent. A basic rule is that the variables of the scaled problem should be of similar magnitude and of order of unity because with in optimization routines convergence tolerances and other criteria are necessarily based on an implicit definition of "small" and "large", and, thus variables with widely varying orders of magnitude may cause difficulties for some minimization algorithms [73]. One simple direct way to determine the scaling factor is to use the typical values for different fields. For some problems, the relative magnitudes of the gradient field components of the cost function differ from each other by several orders of magnitude. In such cases, it is better to scale the gradient of the cost function [73, 115].

5.4 Numerical results of the variational data assimilation

Although the accuracies of the TLM and the adjoint model as well as the gradient calculation of the cost function with respect to the control variables have been verified, it is of primary importance to ensure that the system as a whole was properly constructed and that the minimization of the cost function can be performed using

a reasonably high-resolution system. Hence, two sets of experiments were designed using the LBFGS large-scale unconstrained minimization algorithm of Liu and Nocedal [117]. Both shifted and randomly perturbed initial conditions are used in the first set of experiments to investigate the efficiency of the 4-D VDA as well as the impacts of the length of the assimilation on the convergence rate. In the second set of experiments, the impacts of distributed observations and that of horizontal diffusion are investigated.

In all experiments a horizontal truncation of T42 with 12 levels in the vertical direction is used. For the exact variable transformation from spectral space to physical grid space, or vice versa, a (128×64) Gaussian grid was used, and the dimension of the vector of control variables is 303104. The length of the time assimilation window is 6 hours unless indicated otherwise. All the routines are coded in single precision FORTRAN. The runs were made on the CRAY-YMP supercomputer at Florida State University, for which the relative machine precision ϵ is approximately 10^{-14}.

5.4.1 Numerical results with shifted initial conditions

The experiments are devised as follows: the model-generated values starting from an analysis of the real observational data at 6 UTC 1 June 1988 are used as observations and the initial guess is taken as an analysis of the real observational data at 0 UTC 1 June 1988. We know exactly what the solution is, and the value of the cost function at the minimum must be zero.

Here we assume the observations are available at every time step of the assimilation window and the cost function is defined by Eq. (1.3).

In the first experiment the length of the assimilation window is chosen as one hour. The purpose of this experiment is to verify that all components of the VDA work properly before we conduct any VDA for a longer window of assimilation.

Fig. 5.3 (solid line) displays the variations of the log of the scaled cost function (J_k/J_0) and scaled gradient norm ($\|\vec{g}_k\|/\|\vec{g}_0\|$) with the number of iterations in the minimization process. It would appear initially that the VDA works numerically. Indeed, after 65 iterations, the normalized cost function is reduced by 8 orders of magnitude while the normalized norm of the gradient is reduced by 4 orders of magnitude.

It is a good idea to consider the difference fields between the retrieved and reference (observational) fields before and after the minimization process, since these difference fields measure the ability of the VDA to retrieve the observational fields. Fig 5.11 shows the difference fields between the retrieved fields of divergence, vorticity, temperature and natural log of the surface pressure and the reference (observational) fields at the beginning of the assimilation window after the minimization process. Comparing with the same difference fields (Fig. 5.6) before the minimization process, it is observed that after the minimization the maximum difference values are reduced by four orders of magnitude in all difference fields of divergence, vorticity, temperature and natural log of the surface pressure.

If the length of the assimilation is increased to 6 hours, the VDA also works successfully. The results in Fig 5.3 (dash line) indicate that after 65 iterations, the normalized cost function is reduced by 2.3 orders of magnitude while the normalized norm of the gradient is reduced by 1.5 orders of magnitude.

In Fig 5.4, dotted and long dash lines denote the variation of the log of the rms error between the initial (guess) and exact temperatures with model levels at the beginning and end of time assimilation window before assimilation. The solid and dash lines denote the variation of the log of the rms error between the retrieved and exact temperatures with model levels at the beginning and end of time assimilation window after carrying out VDA. It is interesting to notice that the above rms error is reduced by one order of magnitude at the beginning of time assimilation window

while by 1.5 orders magnitude at the end of time assimilation window. The different reduction in rms errors indicates that our metric for the cost function is not optimal. The consequence is that the shape of the cost function can become strongly elliptic with respect to the metric used and the gradient of the cost function can even become almost orthogonal to the direction of the minimum.

The above experiments imply that the longer the assimilation period, the slower the convergence rate. The slower convergence rate of the minimization in the case of longer assimilation window results from the worsening of the conditioning of the Hessian matrix due to the longer model integration. Thus the use of a longer assimilation window will require more computation not only due to a longer model integration interval but also due to a slower convergence rate. There should exist an optimal assimilation length such that the minimization can retrieve the best initial condition from the observations and the dynamics in the model equations. An investigation in this direction will be very useful for operational implementation of 4-D VDA in NWP, along with selective time weighting of the observations.

5.4.2 Numerical results with randomly perturbed initial conditions

The experiments are devised as follows: the model-generated values starting from an initial condition taken as an analysis of the real observational data at 0 UTC 1 June 1988 are used as observations and the initial guess is obtained by adding a random perturbation of the above initial condition to itself. The random perturbations are obtained by using a standard library randomizer RANF on CRAY-YMP and the magnitudes of the perturbations are at most 10% of the original. Again, we know exactly what the solution is, and the value of the cost function at the minimum must be zero. All the random perturbations used in this chapter are of an uniform distribution.

As in the previous subsection we first consider a case where the length of the assimilation is only one hour and the observations are available at every time step. Fig. 5.8 shows that the VDA performs successfully. After 65 iterations, the normalized cost function is reduced by more than 8 orders of magnitude while the normalized norm of the gradient is reduced by 3.8 orders of magnitude.

From Figs. 5.9 and 5.10 it is observed that after the minimization the maximum difference values are reduced by four orders of magnitude in all difference fields of divergence, vorticity, temperature and natural log of the surface pressure. It is also observed that the reductions in the maximum difference values are different at the beginning and end of the assimilation window. That is the 4-D VDA is more capable to retrieve information in the observations and in the dynamics of the model at the end of the assimilation window for our test problem. It is clear that the 4-D VDA works equally well with either shifted or randomly perturbed initial conditions.

5.4.3 Impact of horizontal diffusion

We have mentioned that the FSUGSM is used in its adiabatic version without any physical parameterizations package. Only horizontal diffusion of vorticity, divergence and temperature terms is retained. The horizontal diffusion terms impact the minimization process [114, 207]. We describe now the impact of these terms on the VDA.

A Laplacian-type of horizontal diffusion is included in the prognostic equations for the vorticity, divergence and temperature fields, respectively. Thus the diffusion is of the form $K\nabla^2 Q$, where K is the diffusion coefficient and Q is vorticity or divergence or temperature. This diffusion is included primarily for numerical reasons, to suppress the growth of the amplitudes of higher wavenumber components during the forecast. This phenomenon, called spectral blocking, has been discussed by Machenhauer [125]. It is caused by neglecting the interactions between waves within the truncation limit

with those waves which lie outside the truncation limit. The diffusion is applied in spectral space, in an implicit manner (i.e. to the prognostic variables at time level $t + \Delta t$) to avoid the growth of a computational mode (linear instability). The numerical values of the diffusion coefficients for vorticity, divergence and temperature fields are 6.0×10^{15} and 6.0×10^{16} for weak and strong horizontal diffusions, respectively.

The same experiments as described in Section 5.4.4 where observations are available at every time step were performed again with strong horizontal diffusion terms and without horizontal diffusion terms. The experimental results obtained without the horizontal diffusion terms and with the strong horizontal diffusion terms present are very similar to these experimental results obtained with weak horizontal diffusion terms. The variations (Fig. 5.3) of the log of the scaled cost function (J_k/J_0) and scaled gradient norm ($\|\vec{g}_k\|/\|\vec{g}_0\|$) with the number of iterations in the minimization process are almost identical to these (also see Fig. 5.3) obtained when weak horizontal diffusion terms are applied to the model. Table 5.1 indicates that rms errors of temperature at different model levels when no diffusion, weak or strong horizontal diffusion terms are applied in the model, respectively. It is observed that the rms errors obtained when weak horizontal diffusion terms are applied in the model are uniformly smaller than these obtained with no diffusion or strong horizontal diffusion terms present in the model. When strong horizontal diffusion terms are applied in the model, the rms errors are the largest amongst the rms errors obtained in the three cases. The maximum and minimum errors occur at the top and the bottom levels, respectively.

All these results indicate that when weak horizontal diffusion terms are included in the adiabatic version of the FSUGSM the minimization yields a more accurate solution of the VDA problem, but when the horizontal diffusion terms become large, the minimization may yield a less accurate solution of the VDA problem (see Fig.5.3).

5.4.4 Impact of observations distributed in time

Here we will investigate the impact on the VDA of having a set of observations distributed in time. Indeed, this experiment allows us to evaluate the ability of 4-D VDA to use information contained in the dynamics.

The experiments carried out below follow the same pattern as those described in Subsection 5.4.1 except that the number of observations is different. The impacts of three different sets of observations, those available at every time step, those available at the beginning and end, and those available at the end of assimilation window are investigated.

If the observations are available at the beginning and end of the assimilation period, then the cost function assumes the following form

$$\begin{aligned} J(U) &= \frac{1}{2}\{W < \vec{X}_k(t_f) - \vec{X}^o(t_f), \vec{X}_k(t_f) - \vec{X}^o(t_f) > \\ &+ W < \vec{X}_k(t_0) - \vec{X}^o(t_0), \vec{X}_k(t_0) - \vec{X}^o(t_0) >\}, \end{aligned} \tag{5.26}$$

Fig 5.13 indicates that after 65 iterations, the normalized cost function and the normalized norm of the gradient are reduced by 3.5 and 1.7 orders of magnitude, respectively. These reductions are larger than the reductions observed in the case where the observations are available at every time step (See Fig. 5.3).

As shown in Fig 5.14, the rms errors between the retrieved and exact temperatures at different levels are reduced by two orders of magnitude at the beginning of time assimilation window and by slightly more than two orders of magnitude at the end of time assimilation window. Similar reductions at the beginning and end of time assimilation window indicate that the cost function defined by Eq. (5.26) leads to more balanced retrievals than the one defined by Eq. (1.3).

It is noted that in Figs. 5.12 and 5.6, after the minimization the maximum difference values of all difference fields of divergence, vorticity, temperature, and natural log of surface pressure are uniformly reduced by two orders of magnitude. The de-

creases in the maximum difference values are one order more than these observed in Fig. 5.5 where the observations are available at every time step.

We also carried out one experiment where the observations are available every two hours, i.e. the observations are available at times 0, 2, 4 and 6h UTC June 1 1988. The results (not shown) are very similar to the results when the observations are available at every time step. This indicates that the impact of having observations during the middle of the assimilation window is weak.

The above results imply that the more observations we add, the smaller is the saturation in the retrieval of the reference fields. This phenomenon has also been observed by Thépaut *et al.* [207].

5.5 Conclusions

4-D VDA with an adiabatic version of the FSUGSM T-42 is presented here. First, we demonstrated the numerical feasibility of the 4-D VDA with a realistic spectral model with horizontal resolution (triangular truncation) T-42 and 12 vertical levels. Second, we investigated both the impact of horizontal diffusion and that of observations distributed in time. The experimental results reported here together with the results of Navon *et al.* [141], Thépaut *et al.* [210] and Courtier and Talagrand [36] indicate the great potential of 4-D VDA for application to operational 3-D NWP models.

In this chapter, the weighting and scaling matrices were calculated from the observational fields and the initial guess fields. A reasonable reduction in cost function was achieved, and the quality of the retrieval (i.e., the retrieved initial conditions based on both observational and dynamical information) is satisfactory. The results

where horizontal diffusion terms are retained in the adiabatic FSUGSM yield an improvement of the accuracy of the retrieved initial condition since the diffusion terms suppress the growth of the amplitudes of higher wavenumber components.

Through different initial conditions and different scenarios of sets of observations distributed in time, we demonstrate the efficiency of the 4-D VDA in extracting the information contained in the dynamics of the model together with the information contained in the observations. Conditioning is an important factor for an operational implementation. The dynamics of the model may lead to different reductions of rms errors at different times and the loss of the conditioning of the problem, i.e. the shape of the cost function may become strongly elliptic with respect to the non-optimal metric used and the gradient can even become almost orthogonal to the direction of the minimum. The solution to this problem is to obtain an adequate knowledge of the structure of the Hessian of the cost function. It seems that the SOA application may be an efficient but very costly way to obtain information about the spectrum of eigenvalues of the Hessian of the cost function with respect to the control variables.

Many issues related to the 4-D VDA are currently under investigation, such as its formulation [37], the relationship of the 4-D VDA and sequential estimation methods [170] and the definition of the cost function etc. A report on the state of art of these research issues may be found in the proceeding of a workshop on variational assimilation, with special emphasis on three-dimensional aspects (ECMWF, U.K., 1993).

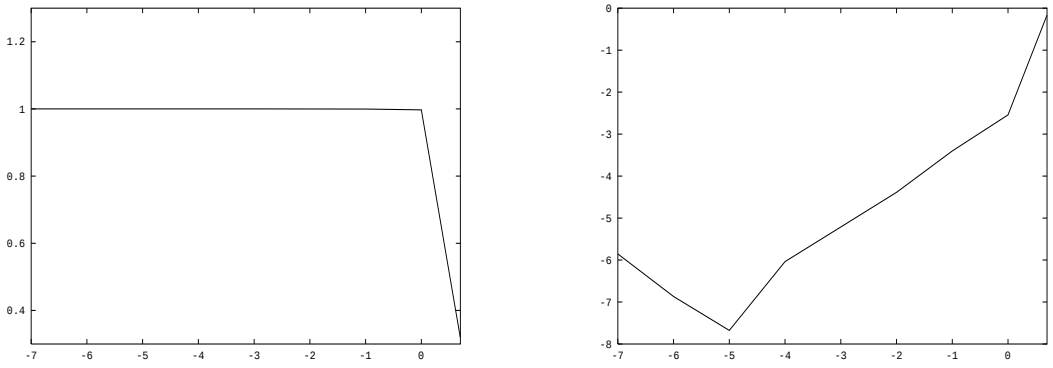

Figure 5.1: Verifications of the correctness of the tangent linear model: (a) variation of ψ with $\log \alpha$, (b) variation of $\log |\psi - 1|$ with $\log \alpha$.

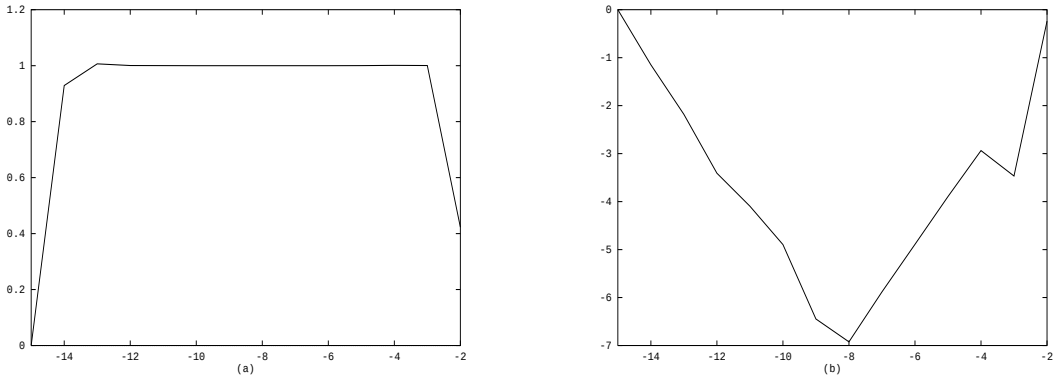

Figure 5.2: Verifications of the correctness of the gradient calculation using Taylor expansion: (a) variation of ϕ with $\log \alpha$, (b) variation of $\log |\phi - 1|$ with $\log \alpha$.

Figure 5.3: Variations of the log of the scaled cost function (J_k/J_0) and scaled gradient norm ($\|\vec{g}_k\|/\|\vec{g}_0\|$) with the number of iterations. The lengths of the assimilation window are one (solid line) and six (dash (with weak diffusion), dotted (with strong diffusion) and long dash (without diffusion) lines) hours and the observations are available at every time step.

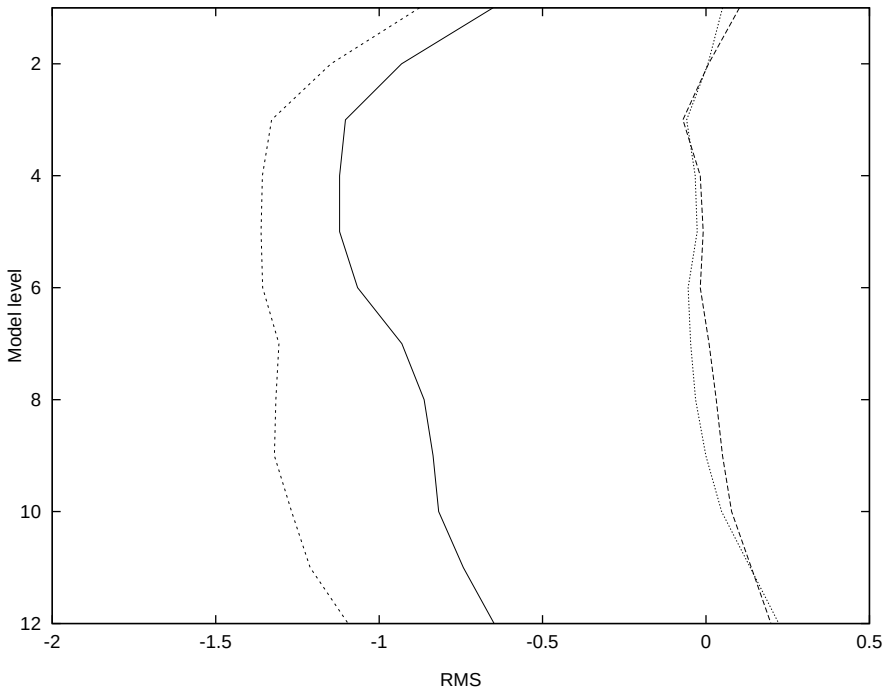

Figure 5.4: Dotted and long dash lines denote the log of rms errors between the initial (guess) and exact temperatures at the beginning and end of time assimilation window before variational data assimilation. Solid and dash lines denote the log of rms errors between the retrieved and exact temperatures at the beginning and end of time assimilation window after variational assimilation. The observations are available at every time step and the initial guess is the analysis of the observational fields at 0 UTC 1 June 1988.

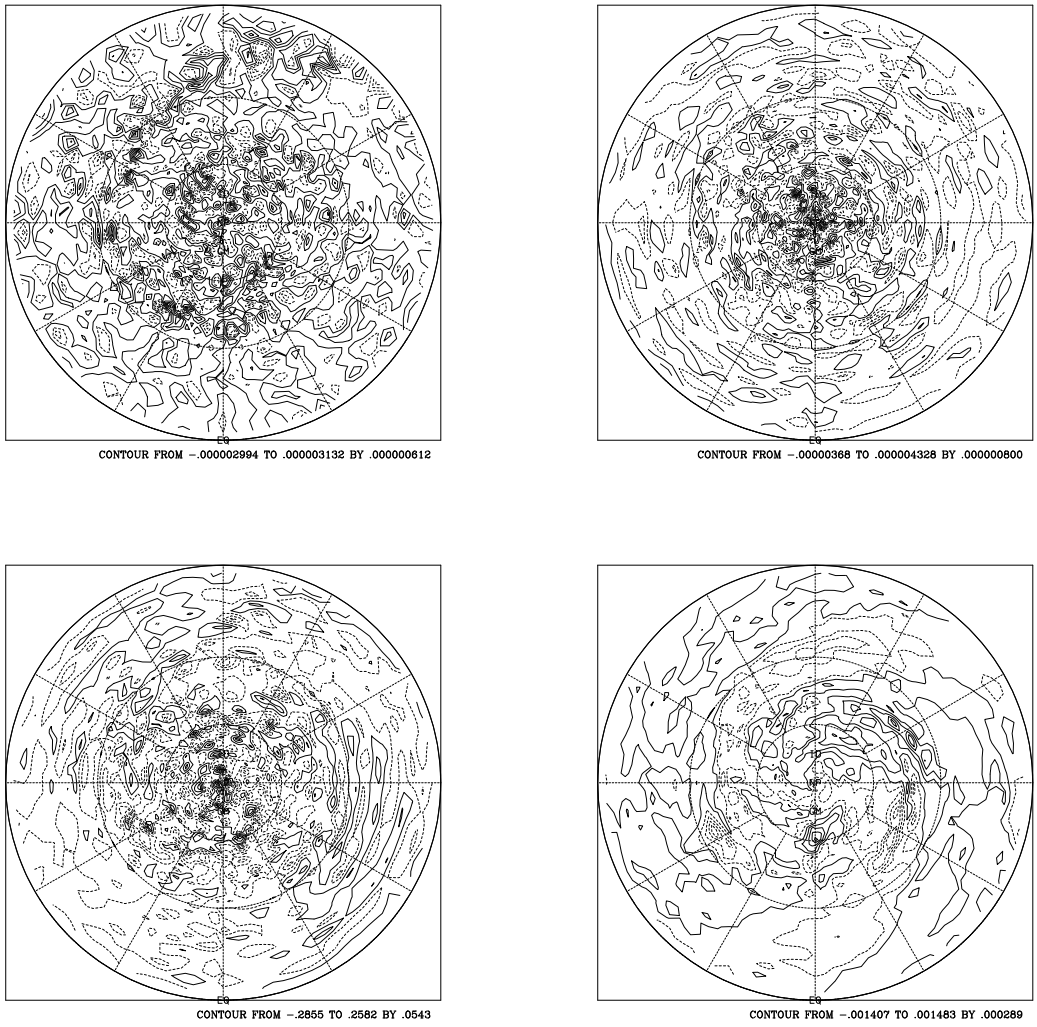

Figure 5.5: Divergence (top left), vorticity (top right), temperature (bottom left) natural log of surface pressure (bottom right) difference fields between the retrieved and reference fields at level 8. The reference and starting fields are the analysis of the observational fields of divergence, vorticity, temperature and natural log of surface pressure at 0 and 6 UTC 1 June 1988, respectively. The observations are available at every time step.

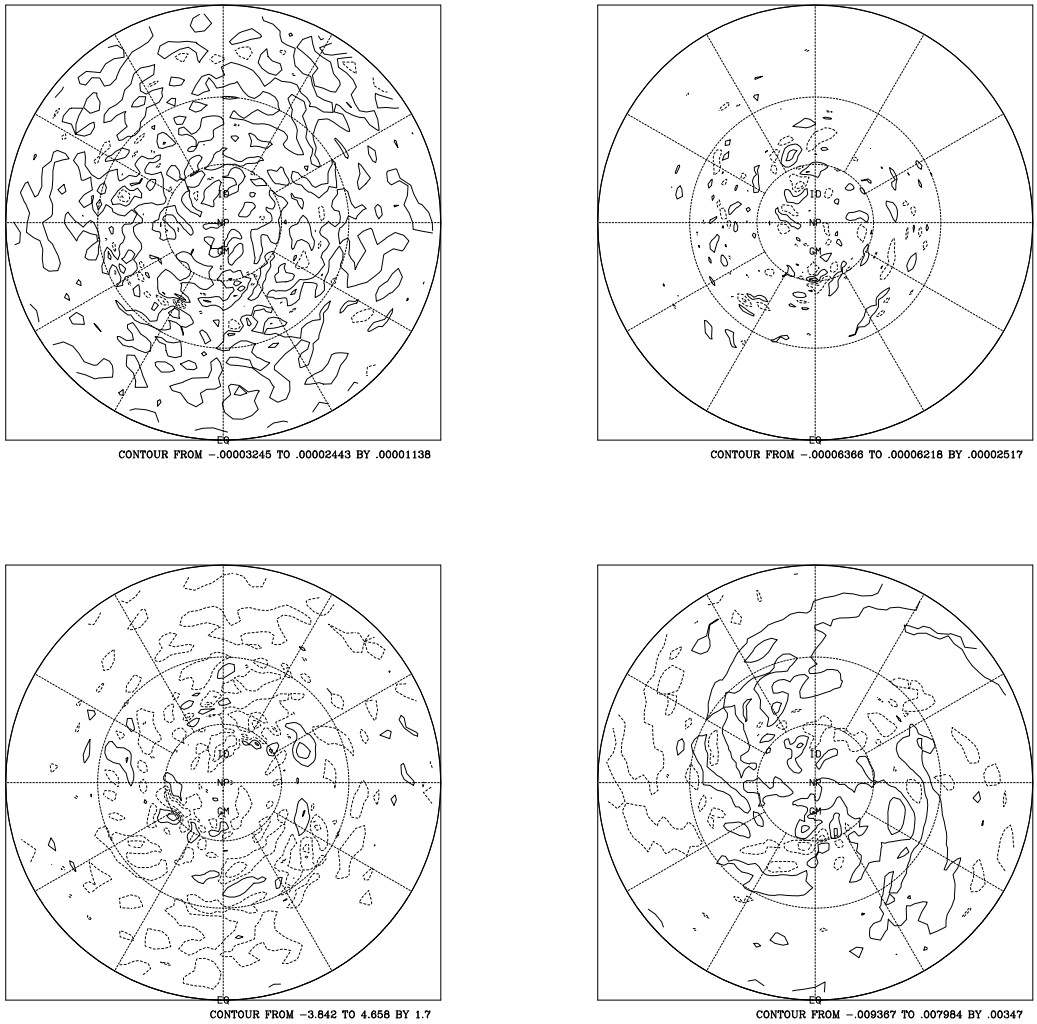

Figure 5.6: Divergence (top left), vorticity (top right), temperature (bottom left) natural log of surface pressure (bottom right) difference fields between starting and reference fields at level 8. The starting and reference fields are the analysis of the observational fields of divergence, vorticity, temperature and natural log of surface pressure at 0 and 6 UTC 1 June 1988, 1988, respectively.

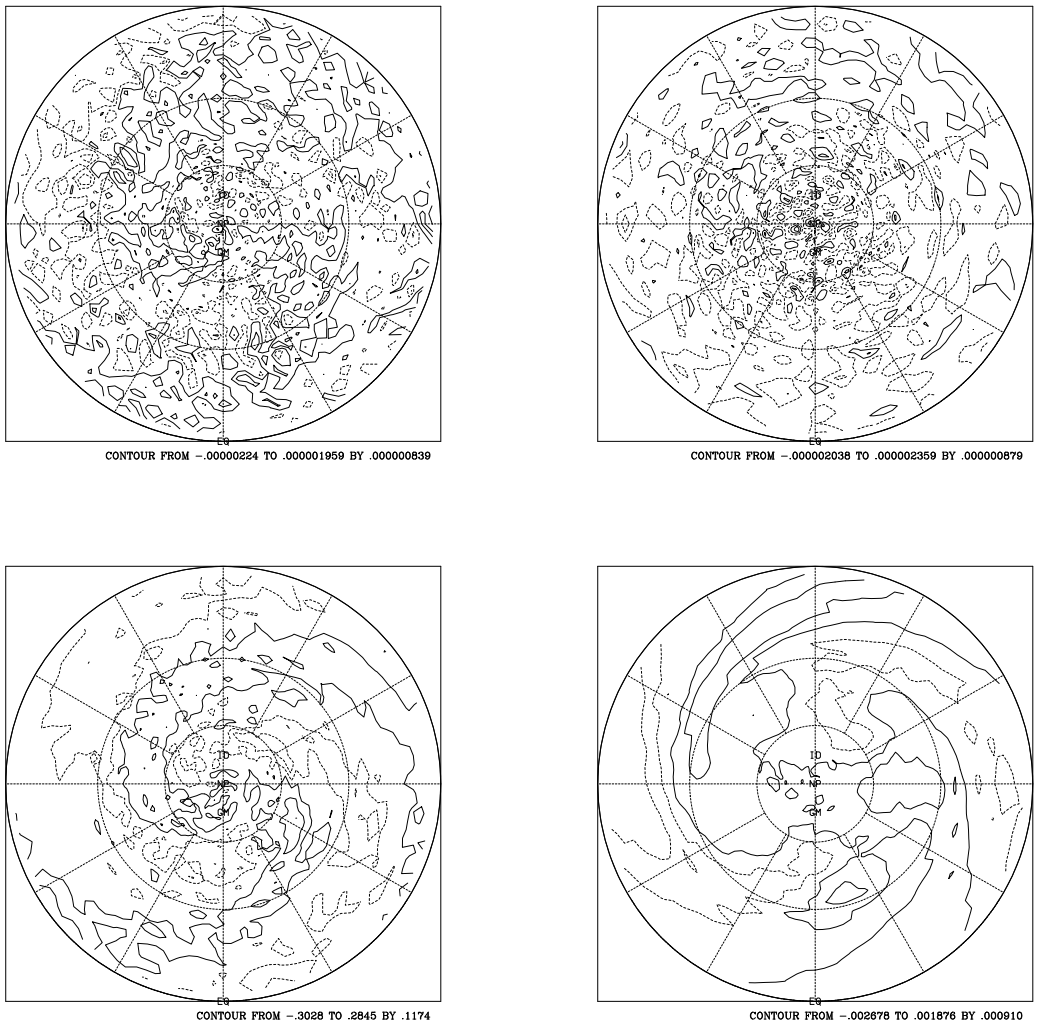

Figure 5.7: Same as Fig. 5.5 except the observations are available at the end of the assimilation window and the length of the assimilation is 1 hour.

Table 5.1: Rms errors of temperature at different model levels. Results in the second, third and fourth columns are the rms errors when no, weak and strong horizontal diffusions are included in the model, respectively.

Model levels	rms_1	rms_2	rms_3
1	0.25333	0.22312	0.23611
2	0.12962	0.11712	0.12267
3	0.85620×10^{-1}	0.86900×10^{-1}	0.64824×10^{-1}
4	0.81224×10^{-1}	0.83431×10^{-1}	0.65400×10^{-1}
5	0.81658×10^{-1}	0.88145×10^{-1}	0.67527×10^{-1}
6	0.91402×10^{-1}	0.98518×10^{-1}	0.82567×10^{-1}
7	0.12172	0.11732	0.12770
8	0.14343	0.13726	0.14319
9	0.15139	0.14624	0.15058
10	0.15543	0.15202	0.15642
11	0.18852	0.18093	0.19081
12	0.23645	0.22469	0.24192

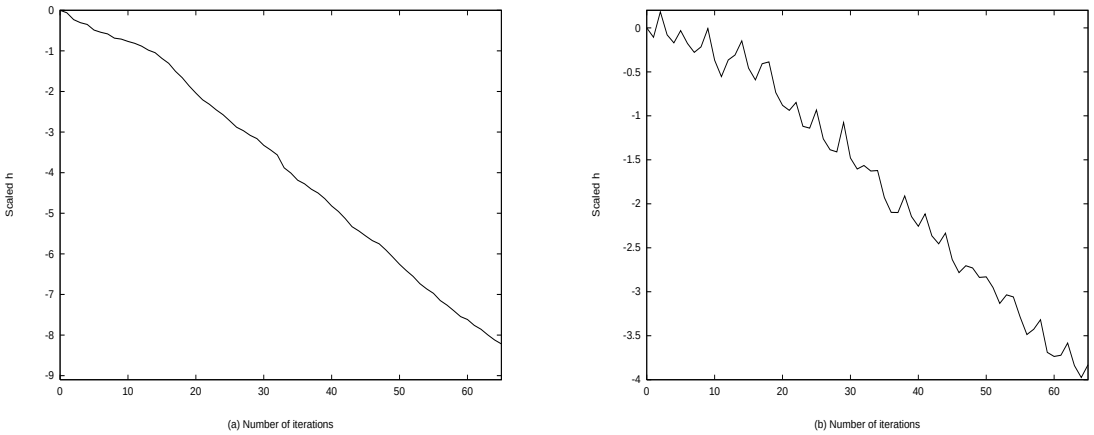

Figure 5.8: Same as Fig. 5.3 except the initial fields are randomly perturbed fields. The length of the assimilation window is 1 hour and the observations are available at every time step of the assimilation window.

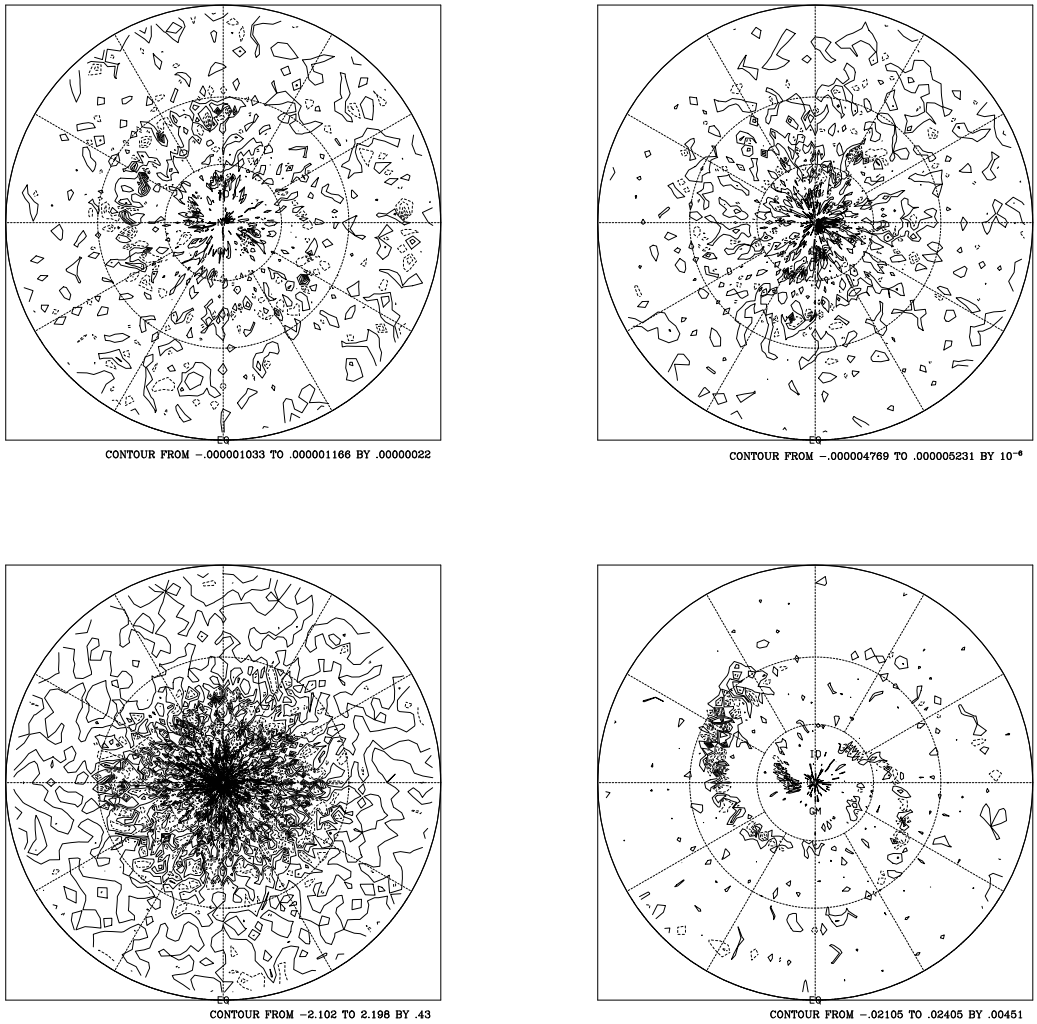

Figure 5.9: Divergence (top left), vorticity (top right), temperature (bottom left) natural log of surface pressure (bottom right) difference fields between randomly perturbed and reference fields at level 8. The randomly perturbed fields are the randomly perturbed fields of the analysis of the observational fields of divergence, vorticity, temperature and natural log of surface pressure at 0 UTC 1 June 1988.

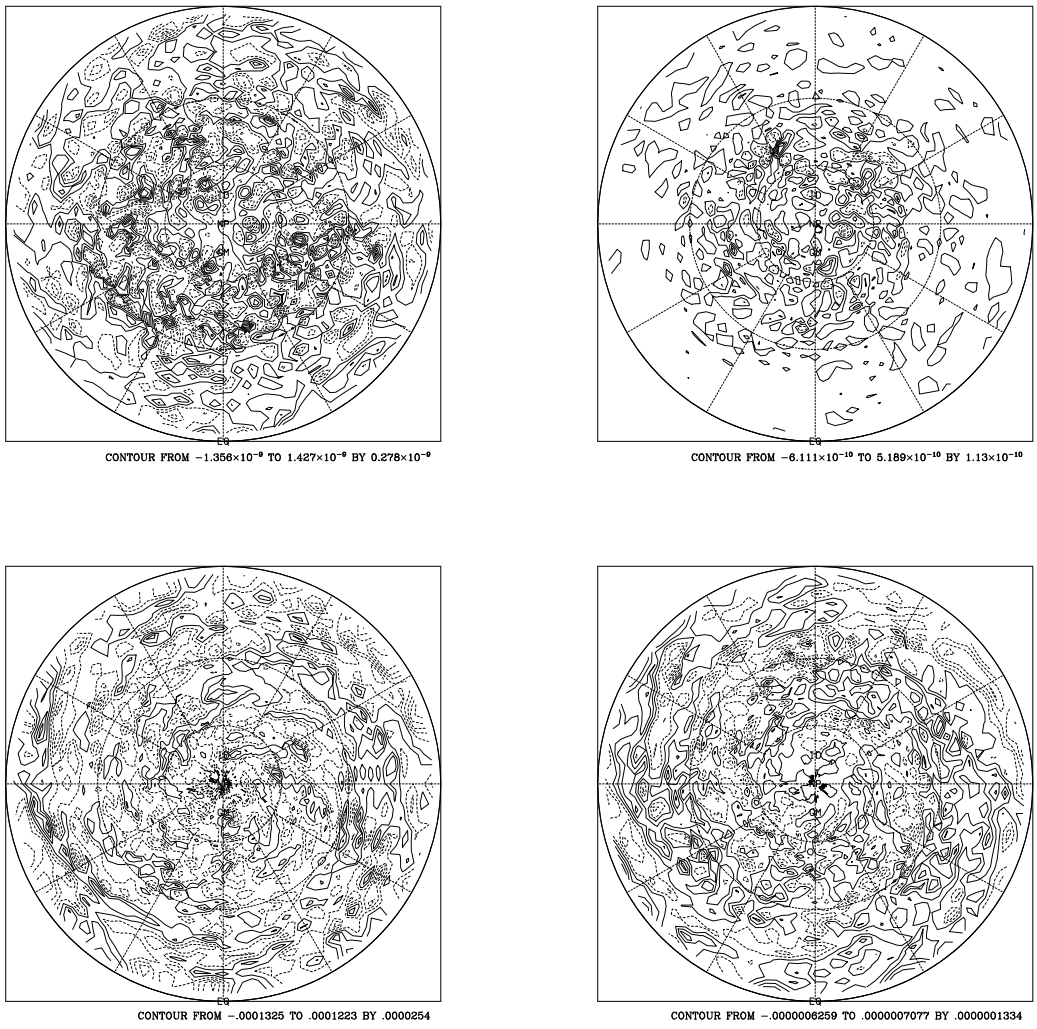

Figure 5.10: Same as Fig. 5.5 except the length of the assimilation window is one hour starting from the randomly perturbed initial conditions.

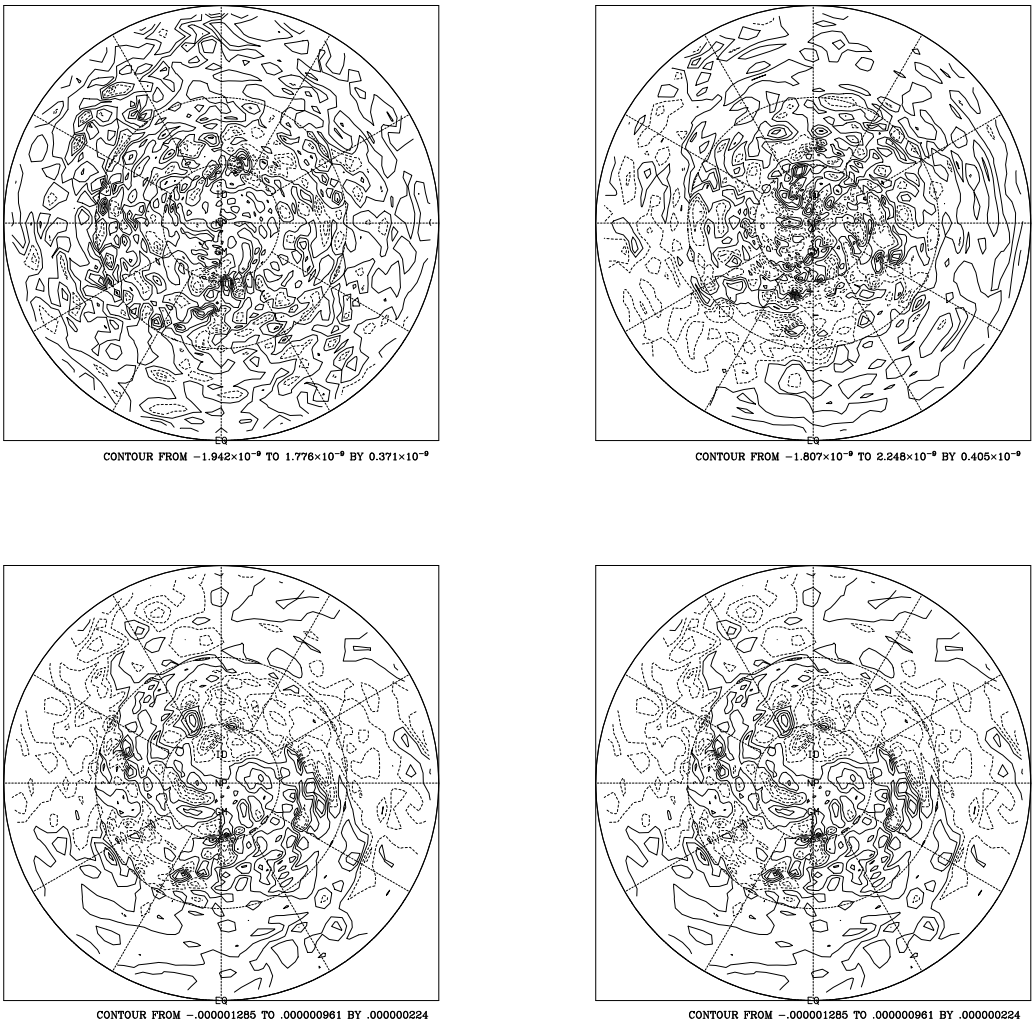

Figure 5.11: Same as Fig. 5.5 except the length of the assimilation window is one hour.

Figure 5.12: Same as Fig. 5.5 except the observations are available only at the beginning and end of assimilation window.

Figure 5.13: Same as Fig. 5.3 except the observations are available only at the beginning and end of assimilation window and the length of the assimilation window is six hours.

Figure 5.14: Same as Fig. 5.4 except The observations are available only at the beginning and end of the assimilation window.

CHAPTER 6

VARIATIONAL NUDGING DATA ASSIMILATION

6.1 Introduction

Among the 4-D data assimilation approaches, variational data assimilation (VDA) nudging data assimilation (NDA) [195], variational continuous assimilation (VCA) [52, 235], Kalman-Bucy and extended Kalman-Bucy filtering methods [29, 30, 31, 43, 163, 90, 91, 212] are considered to be five of the most promising techniques capable of utilizing the ever growing number of asynoptic observations. Briefly, the VDA method seeks to find an optimal initial condition and /or boundary conditions which minimizes in the least square sense the differences between the model solution and observations in a certain assimilation time interval [36, 44, 108, 151, 175, 218, 215]. It uses an optimal control approach based on adjoint model integration to obtain the gradient of the cost function, with respect to the control variables for the minimization procedure efficiently [177]. This approach is cheaper than that of using an explicit finite-difference approximation for the gradient when large-dimensional models are involved. Nevertheless, the cost of the VDA method for real distributed data is still prohibitive for operational applications. Additional research is required to improve the rate of convergence of the minimization part of the algorithm by proper scaling and weighting as well as developing efficient parallel algorithms for VDA. Other issues related to VDA concern determining the optimal length of the assimilation window, the treatment of on-off physical processes such as large-scale precipitation and deep cumulus convection [9, 231, 237, 216], and the control of high-frequency gravity-wave oscillations [226, 230], to mention but a few. Moreover, the adjoint VDA method may also be used to perform parameter estimation and sensitivity analysis.

The NDA method relaxes the model state towards the observations during the assimilation period by adding a non-physical diffusion-type term to the model equations. The nudging terms are defined as the difference between the observation and model solution multiplied by a nudging coefficient. The size of this coefficient is chosen by numerical experimentation so as to keep the nudging terms small in comparison with the dominant forcing terms in the governing equations, in order to avoid the rebounding effect that slows down the assimilation process, yet large enough to impact the simulation. NDA has been successfully used on global scale ([98, 122, 124]) and in a wide variety of research applications on mesoscale models ([47, 88, 172, 217] etc). The NDA method can be thought of as an iterative simplified approximation to the Kalman filter (KF) [119, 122]. The NDA method is a flexible assimilation technique which is computationally much more economical than the VDA method. However, results from NDA are quite sensitive to the *ad hoc* specification of the nudging relaxation coefficient, and it is not at all clear how to choose a nudging coefficient so as to obtain an optimal solution (the theory and examples of NDA are provided in section 6.2).

VCA [52, 235] considers the forecast model error by adding to the right hand sides of the model equations given by Eq. (1.4) a correction term $\lambda\phi$, i.e. a product of a prespecified time dependent variable λ and a spatially dependent variable ϕ,

$$\frac{\partial \vec{X}}{\partial t} = F(\vec{X}) + \lambda\phi. \tag{6.1}$$

The form of λ controls the distribution of the correction over the assimilation window. The objective of the VCA technique is to find the optimal ϕ which minimizes the cost function. Note that ϕ contains the same number of degrees of freedom as the initial state. The VCA is similar to the NDA. However, in the VCA, the correction varies in time only in a prespecified manner. The correction is calculated to optimally fit the

data throughout the assimilation window, rather than relaxing the solution toward the values at the observation time.

Variational nudging data assimilation (VNDA, which is called optimal nudging data assimilation in the work of Zou *et al.* [229]) combines the aforementioned data assimilation schemes in the most efficient way. The original idea of optimal nudging data assimilation was put forward by Le Dimet (personal communication). A parameter-estimation approach is used in the framework of the VDA algorithm to simultaneously determine the best initial conditions for numerical weather prediction (NWP) and optimal coefficients for the NDA scheme. The goal is to find the best initial conditions and optimal nudging coefficients which best assimilate the given observations. It is well known that the best nudging coefficients are those related to a KF in a linear system of equations. Applications of the KF technique to the assimilation of meteorological (or oceanographical) observations has been studied by several authors [31, 43, 163, 65, 66, 67, 68, 171], and present operational NDA procedures can be described as degraded forms of the KF. However, the KF is very costly to implement in practice [31, 56]. In Appendix E, the extended KF [163] and the relationship amongst the KF, VNDA and NDA approaches are described.

We can obtain optimal nudging coefficients in a much more economical way by using the VDA method in a parameter estimation mode; this allows the adjusting of variables other than the initial conditions for either linear or nonlinear systems and employs an adjoint model for gradient calculation. Parameters in the numerical weather prediction model can easily be incorporated in the adjoint VDA procedure and serve as additional control variables. The variational algorithm is formulated here using the nudging coefficients as the control parameters. The nudging coefficients are estimated so that the model solution is as close as possible to the observations.

In this chapter, we present the results of an application of the VNDA to the FSUGSM in its full-physics diabatic operational form. The gradient of the cost func-

tion with respect to the initial conditions is approximately obtained by integrating the adjoint model of the adiabatic version of the FSUGSM [219]. This introduces an error in the gradient calculation but also opens a door for data assimilation of the most sophisticated forecast models by employing imperfect adjoint models. The nudging term included in the model equations takes the model error into account. The nudging term also plays a role in improving the conditioning of the cost function and thus results in a faster convergence of the minimization process.

The FSUGSM and all its physical processes have been described in Chapter 4 and Appendix F, respectively. In subsections 6.2 and 6.2.3, some simple examples of the NDA are first presented and then the NDA and VNDA are described for a general model, respectively. Computational details and numerical results are shown in section 6.3. A comparison of estimated nudging, optimal nudging and variational data assimilation is provided in section 6.3.3. Conclusions as well as topics for further research related to the 4-D VDA with FSUGSM are presented in section 6.4.

6.2 Nudging data assimilation

6.2.1 Some simple examples

In order to get a feeling for the effects of nudging, we consider a few very simple cases in this subsection.

In NDA there is a preforecast integration period during which model variables are driven toward the observations by extra forcing terms (nudging terms) in the model equations. When the actual initial time is reached, the nudging terms are dropped from the model equations and the forecast proceeds. The principal objective is to bring the data and the model in harmony and provide a relatively noise-free start [81]. The schematic representation of the assimilation-forecast cycle of the NDA is provided in Fig. 6.2.

The equations for a particular gridpoint variable α may be expressed in the form

$$\frac{\partial \alpha}{\partial t} = F + N(\alpha, t)\epsilon[\alpha^{(0)} - \alpha], \tag{6.2}$$

where F represents the usual terms in the model equations; $N(\alpha, t)$ is a nonnegative nudging coefficient; and ϵ, a positive weight factor ≤ 1, is a confidence factor with respect to accuracy of the gridpoint value $\alpha^{(0)}(t)$ which is a best estimate based upon observations. If $\alpha^{(0)}(t)$ is assumed to be the true value, ϵ would be 1.

First assume that the physical forcing is zero, and further that N is constant and $\epsilon = 1$. Then (6.2) reduces to

$$\frac{\partial \alpha}{\partial t} = N[\alpha^{(0)} - \alpha], \tag{6.3}$$

or

$$\alpha = \alpha_0 e^{-Nt} + N e^{-Nt} \int_0^t N e^{Nt} \alpha^{(0)} dt, \tag{6.4}$$

where $t = 0$ represents the beginning of the preforecast period and α_0 is the value of α at $t = 0$. If $\alpha^{(0)}(t)$ is assumed to be constant in time toward which α is nudged, the solution of (6.4) is

$$\alpha = \alpha^{(0)} + (\alpha_0 - \alpha^{(0)}) e^{-Nt}. \tag{6.5}$$

Thus α approaches the observed value $\alpha^{(0)}$ as time passes, though of course never reaching it in a finite time T.

As a second case assuming the atmosphere and also the observations are changing linearly with time [i.e., $\alpha^{(0)}(t) = \alpha_0^{(0)} + at$]. Then the solution for this case is

$$\alpha = \alpha^{(0)}(t) + [\alpha_0 - \alpha_0^{(0)}] e^{-Nt} - \frac{a}{N}(1 - e^{-Nt}). \tag{6.6}$$

In this instance it is seen that the initial error, if such exists, is damped toward zero with increasing time (second term on right-hand side) but the third term approaches $-a/N$. Thus at best α approaches the observationally determined value $\alpha^{(0)}(t)$ only to within the constant a/N, which is the ratio of the rate of change of $\alpha^{(0)}$ to nudging constant. Also for a large N, the last term will be small.

The preceding simple cases have been mainly concerned only with a simple variable rather than a complete system of equations and the physical forcing has been omitted. In order to treat a more complete system analytically, shallow water equations with nudging terms added are linearized about a state of rest with constant Coriolis parameter f and no variation in the y-direction [88]. Assuming periodic disturbances of the form

$$\bar{X}' = \sum_m \hat{X}_m \exp(ik_m x),$$

where $\bar{X}'$ is a column vector of the perturbation quantities u', v' and h', $k_m = 2\pi/L_m$, is the mth wavenumber, L_m is the wavelength and the $\hat{X}_m$ are the corresponding amplitudes, the equations for the amplitudes take the form

$$\frac{\partial}{\partial t}\begin{pmatrix} \hat{u}_m \\ \hat{v}_m \\ \hat{h}_m \end{pmatrix} + \begin{pmatrix} N_u & -f & ik_m g \\ f & N_v & 0 \\ ik_m H & 0 & N_h \end{pmatrix}\begin{pmatrix} \hat{u}_m \\ \hat{v}_m \\ \hat{h}_m \end{pmatrix} = \begin{pmatrix} N_u \hat{u}_m^{(0)} \\ N_v \hat{v}_m^{(0)} \\ N_h \hat{h}_m^{(0)} \end{pmatrix} \tag{6.7}$$

or

$$\frac{\partial X}{\partial t} + AX = B. \tag{6.8}$$

Assuming the nudging coefficients and "observations" do not vary in time, the matrix A and vector B are constants, and the solution is given by a constant vector, $A^{-1}B$, and a time-varying part:

$$X = A^{-1}B + \sum_{j=1}^{3} C_j e^{-\sigma_j t}, \tag{6.9}$$

where σ_j are the eigenvalues of A and the C_j are constant vectors that depend on the initial conditions $X(0)$ and $A^{-1}B$. The C_j can be determined given $X(0)$ and $X^{(0)}$ by placing $t = 0$ in (6.9). In the case where the constant "observations" are in geostrophic balance, that is, $\hat{u}_m^{(0)} = 0$ and $\hat{v}_m^{(0)} = (igk_m/f)\hat{h}^{(0)}$, then $X^{(0)} = A^{-1}B$; hence

$$X - X^{(0)} = C_1 e^{-\sigma_1 t} + C_2 e^{-\sigma_2 t} + C_3 e^{-\sigma_3 t}, \tag{6.10}$$

Assuming further that the nudging coefficients are all equal to N, Eq. (6.10) leads to a stationary mode $\sigma_1 = N$, $\sigma_2 = N + i\nu$ and $\sigma_3 = N - i\nu$, where $\nu^2 = \omega^2 + f^2$. Here ν is the frequency of inertial gravity waves and $\omega^2 = k_m^2 gH$. Thus there is a stationary mode damped by nudging coefficient and two damped inertial gravity modes with a resulting approach toward the observational values $X^{(0)}$.

6.2.2 The basic theory of nudging data assimilation

An implicit assumption made in VDA is that the model exactly represents the state of the atmosphere. However, this assumption is clearly not true because any numerical model only approximately represents the state of the atmosphere.

The nudging data assimilation (NDA) technique introduced by Anthes [4] consists in achieving a compromise between the model and observations by considering the state of the atmosphere to be defined by

$$\frac{\partial \vec{X}}{\partial t} = F(\vec{X}) + G(\vec{X}^o - \vec{X}), \tag{6.11}$$

where G is a diagonal matrix with G_q, G_T, G_D, G_ζ and G_S as its diagonal submatrix entries representing adjustable nudging coefficients for the surface pressure, temperature, divergence, vorticity and moisture fields, respectively.

Together with the initial condition

$$\vec{X}(t_0) = U, \tag{6.12}$$

the system (6.11) has an unique solution $\vec{X}(U, G)$.

Let us consider a perturbation U' on the initial condition U in Eq. 6.12. The resulting perturbation on the variable $\vec{X}$, $\hat{X}$ is obtained from Eqs. (6.11) and (6.12) as

$$\frac{\partial \hat{X}}{\partial t} = \frac{\partial F}{\partial \vec{X}}\hat{X} - G\hat{X}, \tag{6.13}$$

$$\hat{X}(t_0) = U'. \tag{6.14}$$

The main difficulty in the NDA scheme resides in the estimation of the nudging coefficient G [194]. If G is too large, the fictitious diffusion term will completely dominate the time tendency and will have an effect similar to replacing the model data by the observations at each time step. Should a particular observation have a large error that prevents obtaining a dynamic balance, an exact fit to the observation is not required since it may lead to a false amplification of observational errors. On the other hand, if G is too small, the observation will have little effect on the solution. In general, G decreases with increasing observation error, increasing horizontal and vertical distance separation, and increasing time separation. In the experiment of Anthes [4] a nudging coefficient of 10^{-3} was used for all the fields for a hurricane model and was applied on all the domain of integration. In the experiment of Krishnamurti *et al.* [98] the relaxation coefficients for the estimated NDA experiment were kept invariant both in space and time, and their values were simply determined by numerical experience. The following values were used:

$$\begin{aligned} G_\zeta,\, G_q &= 10^{-4} s^{-1} \\ G_D &= 0.5 \times 10^{-4} s^{-1}, \end{aligned} \tag{6.15}$$

i.e. the vorticity, divergence and the pressure tendency fields are subjected to the Newtonian relaxation. The implicit dynamic constraints of the model then spread the updated information to other variable fields (temperature and moisture fields) resulting eventually in a set of balanced conditions at the end of the nudging period.

We employ a parameter-estimation approach [8] designed to obtain the best initial conditions and optimal nudging coefficients G^* [229]. They are optimal in the sense that the difference between the model solution and the observations will be small. The values in (6.16) will be used both as the nudging coefficient for estimated NDA experiment (usual NDA using *ad hoc* nudging parameters) and as the initial guess

in a variational parameter-estimation approach aimed at obtaining the best initial conditions and optimal nudging coefficients.

6.2.3 Variational nudging data assimilation

By fitting the model solutions to the observational data, the unknown parameters of the model can be obtained simultaneously by minimizing a cost function that measures the misfit between the model results and observations, in which the model parameters are the control variables along with the initial condition vector in our case [45, 104, 188]. For example, the barotropic gravity-wave speed in a two-dimension reduced-gravity, linear-transport model for the equatorial Pacific Ocean was used as a parameter control variable [193]. In the work of Panchang and O'Brien [159] the friction coefficient for a one-dimension tidal-flow model was the parameter optimally estimated from observations. In the work of Lardner *et al.* [104], the bottom drag coefficient and depth correction for a two-dimensional hydrodynamical model of the Arabian Gulf were the parameters optimally estimated.

The application of the variational approach to determine model parameters is conceptually similar to that of determining the initial conditions. In the following we will present a brief illustration of the method.

For the VNDA, the cost function J can be defined as

$$J(U,G) = \frac{1}{2}\sum_{i=0}^{M} < W(\vec{X}_i - \vec{X}_i^o), (\vec{X}_i - \vec{X}_i^o) > + < K(G - \hat{G}), G - \hat{G} >, \qquad (6.16)$$

where $\hat{G}$ denotes the estimated nudging coefficients and the K is specified weighting matrix. Here observations are assumed to be available everywhere on all model grid points for simplicity. For a more realistic case, a transform operator C from $\vec{X}$ to $\vec{X}^o$ should be included in (6.16) and hence G will no longer be a diagonal matrix with constant terms in each block. The second term plays a double rule. On one hand it ensures that the new values of the nudging parameters are not too far away from the

estimated quantity. On the other hand it enhances the convexity of the cost function since this term contributes a positive term, K, to the Hessian matrix of J [193].

The solutions of the VNDA procedure can be defined by the best initial condition U^* and optimal nudging coefficients G^* such that

$$J(U^*, G^*) \leq J(U, G), \qquad \forall G,\ U. \tag{6.17}$$

The problem of extracting the dynamical state from observations is now identified as the mathematical problem of finding initial conditions or external forcing parameters that minimize the cost function.

Owing to the dynamical coupling of the state variables to the forcing parameters, the dynamics can be enforced through the use of a Lagrange function constructed by appending the model equations to the cost function as constraints, thus avoiding the repeated application of the chain rule when differentiating the cost function. If the implicit time differencing equations of Eq (6.11) are given by the following equations

$$\vec{X}_1 = \vec{X}_0 + \Delta t F(\vec{X}_0) + \Delta t G(\vec{X}_1^o - \vec{X}_1), \tag{6.18}$$

$$\vec{X}_{i+1} = \vec{X}_{i-1} + 2\Delta t F(\vec{X}_i) + 2\Delta t G(\vec{X}_{i+1}^o - \vec{X}_{i+1}), \tag{6.19}$$

for $i = 1$, ..., $M-1$, then the Lagrange function may be defined by

$$\begin{aligned} L &= J(U) \\ &+ < P_0, \vec{X}_1 - \vec{X}_0 - \Delta t F(\vec{X}_0) - \Delta t G(\vec{X}_1^o - \vec{X}_1) > \\ &+ \sum_1^{M-1} < P_i, \vec{X}_{i+1} - \vec{X}_{i-1} - 2\Delta t F(\vec{X}_i) - 2\Delta t G(\vec{X}_{i+1}^o - \vec{X}_{i+1}), \end{aligned} \tag{6.20}$$

where P_i is a vector of Lagrange multiplies, which are identical to the adjoint model variables [206]. The Lagrange multiplies are not specified but computed in determining the best fit. At the minimum point the gradient of the Lagrange function must be zero. This leads to the following first-order condition:

$$\frac{\partial L}{\partial X_i} = 0, \tag{6.21}$$

$$\frac{\partial L}{\partial P_i} = 0, \tag{6.22}$$

$$\frac{\partial L}{\partial G} = 0, \tag{6.23}$$

The solution of Eqs (6.21-6.23) is called the stationary point of L. Eq (6.23) results in (the notation here has some problems)

$$\begin{aligned} & \nabla_G J - \Delta t P_0(\vec{X}_1^o - \vec{X}_1) \\ - & \sum_1^{M-1} 2\Delta t P_i(\vec{X}_{i+1}^o - \vec{X}_{i+1}), \end{aligned} \tag{6.24}$$

which yields the gradient of the cost function with respect to G

$$\begin{aligned} \nabla_G J &= \Delta t P_0(\vec{X}_1^o - \vec{X}_1) \\ &+ \sum_1^{M-1} 2\Delta t P_i(\vec{X}_{i+1}^o - \vec{X}_{i+1}), \end{aligned} \tag{6.25}$$

Eq (6.22) recovers discrete form of the original model Eq (6.11), while (6.21) yields the adjoint equations

$$P_{M-1} + 2\Delta t G^* P_{M-1} + W(\vec{X}_M - \vec{X}_M^o) = 0, \tag{6.26}$$

$$P_{M-2} + 2\Delta t G^* P_{M-2} - 2\Delta t\Big(\frac{\partial F}{\partial \vec{X}_{M-1}}\Big)^* P_{M-1} + W(\vec{X}_{M-1} - \vec{X}_{M-1}^o) = 0, \tag{6.27}$$

and

$$P_{i-1} + 2\Delta t G^* P_{i-1} - 2\Delta t\Big(\frac{\partial F}{\partial \vec{X}_i}\Big)^* P_i - P_{i+1} + W(\vec{X}_i - \vec{X}_i^o) = 0, \tag{6.28}$$

for $i = M-2$, $M-3$, ..., 1.

As we know from previous chapters, the value of the adjoint variable vector at the initial time is the gradient of the cost function with respect to the initial conditions

$$\nabla_U J = P(t_0). \tag{6.29}$$

Therefore the gradient of the cost function with respect to both the initial conditions and the nudging coefficients is

$$\nabla J = (\nabla_U J, \nabla_G J)^t. \tag{6.30}$$

Having obtained the value of the cost function J by integrating the model (6.11) forward, and the value of the gradient by integrating the adjoint Eqs (6.26, 6.27, 6.28) backwards in time, any efficient large-scale unconstrained minimization method may be employed to minimize the cost function with respect to the vector of control variables and finally obtain the best initial conditions and an optimal parameter estimation.

It is worth noting that if G is zero the Eqs (6.26, 6.27, 6.28) reduce to the usual first order adjoint equations and the VNDA reduces to the VDA. In the case in which the cost function is defined as an integral, the derivation of the gradient of the cost function with respect to the nudging coefficients is provided in Appendix D.

6.2.4 Identifiability and stability of parameter estimation

If the parameter estimation is going to be successful, an important question must be addressed; under what conditions can one expect the proposed estimation method to yield unique and stable results [22, 23, 24, 104, 153, 159, 193]? Now we discuss these issues.

Strictly speaking, if the method is going to work, the problem should be well-posed. The inverse problem is often ill-posed, but there are situations under which a meaningful solution can be found, although in a limited sense. It will be therefore important to be able to recognize the circumstances which allow a solution of the problem to be found. The inverse problem can be defined as follows: let a functional relationship $\vec{X} = f(G)$ be given between a parameter vector G which in our case is the nudging coefficients and $\vec{X}$ where $\vec{X}$ represents the state variable of our problem. The inverse problem will then be to determine the nudging coefficients G on the basis of $\vec{X}$ and the inverse relationship $G = R(\vec{X})$. This problem is said to be well-posed if and only if (1) to every $\vec{X}$ there corresponds a solution G; i.e. a solution exists; (2) the solution is unique for any given $\vec{X}$; and (3) the solution depends continuously on

$\vec{X}$; i.e. the solution is stable. If the inverse problem fails to satisfy one or more of these three requirements, it is said to be ill-posed.

Uniqueness and parameter identifiability [93] can be defined in the following way. Let $\vec{X}_1 = f(G_1)$ and $\vec{X}_2 = f(G_2)$ be two solutions of the forward problem, and $G_1 = R(\vec{X}_1)$ and $G_2 = R(\vec{X}_2)$ be two solutions of the inverse problem, then

$$\|\vec{X}_1 - \vec{X}_2\| = 0 \Leftrightarrow \|G_1 - G_2\| = 0, \tag{6.31}$$

where $\|\cdot\|$ represents a norm over the appropriate space. In practical problems $\vec{X}$ is only given at discrete points in space and time, and R represents a minimization of a functional J as given in previous subsection. While uniqueness refers to the inverse problem, R, identifiability refers to the forward problem, f. If two sets of parameters lead to the same function $\vec{X}$, the parameters are said to be unidentifiable. Uniqueness on the other hand is concerned with the problem whether different parameters may be found from a given $\vec{X}$, if so the parameters are nonunique.

Stability can be defined in the following way. For every $\epsilon > 0$ there exists a δ such that for $G_1 = R(\vec{X}_1)$ and $G_2 = R(\vec{X}_2)$ one has

$$\|\vec{X}_1 - \vec{X}_2\| < \delta \Rightarrow \|G_1 - G_2\| < \epsilon. \tag{6.32}$$

Eq. (6.32) states that small errors in the variables must not lead to large changes in the computed parameter. Tarantola [203] discusses the concepts of uniqueness and stability in more detail.

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

the parameter values are bouncing back and forth between high and low values [193]. Lardner *et al.* [104] pointed out that in most cases it was necessary to include a term in the cost function that penalizes large variations in the estimated parameters.

The uniqueness of the solution implies that the cost function J is convex. The requirement for this condition is that the Hessian matrix is positive definite [73]. Writing the cost function as

$$J = J_U + J_G, \tag{6.33}$$

where J_U represents the first term in Eq. (6.16), while J_G represents the last term or the term representing prior information about the parameters (nudging coefficients). The Hessian is represented by

$$\frac{\partial^2 J}{\partial G^2} = \frac{\partial^2 J_U}{\partial G^2} + \frac{\partial^2 J_G}{\partial G^2}. \tag{6.34}$$

The first term gives

$$\frac{\partial^2 J_U}{\partial G^2} = \sum_{i=0}^{M} < W \frac{\partial \vec{X}_i}{\partial G}, \frac{\partial \vec{X}_i}{\partial G} > + < W(\vec{X}_i - \vec{X}_i^o), \frac{\partial^2 \vec{X}_i}{\partial G^2} >, \tag{6.35}$$

which can be positive or negative. So the first term in the minimization criterion does not guarantee that the cost function is convex. The last term in Eq. (6.34) gives

$$\frac{\partial^2 J_G}{\partial G^2} = K, \tag{6.36}$$

which is clearly a positive term. Adding prior information about the parameters therefore increases the chance that the cost function will be convex. Of course there is no guarantee that the term in Eq. (6.36) will make the Hessian positive definite. Carrera and Neuman [23] discuss the effect of prior information in a few simple examples of estimation of aquifer parameters. Their examples clearly show that the addition of prior information leads to unique solutions.

6.3 Numerical results of the variational data assimilation

The algorithmic procedure of the VNDA assumes the following form:

1. Set the iteration number $k = 0$ and initialize the state variable $\vec{X}$ and the nudging coefficient vector G as U_0 and G_0, respectively, where G_0 is given by (6.16).

2. Test (U_k, G_k) for convergence. If the following convergence criterion is satisfied

$$\|G_k\| \leq \epsilon \|U_k\|, \tag{6.37}$$

   where ϵ is a predetermined small number, then stop. Otherwise continue.

3. Integrate the model (6.11) forward in time and calculate the value of the cost function defined by

$$\begin{aligned} J(U,G) &= \frac{1}{2}\{W < \vec{X}_k(t_f) - \vec{X}^o(t_f), \vec{X}_k(t_f) - \vec{X}^o(t_f) > \\ &+ W < \vec{X}_k(t_0) - \vec{X}^o(t_0), \vec{X}_k(t_0) - \vec{X}^o(t_0) > \}, \end{aligned} \tag{6.38}$$

   if the observations are available only at the beginning and end of the assimilation window , or defined by Eq (6.16) if the observations are available at every time step.

4. Integrate the adjoint Eqs (6.26, 6.27, 6.28) backwards in time and calculate the gradient of the cost function with respect to the control variables U and G using (6.30).

5. Apply the limited-memory quasi-Newton (LBFGS) large-scale unconstrained minimization algorithm of Liu and Nocedal [117] to obtain a set of new values for U and G: U_{k+1} and G_{k+1}. Set $k = k + 1$ and go to step 2.

In our experiments, no prior information was presumed ($K = 0$).

For the nudging terms added to the model, an implicit time-differencing scheme was used in order to ensure computational stability for any value assumed by the nudging coefficients. The time integrations are carried out in two steps. The tendency without the nudging term, $\vec{X}^*(t + \Delta t)$, is first calculated. The Newtonian term is then expressed in finite difference form using the relation

$$\frac{\vec{X}(t + \Delta t) - \vec{X}^*(t + \Delta t)}{2\Delta t} = G[\vec{X}^O(t + \Delta t) - \vec{X}(t + \Delta t)], \tag{6.39}$$

or

$$\vec{X}(t + \Delta t) = \frac{\vec{X}^*(t + \Delta t) + 2\Delta t G \vec{X}^O(t + \Delta t)}{1 + 2\Delta t G}. \tag{6.40}$$

It is first necessary to obtain the correct gradient of the cost function with respect to the nudging coefficients in the parameter-estimation procedure before carrying out the minimization of the cost function. Since otherwise all our calculations will be erroneous. The Taylor formula applied in the direction $G' = \nabla_G J(G)$ yields

$$\psi(\alpha) = \frac{J(G + \alpha G') - J(G)}{\alpha G'^t \cdot \nabla_G J} = 1. + O(\alpha \|G'\|). \tag{6.41}$$

For small α but not too close to the machine accuracy, the value of $\psi(\alpha)$ should be close to one if the gradient $\nabla_G J$ is correctly calculated. Fig. 6.1 indicates that for α between 10^{-11} and 10^{-22} an unit value of $\phi(\alpha)$ is obtained and therefore the correctness of the gradient calculation has been verified.

6.3.1 Variational nudging data assimilation with the adiabatic version of the FSU global spectral model

Before conducting the VNDA with the FSUGSM in its full-physics operational form, it is important to ensure that the VNDA performs successfully with the adiabatic version of the FSUGSM since the adjoint model is derived from the adiabatic

version of the FSUGSM. As stated previously, the gradient calculation is exact except for computational errors.

The experiments are devised as follows: the model-generated values starting from an analysis of the real observational data at 6 UTC 1 June 1988 are used as observations and the initial guess for initial conditions is taken as an analysis of the real observational data at 0 UTC 1 June 1988. The initial guess for nudging coefficients is given by (6.16).

The experiments carried out here involved nudging the surface pressure, divergence and vorticity. Figs. 6.6 (dash line) and 6.4 display the variations of the log of the scaled cost function (J_k/J_0) and scaled gradient norm ($\|\vec{g}_k\|/\|\vec{g}_0\|$) with the number of iterations in the minimization process when the observations are available at every time step and only at the beginning and end of the assimilation window, respectively. After either 24 or 18 iterations, the normalized cost function and the normalized norm of the gradient are reduced by more than 2.5 and 2 orders of magnitude, respectively, irrespective which set of observations is used. At this stage the prescribed convergence criterion given by (6.37) with $\epsilon = 10^{-5}$ is satisfied. In the case where the observations are available at every time step of the assimilation window, by continuing the minimization process, the scaled cost function and scaled gradient norm decreased another 2 and 1 orders of magnitude after 65 iterations, respectively. Both the cost function and the gradient norm continue to decrease even after 65 iterations. We note that due to the nudging effect, it appears that when more observations are added, the convergence rate becomes faster. The opposite is true for the VDA experiments carried out in Chapter 5. Comparing Figs. 6.6 and 6.4 with Figs. 5.3 and 5.13, it is observed that the VNDA is indeed much faster than the VDA for our test problem. In both of VNDA experiments the minimization requires less than 25 iteration to satisfy the prescribed convergent criterion while the minimizations requires more than 65 iterations to converge in both of the VDA experiments. This

is due to the fact that the conditioning of the Hessian is improved when a nudging term is added to the forecast model. These figures indicate that VNDA yields higher quality solutions requiring much less iterations. If we notice that the only additional cost for VNDA is three more control variables (nudging coefficients) and the calculation of the gradient of the cost function with respect to these variables, the VNDA is really cost effective in obtaining a speed up of the convergence of the minimization process.

It is a good idea to consider the difference fields between the retrieved fields and the observation fields before and after the minimization process, since these difference fields measure the ability of the VNDA to retrieve the observational fields. Figs. 6.8 and 6.5 show the difference fields between the retrieved fields of divergence, vorticity, temperature and natural log of the surface pressure and the corresponding reference (observational) fields at the beginning of the assimilation window after the minimization process for the two experiments. Comparing with the same difference fields (Fig. 5.6) before the minimization process, it is observed that after minimization the maximum difference values are reduced by one order of magnitude in all difference fields of divergence, vorticity, temperature and natural log of the surface pressure in both cases.

Figs. 6.3 and 6.7 display the evolution of the nudging coefficients, $G_{\ln p}$, G_ζ and G_D, in the VNDA procedure. During the first 15 iterations, all the nudging coefficients experience a dramatic increase for both cases. After 15 iterations the increases level-off. The optimal values of the nudging coefficients are provided in Table 6.1.

It is observed that the optimal nudging coefficients are two orders of magnitude bigger than the initial guesses given by (6.16), which were suggested by Krishnamurti *et al.* [98]. By noticing that the horizontal truncation of our experiment (T42) is smaller than theirs (T106), the values obtained appear to be very reasonable since it

is general experience that the nudging coefficients increase with decreasing horizontal resolution.

We note here that the optimal nudging coefficients do not tend to infinity as the minimization proceeds. This is due to the fact that when the minimization process proceeds, both the initial state U and the nudging coefficients G are updated and the optimal solution of the minimization problem is obtained due to the combined effect of the model dynamics and the nudging term.

We also carried out an experiment starting from the randomly perturbed initial conditions described in section 5.4.2 with the observations available at every time step. The results are similar to the results described above and are displayed in Figs. 6.9, 6.10 and 6.11.

All previous experiments were carried out starting from an analysis of the observational fields of divergence, vorticity, temperature and natural log of surface pressure at 0 UTC 1 June 1988. These analysis fields are not initialized. One may assume the existence of spurious gravity waves (high frequency "noise") in them. The nonlinear normal mode initialization (NNMI) permits an explicit filtering of the unwanted high frequency oscillations (implied by the initial data) consistent with the framework of the particular dynamical prediction model being used. Therefore, in order to investigate the impact of gravity oscillations on the VNDA, we apply the NNMI to the above analysis fields and perform the same experiments starting from the resulting fields of the NNMI. The results are similar to the results described above. Fig. 6.6 (solid lines) shows the results of VNDA when the observations are available at every time step of the assimilation window. It is clear that there are no sizable differences whether the VNDA starts from either the analysis of the observational fields or from the fields initialized by the NNMI, which means there are almost no spurious gravity waves in the analysis of the observational fields.

6.3.2 Variational nudging data assimilation with the FSU global spectral model in its full-physics operational form

In this subsection, we will apply the VNDA to the FSUGSM in its full-physics operational form while using the same adjoint model of the adiabatic version of the FSUGSM to calculate the gradient of the cost function. An implicit assumption made here is that the gradient of the cost function thus calculated is a reasonable approximation to the exact gradient such that the VNDA could proceed successfully.

The same experiments as described in subsection 6.3.1 are performed when observations are available at every time step and only at the beginning and end of the assimilation window, respectively. In Fig. 6.12 the cost function and the gradient norm are shown as functions of the number of iterations for all the experiments. The values of the cost function and gradient norm are normalized by their initial values, respectively. After 9 iterations the cost function and gradient norm are reduced by more than 1.8 and 1.7 orders of magnitude, respectively, in the case where observations are available at every time step of the assimilation window. Based on the behaviour of the gradient norm and the cost function, the minimization is satisfactory, i.e. the cost function and gradient norm do not decrease significantly after 8 iterations.

In all aforementioned assimilation experiments we have also calculated the cost functions corresponding to each variable (divergence, vorticity, temperature and pressure). It is noticed that the decreases of the different components of the cost function are uniform and similar to the decrease in the total cost function.

The nudging coefficients as a function of the number of iterations from the present experiment are shown in Fig. 6.13. The optimal nudging coefficients obtained in this case are $G_{\ln p} = 0.9553 \times 10^{-3}$, $G_{\zeta} = 0.6081 \times 10^{-2}$ and $G_D = 0.3159 \times 10^{-2}$. They are very similar to the corresponding optimal nudging coefficients obtained in the previous subsection. The behavior of the nudging coefficients as a function of the

number of iterations is also very similar in all the experiments performed in this thesis.

It is important to note that, even though an approximate gradient is used in the minimization, the VNDA works successfully. This may lead to a possible operational implementation of the 4-D VDA for sophisticated operational models while using a simplified adjoint model to calculate the gradient. If full adjoint models of the most complex models in their full-physics operational form were used, the convergence rate would be improved and the minimum would be different from that obtained using simplified adjoint models. But it is reasonable to assume that they will not be far from each other. All these results indicate that the assumption mentioned at the beginning of this subsection seems to be very reasonable. Courtier *et al.* [37] and Zupanski [235] also employed simplified adjoint models to obtain the gradient of the cost function with respect to control variables while integrating the nonlinear model in its full-physics operational form. Their results agree with ours.

The moisture field is not part of the control variables vector in the present experiment. Only divergence, vorticity, temperature and natural log of surface pressure fields are updated explicitly during each iteration of the minimization process. The moisture field is implicitly updated through the forecast model integrations, i.e. the impact of the improved fields of divergence, vorticity, temperature and natural log of surface pressure progressively propagates to the moisture field through the forward model integrations.

6.3.3 Comparisons of estimated nudging, optimal nudging and variational data assimilation

In this section we compare the NDA and VNDA techniques when observations are available at every time step of the assimilation window. The FSUGSM is used in its full-physics operational form. The basic experiments that follow consist of four types of assimilation:

1. A control assimilation without nudging,

2. An assimilation with estimated nudging coefficients as suggested by Krishnamurti *et al.* [98],

3. An assimilation with optimal nudging coefficients obtained in subsection 6.3.2 by an optimal parameter-estimation procedure using the adjoint technique,

4. VNDA minimizing a cost function measuring the distance between the model solution and observations. An overview of the aforementioned assimilations are displayed in Fig. 6.2.

With the estimated and optimal nudging coefficients we may carry out two parallel NDAs. For the sake of comparison, and with a view to obtain a better insight into the ability of the VNDA procedure, we also carried out a similar VNDA experiment. The ensuing 12 hour integrations from the retrieval are used for carrying out a comparison with the corresponding results of NDA schemes.

We do not intend to compare the NDA with the VNDA approaches directly since the two models have their own arbitrariness in the definition of the cost function, the choice of the minimization algorithm, the determination of the stopping criteria, the length of assimilation window and the choice of nudging coefficients. However, the ability to reconstruct as accurately and economically as possible the state of the flow is of paramount importance.

Rms errors are computed for the aforementioned assimilation experiments. Distributions of the rms differences for temperature fields between true solution and the forecast at the end of the forecast period are displayed in Figs. 6.14, respectively. As expected [229], the NDA with estimated nudging coefficients yields the poorest results while the VNDA yields the best results.

6.4 Conclusions

In this chapter we took advantage of the VDA's ability to perform an optimal parameter estimation to obtain both optimal nudging coefficients and the best initial conditions simultaneously. That is the nudging coefficients are part of the control variables vector. Both the initial conditions and the nudging coefficients are updated after each iteration of the minimization process. Since at every iteration the model integration is carried out with current estimated nudging coefficients used in the nudging term, the conditioning of the Hessian is improved. Thus variational nudging assimilation results in faster convergence rate compared with the 4-D VDA. The resulting optimal nudging coefficients can be applied in NDA, thus the term optimal NDA. The optimal NDA is practically implementable and performs very well in as far as the convergence and quality of the resulting assimilation state are concerned. This procedure is much more economical than the VDA approach due to tremendous computational cost of the VDA. Therefore the optimal NDA is also a good candidate as a future data assimilation scheme where the observations in a certain period of time can be effectively incorporated into the model so as to obtain the best initial conditions. The relationship amongst the KF, VNDA and NDA approaches is described in Appendix E, can also be found in reference [229].

In subsection 6.3.2, we carried out VNDA experiments using the FSUGSM in its full-physics operational form and using a simplified adjoint model of the adiabatic

version of the FSUGSM to calculate the gradient approximately. The results thus obtained are realistic and both the cost function and the gradient norm in all experiments show sufficient decreases. Although the moisture field is not explicitly updated, the resulting moisture field from VNDA is satisfactory. The main conclusion of this chapter is that the VNDA could be performed successfully using most sophisticated models while using a simplified adjoint model to calculate the gradient.

It might be desirable for nudging coefficients G varying both in time as well as over the horizontal domain and on the different vertical levels. Zou *et al.* [229] demonstrated the ability of the VNDA to obtain optimal nudging coefficients varying in space in the framework of a parameter-estimation approach. In their experiment, the spatial variability of the optimal nudging coefficients is very small. Therefore, using constant optimal nudging coefficients for each variable is satisfactory for perfect observations' assimilation. When the nudging coefficients have too many degrees of freedom, physical constraints should be added to the choice of the nudging coefficients. We will pay special attention to this topics in the future.

Table 6.1: Optimal nudging coefficients obtained by variational nudging data assimilation when the observations are available at every time step (case 1), and when the observations are available only at beginning and end (case 2) of the assimilation window, respectively.

Cases	G_ζ	G_D	$G_{\ln p}$
1	0.5869×10^{-2}	0.3915×10^{-2}	0.1730×10^{-2}
2	0.5437×10^{-2}	0.3921×10^{-2}	0.2467×10^{-2}

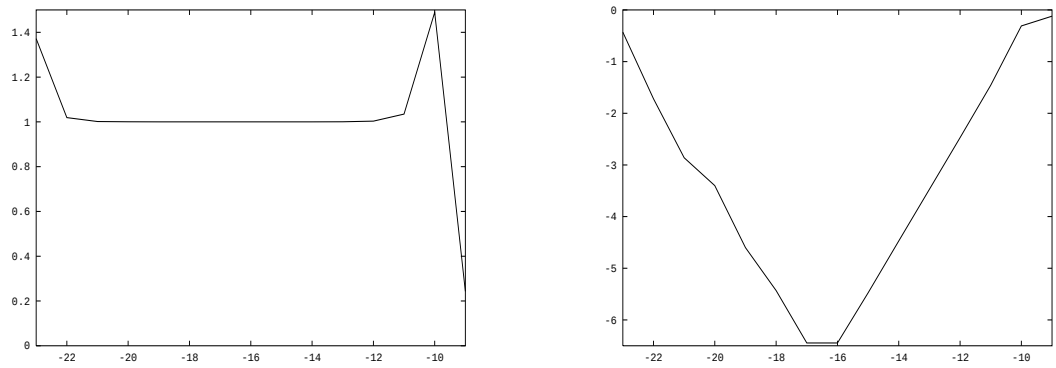

Figure 6.1: Verification of the correctness of the gradient calculation using Taylor expansion: (a) variation of ϕ with $\log \alpha$, (b) variation of $\log |\phi - 1|$ with $\log \alpha$.

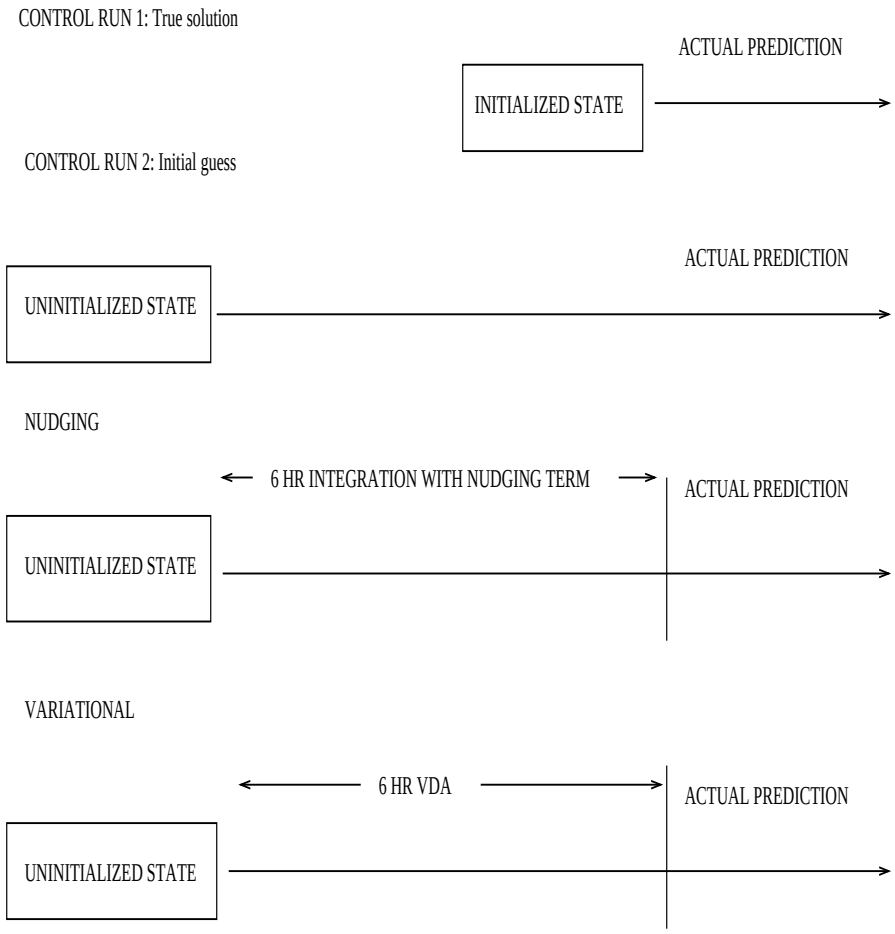

Figure 6.2: Schematic representation of the assimilation-forecast cycle.

Figure 6.3: Variation of the nudging coefficients, $G_{\ln p}$ (dash line), G_ζ (solid line) and G_D (long dash line), with the number of iterations. Both the initial conditions and the nudging coefficients $G_{\ln p}$, G_ζ and G_D serve as control variables. The observations are available only at the beginning and end of the assimilation window.

Figure 6.4: Variation of the *log* of the scaled cost function (J_k/J_0) and the scaled gradient norm ($\|\vec{g}_k\|/\|\vec{g}_0\|$) with the number of iterations using LBFGS algorithm. The observations are available only at the beginning and the end of the assimilation window.

Figure 6.5: Divergence (top left), vorticity (top right), temperature (bottom left) and natural log of surface pressure (bottom right) difference fields between the retrieved fields and the reference fields at level 8. The starting fields are the analysis fields of divergence, vorticity, temperature and natural log of surface pressure at 0 UTC 1 June 1988. Both the initial conditions and the nudging coefficients $G_{\ln p}$, G_ζ and G_D serve as control variables. The observations are available only at the beginning and end of the assimilation window.

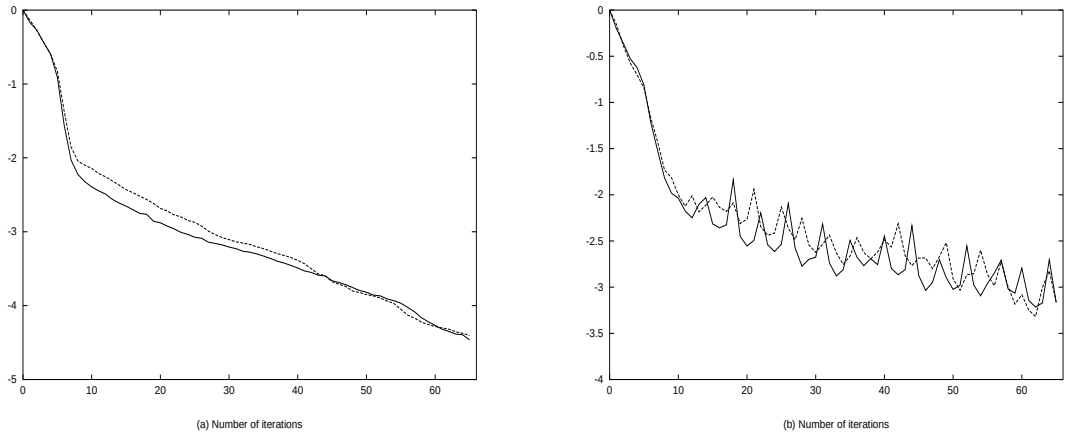

Figure 6.6: Same as Fig. 6.4 except that the observations are available at every time step of the assimilation window. Solid and dash lines indicate the results obtained starting from the analysis of the observational fields and the resulting fields after applying the NNMI to it, respectively.

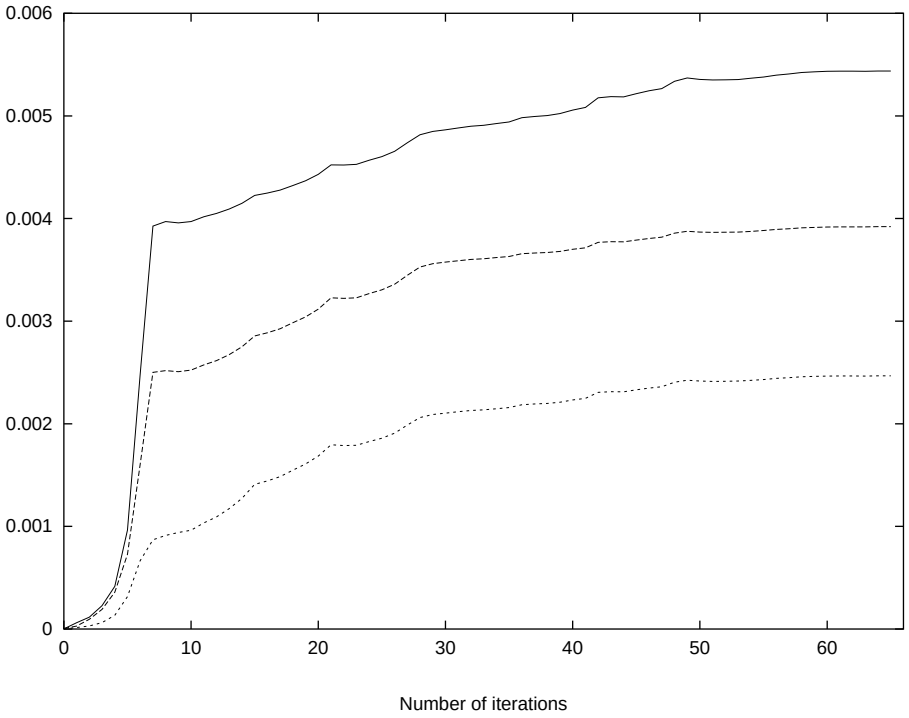

Figure 6.7: Same as Fig. 6.3 except that the observations are available at every time step of the assimilation window.

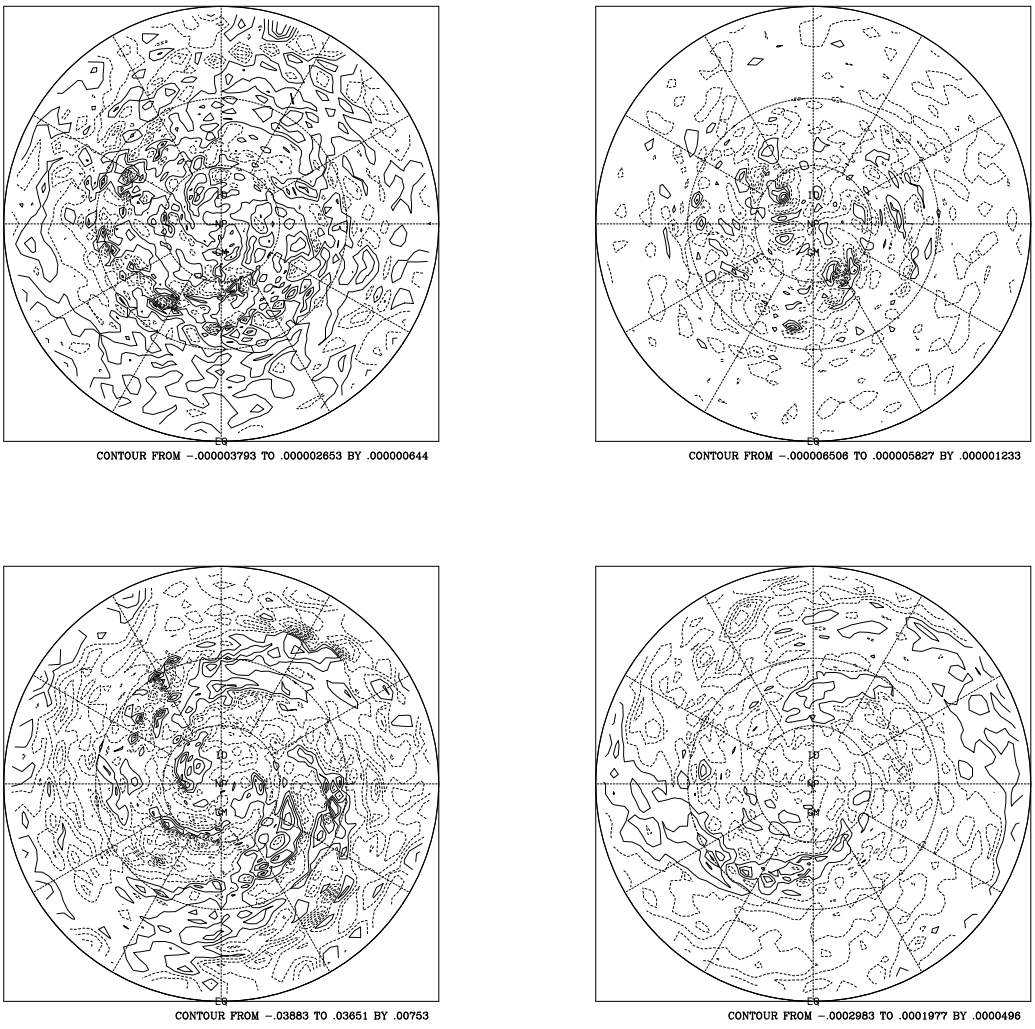

Figure 6.8: Same as Fig. 6.5 except that the observations are available at every time step of the assimilation window.

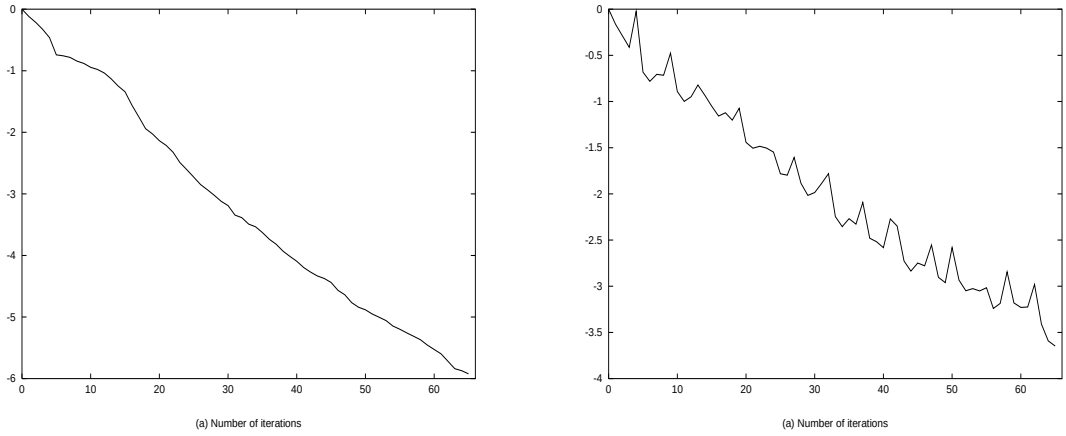

Figure 6.9: Same as Fig. 6.4 except that the initial guess is the randomly perturbed initial conditions, and the observations are available at every time step of the assimilation window.

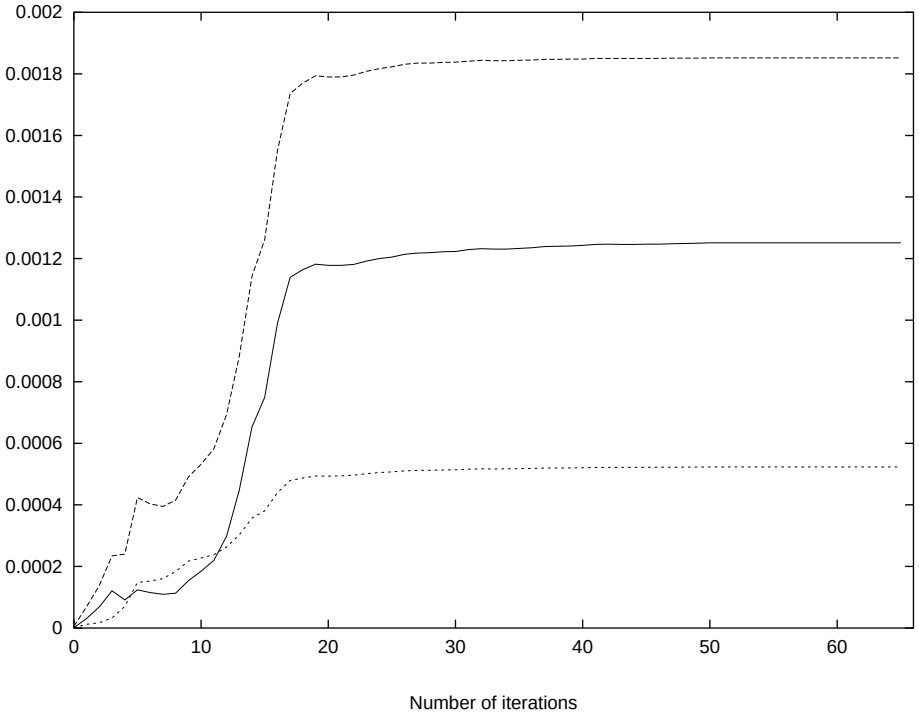

Figure 6.10: Same as Fig. 6.3 except that the initial guess is the randomly perturbed initial conditions, and the observations are available at every time step of the assimilation window.

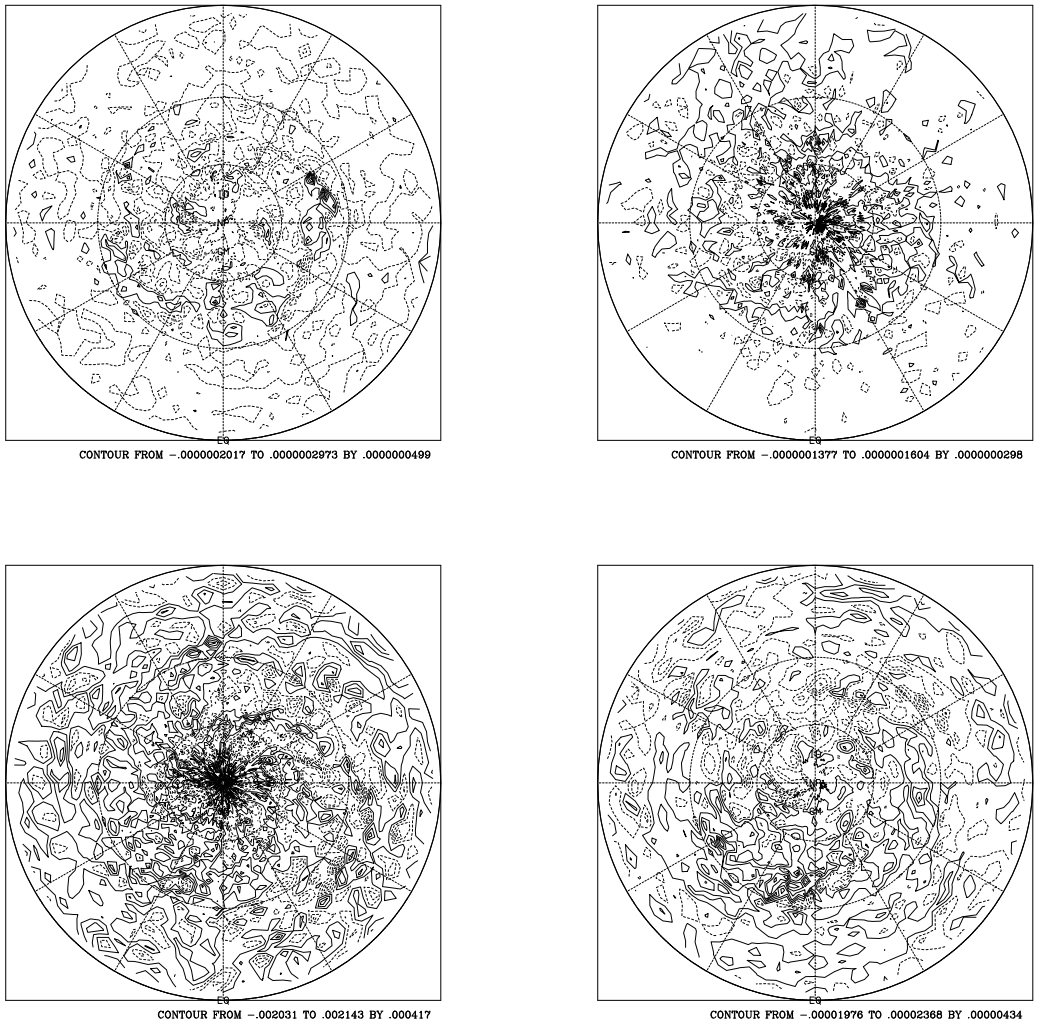

Figure 6.11: Same as Fig. 6.5 except that the initial guess is the randomly perturbed initial conditions, and the observations are available at every time step of the assimilation window.

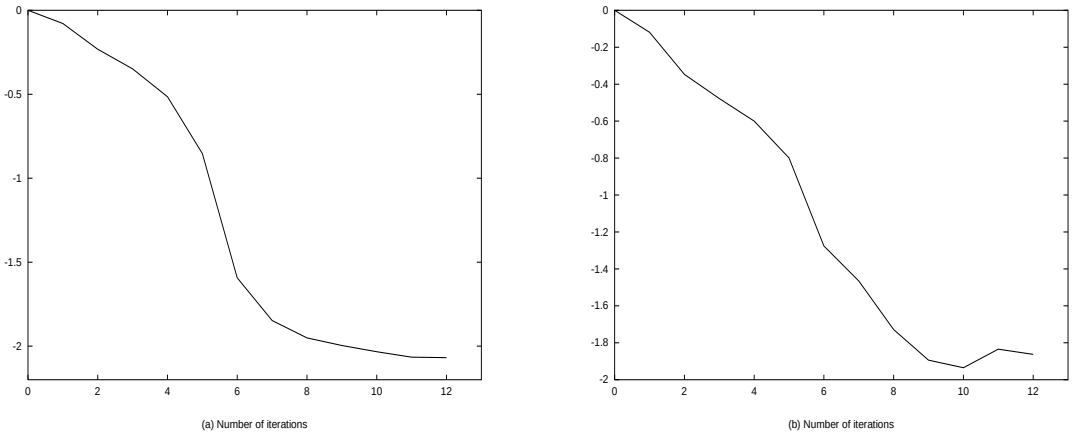

Figure 6.12: Same as Fig. 6.4 except that the Florida State University global spectral model in its full-physics form is used, and the observations are available at every time step of the assimilation window.

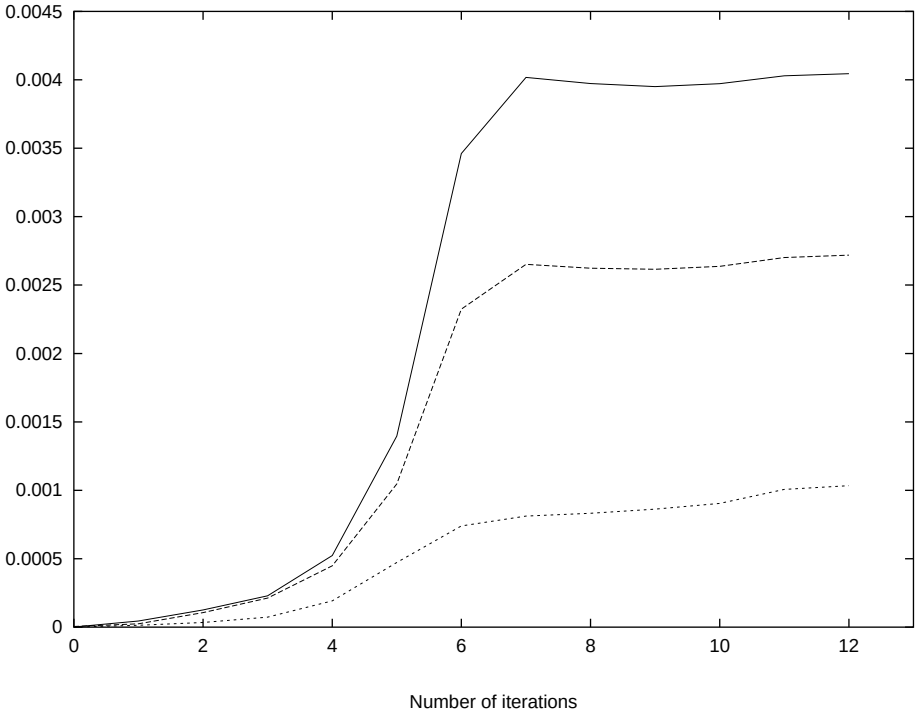

Figure 6.13: Same as Fig. 6.3 except that the Florida State University global spectral model in its full-physics form is used, and the observations are available at every time step of the assimilation window.

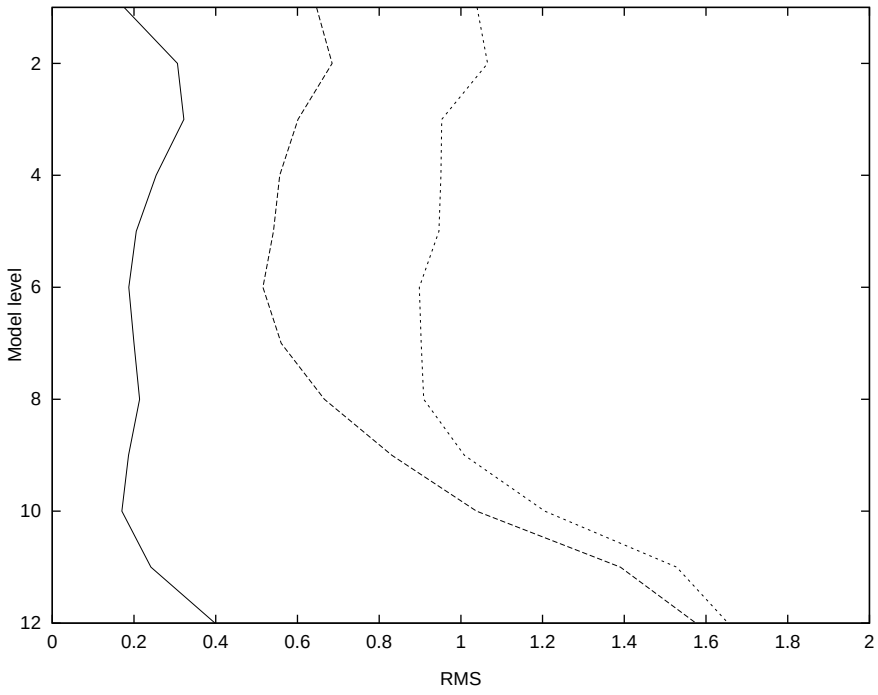

Figure 6.14: Solid, long dash and dash lines denote the rms errors between the forecast and exact temperatures at the end of the forecast period using the VNDA, ONDA and NDA methods, respectively.

CHAPTER 7

SUMMARY AND CONCLUSIONS

In this thesis, we have systematically studied 4-D VDA applied to numerical weather prediction problems, i.e. solving the problems related to finding the best initial conditions such that the sum of weighted squares of the difference between model solution and the observations is minimized over a period of time. The 4-D VDA uses iterative large-scale unconstrained minimization algorithms in the framework of optimal control theory of PDEs to iteratively find aforementioned best initial conditions to a forecast model. Each iteration of the minimization algorithm requires the calculation of the gradient of the cost function with respect to the control variables. Theoretically, this gradient could be calculated by finite-difference approximation or automatic differentiation. But these approaches are too costly. Here the gradient of the cost function with respect to the control variables is found by integrating a first order adjoint model backwards in time from the final time to the initial time of the assimilation window while introducing the weighted differences between the model solution and the observations into the adjoint model whenever the observations are available. This gradient is then used to find a descent direction along which the cost function could be reduced. A step size will be calculated such that a maximum reduction in the cost function in the descent direction is obtained. From the step size and the descent direction, the approximate initial conditions to the forecast model are updated and the process of determining the gradient, descent direction, step size and the updated initial conditions is repeated until the minimization algorithm converges, i.e. satisfies a prescribed convergence criterion. The updated initial conditions at the

end of iterative minimization procedure are the best initial conditions minimizing the cost functional, while satisfying the forecast model equations as a strong constraint.

In Chapters 2 and 3, the thesis focuses on the analysis of the minimization process and the improvement of large-scale unconstrained minimization algorithms. Since these issues are strongly related to the structure of the Hessian of the cost function, a SOA theory is developed and applied to a shallow water equations model. The SOA model is similar to the FOA model in form (see Fig. 2.1). Given a perturbation vector U' on the initial condition vector U, the Hessian (H) vector product HU' is obtained by integrating the SOA model backwards in time from the final time to the initial time of the assimilation window while introducing the first order adjoint variables and the weighted tangent linear variables into the SOA model at each time step. Comparing with the finite-difference scheme (3.19), the above adjoint Hessian/vector calculation strategy yields a more accurate Hessian/vector product since it does not require the user to select a finite-difference parameter h as the finite-difference scheme does. The numerical cost of using the SOA approach is roughly the same as that of using the finite-difference approach, i.e. both approaches require four model integrations and the numerical cost of each model integration is approximately the same as any other model integrations. There are many applications of the SOA approach in calculating the Hessian/vector product. In this thesis, we applied the SOA approach to (a) the calculation of eigenvalues of the Hessian; (b) the calculation of the sensitivity of the cost function with respect to observational errors; (c) the implementation of the large-scale truncated Newton method [134] using a two dimensional limited-area shallow water equations model. The results presented here demonstrate that (a) the power method can be efficiently implemented to obtain estimates of the eigenvalues and the condition numbers of the Hessian matrix with the matrix/vector product calculated using SOA approach. The variation of the extremal eigenvalues with the number of the iterations indicates that most changes of the largest eigenvalue occur during the

first few iterations of the minimization procedure, which might explain why most of large-scale features are reconstructed earlier than the small scale features in the VDA retrieval solution during the minimization process [149]. The fact that the smallest eigenvalues of the Hessians of the cost function remain positive during the minimization process indicates the uniqueness of the optimal solution; (b) the sensitivity of the cost function with respect to observational errors depends on the time when the errors occur, the specific field containing the errors, and the spatial location where the errors occur. The cost function is more sensitive to the observational errors occurring at the beginning of the assimilation window, to errors in the geopotential field, and to errors at these grid point locations where intensive events occur; (c) although the ATN algorithm and TN algorithm of Nash [134] differ from each other only in the use of a more accurate Hessian/vector product for carrying out the large-scale unconstrained optimization required in VDA, the ATN algorithm yields results which are twice as fast as these obtained by the TN algorithm both in terms of number of iterations (i.e. convergence rate) as well as in terms of CPU time. Further the ATN algorithm turns out to be faster than the LBFGS method in terms of CPU time for the problem tested.

In Chapter 4 we described the FSUGSM in both physical and spectral spaces, its semi-implicit time differencing algorithm and its vertical discretization. In Chapter 5, the TLM and FOA models of the FSUGSM were first developed and then were verified for their correctness. The FOA model was then used to obtain the gradient of the cost function with respect to the control variables. The correctness of the calculation of this gradient was verified by using a Taylor expansion. Finally, we conducted a series of experiments to demonstrate the numerical feasibility of 4-D variational data assimilation with the FSUGSM. The 4- VDA in this chapter starts from two types of initial conditions: an analysis of the real observational data and a randomly perturbed initial condition of the above analysis. It is demonstrated that

the VDA performs successfully with different scenarios of observations distributed in time irrespective which type of initial conditions is used. The impact of observations being available at the beginning of the assimilation window is dramatic in so far as the decrease of the cost function is concerned since information on the entire field is in effect available. When observations are available both at the beginning and the end of the assimilation window, adding more observations in the middle of the assimilation window has only a weak impact on the convergence rate of the minimization process. It is also demonstrated that the presence of horizontal diffusion in the model yields a more accurate solution to the variational data-assimilation problem.

In Chapter 6, the model forecast error is considered by adding a non-physical diffusion-type term (i.e. the nudging term) to the model equations. The magnitude of this nudging term is also optimally estimated in the process of the minimization of a cost functional. We took advantage of the VDA's ability to perform optimal parameter estimation to obtain both optimal nudging coefficients and the best initial conditions simultaneously, i.e. the nudging coefficients are part of the control variables vector. Both the initial conditions and the nudging coefficients are updated after each iteration of the minimization process. Since at every minimization iteration a model integration has to be carried out with current estimated nudging coefficients used in the nudging term, the conditioning of the the Hessian matrix is improved. Thus variational nudging assimilation results in a faster convergence rate compared with 4-D VDA. The resulting optimal nudging coefficients are applied in the NDA, thus the term optimal NDA. The optimal NDA is practically implementable and performs very well in as far as the convergence and quality of the resulting assimilation state are concerned. This procedure is much more economical than the VDA owing to the tremendous computational cost of the VDA. Therefore the optimal NDA constitutes a good candidate for future data assimilation schemes where observations in a certain time window can be effectively incorporated into the model so as to obtain the best

initial conditions, i.e. physical initialization. Zou *et al.* [229] proved this method to be related to the KF method.

In subsection 6.3.2, we carried out VNDA experiments using the FSUGSM in its full-physics operational form and using a simplified adjoint model of the adiabatic version of the FSUGSM to calculate the gradient approximately. The results thus obtained are realistic and both the cost function and gradient norm in all experiments show a sufficient decrease. When the physical processes are included in forecast models, the numerical cost of model integration increases dramatically. Calculation of the gradient of function using a simplified model reduces the computational cost of the 4-D VDA at the cost of an error in the gradient. The main conclusion of this chapter is that the VNDA could be performed successfully using most sophisticated models while using a simplified adjoint model to calculate the gradient.

The VNDA converges faster than the VDA. It is also easy to implement. But the numerical cost is still prohibitive for operational implementation of the method. Preconditioning is an alternative to speed up the minimization process, but it is not easy to implement in practice. Due to the large dimension and high nonlinearity of the VDA problems, it is only possible to improve the conditioning to a limited extent. Therefore it is important to study new ways to formulate the problem and new algorithms to perform the minimization. Courtier *et al.* [37] has proposed to update the perturbations on the basic state instead of updating the state variable itself (incremental approach). It should be pointed out that the evolution of the perturbations is approximately described by the TLM. Thus the formulation is not exact. But it might be scientifically acceptable. To develop new algorithms, one should consider how to implement Newton's algorithm in the VDA. Newton's algorithm has a quadratic convergence rate. Thus it might be the only algorithm which could lead to operational implementation of the 4-D VDA before a generation of significantly faster high performance parallel computers being developed.

Developing high performance parallel computing algorithms is another alternative [21, 150]. For instance, in the adjoint truncated Newton algorithm, there are four model equations involved, namely, nonlinear model integration, TLM integration, FOA model integration and SOA model integration. Once the basic model integration is carried out, the TLM integration and FOA model integration can be carried out in a parallel manner. This strategy will improve the efficiency of the ATN algorithm. It is also possible to partition the computational domain into subdomains (domain decomposition) and carry out the VDA in each subdomain in parallel [21, 150], both for the forward nonlinear model as well as for the linear adjoint model.

APPENDIX A

A SIMPLE EXAMPLE WITH EXACT SOLUTIONS

In this appendix we will illustrate the whole process of using adjoint model techniques in the 4-dimensional data assimilation by presenting a very simple linear example. Because the example has exact solutions for the model equation, TLM, FOA and SOA models, the theory introduced in the Chapters 1 and 2 will be clearly illustrated.

Let us consider the following 1-dimensional model equation

$$\frac{\partial X}{\partial t} = -X, \tag{A.1}$$

$$X(0) = U, \tag{A.2}$$

where the time t runs from 0 to 1. Using a direct integration the solution of the model can be obtained as

$$X = Ue^{-t}. \tag{A.3}$$

Let us further suppose that the observations are model generated with the initial condition

$$X(0) = 1, \tag{A.4}$$

then from Eq. (A.3), the observation may be written as

$$X^o = e^{-t}. \tag{A.5}$$

Let us define the cost function as

$$J(U) = \frac{1}{2}\int_0^1 < (X - X^o), (X - X^o) > dt, \tag{A.6}$$

then

$$J(U) = \frac{(U-1)^2}{4}(1 - e^{-2}). \tag{A.7}$$

The first order adjoint model of Eqs. (A.1) and (A.2) may be written as

$$-\frac{\partial P}{\partial t} = -P + (X - X^o), \tag{A.8}$$

i. e.

$$-\frac{\partial P}{\partial t} = -P + (U - 1)e^{-t}, \tag{A.9}$$

$$P(1) = 0, \tag{A.10}$$

where P is the adjoint variable. The gradient of the cost function with respect to the initial conditions is given by

$$\nabla_U J = P(0). \tag{A.11}$$

Eq. (A.9) has an analytic solution of the following form

$$P(t) = \overline{P}e^t - e^t \int_0^t e^{-\tau}(U - 1)e^{-\tau} d\tau, \tag{A.12}$$

$$= \overline{P}e^t + \frac{(U - 1)}{2}(e^{-t} - e^t), \tag{A.13}$$

where $\overline{P}$ is a constant to be determined by the final condition (A.10). Therefore $\overline{P} = [(U - 1)/2](1 - e^{-2})$, and Eq. (A.10) yields

$$P(t) = \frac{(U - 1)}{2}(1 - e^{-2})e^t + \frac{(U - 1)}{2}(e^{-t} - e^t), \tag{A.14}$$

and

$$P(0) = \frac{(U - 1)}{2}(1 - e^{-2}). \tag{A.15}$$

Eq. (A.15) is exactly the gradient of the cost function (A.7) with respect to the initial condition U.

Let us now consider a perturbation, U', on the initial condition for X, U. The resulting perturbations for the variables X and P, $\hat{X}$ and $\hat{P}$, are obtained from Eqs. (A.1), (A.2) and (A.9) as

$$\frac{\partial \hat{X}}{\partial t} = -\hat{X}, \tag{A.16}$$

$$\hat{X}(0) = U', \tag{A.17}$$

$$-\frac{\partial \hat{P}}{\partial t} = -\hat{P} + \hat{X}, \tag{A.18}$$

$$\hat{P}(1) = 0, \tag{A.19}$$

Eqs. (A.18)-(A.19) and (A.16)-(A.17) are the SOA equations and the TLM for the model Eqs. (A.1) and (A.2), respectively. They have the following exact solutions

$$\hat{X} = U' e^{-t}, \tag{A.20}$$

and

$$\hat{P}(t) = \frac{U'}{2}(1 - e^{-2})e^{t} + \frac{U'}{2}(e^{-t} - e^{t}), \tag{A.21}$$

respectively. $\hat{P}(0) = U'(1-e^{-2})/2$ is exactly the product of the second order derivative of the cost function with respect to the initial condition $X(0) = U$ and the small perturbation U'.

It should be pointed out that only when the original model equations are linear, the TLM and SOA model yield an exact evolution of the perturbation due to the perturbation U' on the initial condition U and an exact Hessian/vector product, respectively, but the FOA model always yields an exact gradient of the cost function with respect to the control variables.

Since the model is linear with respect to the state variable X, the cost function J is a quadratic function of the initial condition U. If the model is not linear with respect to the state variable, then the cost function is generally not quadratic in the initial condition. Say for instance in the above example, let $F(X) = -X^2$, while the other conditions remain the same, then the model solution is

$$X(t) = \frac{U}{(tU + 1)}, \tag{A.22}$$

and the cost function may then be written as

$$J(U) \;=\; \frac{1}{2}\int_0^1 (\frac{U}{(tU+1)} - \frac{1}{(t+1)})^2 dt$$

$$= \frac{1}{2}[1 + U + \frac{2U}{(1-U)} ln \frac{(U+1)}{2} - \frac{(1+3U)}{2(U+1)}], \tag{A.23}$$

which is absolutely not a quadratic function in the initial condition U.

APPENDIX B

A NONLINEAR MODEL WITH EXACT SOLUTIONS.

This appendix will illustrate the adjoint process by using a quadratic model. It is emphasized that when the model equations are nonlinear, the corresponding TLM is a second order accurate approximation to the true evolution of the perturbations due to the perturbation on the initial condition of the model equations, the FOA model yields an exact gradient of the cost function with respect to the control variables, and the SOA model yields a second order accurate approximation to the exact Hessian/vector product.

Let us now consider the following 1-dimensional nonlinear model equation

$$\frac{\partial X}{\partial t} = -X^2, \tag{B.1}$$

$$X(0) = U, \tag{B.2}$$

where time t changes from 0 to 1. The exact solution of the model is

$$X(t) = \frac{U}{(tU+1)}. \tag{B.3}$$

If the observations are model generated with the initial condition

$$X(0) = 1, \tag{B.4}$$

then from Eq. (B.3), the observations may be written as

$$X(t) = \frac{1}{(t+1)}. \tag{B.5}$$

Let us define the cost function as

$$J(U) = \frac{1}{2}\int_0^1 < (X - X^o), (X - X^o) > dt, \tag{B.6}$$

i.e.

$$\begin{aligned} J(U) &= \frac{1}{2}\int_0^1 (\frac{U}{(tU+1)} - \frac{1}{(t+1)})^2 dt \\ &= \frac{1}{2}[1+U+\frac{2U}{(1-U)}\ln\frac{(U+1)}{2} - \frac{(1+3U)}{2(U+1)}]. \end{aligned} \tag{B.7}$$

The first order adjoint model of Eqs. (B.1) and (B.2) may be written as

$$-\frac{\partial P}{\partial t} = (-2X)^* P + (X - X^o), \tag{B.8}$$

i. e.

$$-\frac{\partial P}{\partial t} = -\frac{2U}{Ut+1}P + \frac{U}{(tU+1)} - \frac{1}{(t+1)}, \tag{B.9}$$

$$P(1) = 0, \tag{B.10}$$

where P is the adjoint variable. The gradient of the cost function with respect to the initial conditions is given by

$$\nabla_U J = P(0). \tag{B.11}$$

Eqs. (B.9-B.10) have an analytic solution of the following form

$$\begin{aligned} P(t) &= (tU+1)^2[\int (\frac{1}{(t+1)} - \frac{U}{(tU+1)})\frac{1}{(tU+1)^2}dt + c] \\ &= (tU+1)^2[\int \frac{dt}{(t+1)(tU+1)} + \frac{1}{2(tU+1)^2} + c] \\ &= (tU+1)^2[\frac{1}{(1-U)(tU+1)} + \frac{1}{2(tU+1)^2} + c], \end{aligned} \tag{B.12}$$

where c is a constant to be determined by the final condition (B.10). Therefore

$$c = -\frac{1}{1-U^2} - \frac{1}{(1-U)^2} ln(\frac{2}{U+1}) - \frac{1}{2(U+1)^2}, \tag{B.13}$$

and Eq. (B.12) yields

$$P(0) = \frac{1}{1-U} + \frac{1}{2} - \frac{1}{1-U^2} - \frac{1}{(1-U)^2}\ln(\frac{2}{U+1}) - \frac{1}{2(U+1)^2}. \tag{B.14}$$

Eq. (B.14) is exactly the gradient of the cost function (B.7) with respect to the initial condition U.

Let us now consider a perturbation, U', on the initial condition for X, U. The resulting perturbations for the variables X and P, $\hat{X}$ and $\hat{P}$, are obtained from Eqs. (B.1), (B.2) and (B.9-B.10) as

$$\frac{\partial \hat{X}}{\partial t} = -2X\hat{X}, \tag{B.15}$$

$$\hat{X}(0) = U', \tag{B.16}$$

$$-\frac{\partial \hat{P}}{\partial t} = (-2X)^*\hat{P} + \hat{X}, \tag{B.17}$$

$$\hat{P}(1) = 0, \tag{B.18}$$

Eqs. (B.15-B.16) and (B.17)-(B.18) are the TLM and SOA model of the model Eqs. (B.1) and (B.2), respectively. Substituting $X = U/(tU+1)$ into Eq. (B.15) one obtains the following exact solution of the TLM (B.15)

$$\hat{X} = \frac{U'}{(tU+1)^2}, \tag{B.19}$$

which is a quadratic approximation to

$$\frac{U+U'}{t(U+U')+1} - \frac{U}{t(U)+1}, \tag{B.20}$$

the exact evolution of the perturbation due to the perturbation U' on the initial condition U. Table B.1 summarizes the errors between the true evolution given by Eq. (B.20) and the approximation given by Eq. (B.19) at $U = 0.9$ and $t = 0.5$ due to different perturbations on the initial condition U. Table B.1 clearly indicates that

$$\frac{U+U'}{t(U+U')+1} - \frac{U}{t(U)+1} = \frac{U'}{(tU+1)^2} + O(\|U'\|^2). \tag{B.21}$$

Substituting $X = U/(tU+1)$ and the TLM solution (B.19) into SOA Eq. (B.17), one obtains

$$-\frac{\partial \hat{P}}{\partial t} = \frac{-2U}{(tU+1)}\hat{P} + \frac{U'}{(tU+1)^2}. \tag{B.22}$$

Table B.1: Errors between the true evolution given by Eq. (B.20) and the approximation given by Eq. (B.19) at $U = 0.9$ and $t = 0.5$ due to different perturbations on the initial condition U. $\vec{X}_2$ and $\vec{X}_1$ denote the model solutions starting from the initial conditions $U + U'$ and U, respectively.

$\lvert\vec{X}_2 - \vec{X}_1 - \hat{X}\rvert$	U'
$1.2195493765050 \times 10^{-1}$	1.
$1.5854141894547 \times 10^{-3}$	0.1
$1.6344476184732 \times 10^{-5}$	0.01
$1.6395183253561 \times 10^{-7}$	0.001
$1.6400272229676 \times 10^{-9}$	0.0001
$1.6399903435054 \times 10^{-11}$	0.00001
$1.6206145007512 \times 10^{-13}$	0.000001
$1.2021521456654 \times 10^{-15}$	0.0000001

Eqs. (B.22) and (B.18) have an analytical solution of the following form

$$\begin{aligned}\hat{P}(t) &= (tU+1)^2[\int -\frac{U'}{(tU+1)^2})\frac{1}{(tU+1)^2}dt + c] \\ &= U'(tU+1)^2[\int -\frac{dt}{(tU+1)^4} + c] \\ &= \frac{U'(tU+1)^2}{U}[\frac{1}{3(tU+1)^3} + c] \end{aligned} \tag{B.23}$$

where $c = -1/3(U+1)^3$ is a constant determined by the final condition (B.18). Therefore the exact solution of the SOA model is

$$\hat{P}(t) = \frac{U'(tU+1)^2}{U}[\frac{1}{3(tU+1)^3} - \frac{1}{3(U+1)^3}], \tag{B.24}$$

which yields

$$\hat{P}(0) = \frac{U'}{3U}[1 - \frac{1}{(U+1)^3}], \tag{B.25}$$

According to the SOA theory, Eq. (B.25) should be an approximation to the exact Hessian/vector product, $[\partial^2 J/\partial U^2]U'$,

$$\begin{aligned}\frac{\partial^2 J}{\partial U^2}U' &= \{\frac{1}{(1-U)^2} - \frac{2U}{(1-U^2)^2} \\ &- \frac{2}{(1-U)^3}\ln\frac{2}{(U+1)} + \frac{1}{(1-U)^2(1+U)} + \frac{1}{(1+U)^3}\}U', \quad \text{(B.26)}\end{aligned}$$

i.e.

$$\frac{\partial^2 J}{\partial U^2}U' = \hat{P}(0) + O(\|U'\|^2). \quad \text{(B.27)}$$

Table B.2 displays the errors defined by $[\partial^2 J/\partial U^2]U' - \hat{P}(0)$ for the present example. This Table indicates that the Hessian/vector product calculated by the SOA integration is second order accurate.

Table B.2: The absolute errors between the exact and approximate (obtained by second order adjoint integration) Hessian/vector product.

$\lvert\frac{\partial^2 J}{\partial^2 U}U' - \hat{P}(0)\rvert$	U'
$5.866191613327 \times 10^{-1}$	0.99
$1.2915525817253 \times 10^{-3}$	0.1
$1.1592929336223 \times 10^{-5}$	0.01
$1.1350261520048 \times 10^{-7}$	0.001
$5.0755928533783 \times 10^{-8}$	0.0001

APPENDIX C

TN: NASH'S TRUNCATED NEWTON METHOD.

The Truncated Newton algorithm of Nash [130, 134] is one of the most efficient large-scale unconstrained minimization algorithms. Schlick and Fogelson [185, 186] modified the truncated Newton algorithm and developed a FORTRAN package to solve problems for which some separability and sparsity structure of the Hessian is available. Here we give a general description of the truncated Newton algorithm of Nash [130, 134]. The algorithmic form of Nash's T-N method is:

1. Supply $\vec{X}_0$, an initial approximation to the minimizer $\vec{X}$ and set k=0.

2. Test $\vec{X}_k$ for convergence. Terminate if the following criteria are satisfied:

$$\|\vec{g}_k\| < 10^{-5} \cdot \|\vec{g}_0\|, \tag{C.1}$$

   where $\|\vec{g}_k\|$ is the Euclidean norm of the gradient of the cost function J_k with respect to the control variables, then stop. Otherwise continue.

3. Solve the Newton equations system

$$G_k \vec{d}_k = -\vec{g}_k, \tag{C.2}$$

   where G_k is an approximation to the Hessian matrix, using a modified-Lanczos algorithm with preconditioning [134] (also see the additional remarks at the end of this appendix). If the Hessian matrix is positive definite, the modified-Lanczos method is equivalent to the linear conjugate-gradient algorithm of Hestenes and Stiefel [84]. The modified-Lanczos method can be described briefly as follows:

(a) Start the Lanczos process with some vector $\vec{v}_1$, which is chosen as a multiple of $-\vec{g}$ and has $\vec{v}_1^T\vec{v}_1 = 1$. Set

$$\vec{d}_1 = -\vec{g}_k. \tag{C.3}$$

(b) For $q = 1, 2, \ldots$, calculate the matrix-vector product $\mathbf{G}\vec{v}_q$ which is approximated by a finite-differencing scheme along the gradient $\vec{g}$ (we omit the subscripts k for clarity):

$$\mathbf{G}\vec{v}_q = \frac{\vec{g}\left(\vec{X} + h\vec{v}\right) - \vec{g}\left(\vec{X}\right)}{h}, \tag{C.4}$$

where h is some small positive value such as the cubic root of machine precision. Note that this formula requires only one additional evaluation of the gradient in each inner iteration.

(c) Using the Lanczos recurrence relations

$$\beta_{j+1}\vec{v}_{j+1} = \mathbf{G}\vec{v}_j - \alpha_j\vec{v}_j - \beta_j\vec{v}_{j-1}, \qquad j = 1, 2, \ldots, q-1 \tag{C.5}$$

After q-th step this expression may be written as

$$\mathbf{G}\mathbf{V}_q = \mathbf{V}_q\mathbf{T}_q + \beta_{q+1}\vec{v}_{j+1}\vec{e}_q^T \tag{C.6}$$

where

$$\mathbf{T}_q = \begin{pmatrix} \alpha_1 & \beta_2 & & & & & \\ \beta_2 & \alpha_2 & \beta_3 & & & & \\ & \beta_3 & \alpha_3 & \cdot & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & \beta \\ & & & & & \beta_q & \alpha_q \end{pmatrix} \tag{C.7}$$

is a tridiagonal matrix approximating $\mathbf{G}$ and $\alpha_j = \vec{v}^T \mathbf{G} \vec{v}_j$. The value $\beta_{j+1} (\leq 0)$ is chosen so that $\|\vec{v}_{j+1}\|_2 = 1$, $\vec{e}_q$ is the q-th column of the $q \times q$ identity matrix.

(d) Calculate the Cholesky factorization factors of $\mathbf{T}_q$, $\mathbf{L}_q$ and $\mathbf{D}_q$, where

$$\mathbf{T}_q = \mathbf{L}_q \mathbf{D}_q \mathbf{L}_q^T. \tag{C.8}$$

(e) Determine the $(q+1)$ -th approximation for the k^{th} iterate of the search direction by the expression

$$\vec{d}_{q+1} = \vec{d}_q - \frac{\vec{r}_q^T \vec{v}_q}{D_q} \vec{u}_q. \tag{C.9}$$

The value of $\vec{u}_q$ is computed from the recurrence relations

$$\vec{u}_1 = \vec{v}_1, \quad \vec{u}_q = \vec{v}_q - l_q \vec{u}_{q-1}, \tag{C.10}$$

where l_q is subdiagonal element $(q-1)$ of the lower bidiagonal matrix $\mathbf{L}_q$, D_q is the q^{th} diagonal element and $\vec{r}_q$ is the residual of the Newton equations, $\vec{r}_q = \mathbf{G}\vec{d}_q + \vec{g}$.

(f) If the prescribed truncation criteria

$$1 - \frac{\vec{g}\vec{d}_q + \vec{r}_q \vec{d}_q}{\vec{g}\vec{d}_{q+1} + \vec{r}_{q+1}\vec{d}_{q+1}} \leq \frac{0.5}{q} \tag{C.11}$$

is satisfied, stop the inner iteration; otherwise continue the inner loop from step b).

4. Set $k = k + 1$ and update

$$\vec{X}_{k+1} = \vec{X}_k + \alpha_k \vec{d}_k, \tag{C.12}$$

where α_k is the step-size obtained by conducting a line search

$$J(\vec{X}_k + \alpha_k \vec{d}_k) = \min_{\alpha} J(\vec{X}_k + \alpha \vec{d}_k), \tag{C.13}$$

using Davidon's cubic interpolation method for the line search of the step size and which satisfies the following Wolfe conditions:

$$f(\vec{X}_k + \alpha_k \vec{d}_k) \leq f(\vec{X}_k) + \beta' \alpha_k \vec{g}_k^T \vec{d}_k, \tag{C.14}$$

and

$$\frac{\vec{g}_k(\vec{X}_k + \alpha_k \vec{d}_k)^T \vec{d}_k}{\vec{g}_k^T \vec{d}_k} \leq \beta, \tag{C.15}$$

where $\beta' = 0.0001$, $\beta = 0.9$. Go to step 2.

Additional Remarks: In the TN method the Newton equation (C.2) is solved by performing a limited number of iterations of the linear conjugate-gradient method. The iterations are terminated ("truncated") before the system is solved exactly. If a single linear iteration is used, $\vec{d}_k$ will be the steepest-descent direction $-\vec{g}_k$; if the sequence is not truncated, $\vec{d}_k$ will be the Newton direction. Thus, the algorithm computes a vector that interpolates between the steepest-descent direction and the Newton direction.

The TN method will be successful only when a good direction can be produced in a rather small number of linear conjugate-gradient iterations, and hence the use of a good preconditioning is essential. Because the number of iterations required to solve the Newton's equation is equal to the number of distinct eigenvalues in exact arithmetic, the idea of preconditioning transforms the Newton equations to a related system possessing a more favorable eigenvalue structure. Ideally, the condition number will be low, and the eigenvalues will cluster near 1. Theoretically, this facilitates the convergence of the C-G method. Newton's equations (C.2) are equivalent to the following equation

$$(\mathbf{M}^{-\frac{1}{2}}\mathbf{G}\mathbf{M}^{-\frac{1}{2}})\mathbf{M}^{\frac{1}{2}}\vec{X} = -\mathbf{M}^{-\frac{1}{2}}\vec{g}. \tag{C.16}$$

Let $\mathbf{R}$ denote the matrix $\mathbf{M}^{-\frac{1}{2}}\mathbf{G}\mathbf{M}^{-\frac{1}{2}}$; then $\mathbf{R}$ has the same eigenvalues as $\mathbf{M}^{-\frac{1}{2}}\mathbf{G}$, since $\mathbf{M}^{-\frac{1}{2}}\mathbf{R}\mathbf{M}^{-\frac{1}{2}} = \mathbf{M}^{-1}\mathbf{G}$.

The preconditioning matrix $\mathbf{M}$ may be chosen in several ways. One source is from the limited memory quasi-Newton method, where the matrix $\mathbf{H}_{k+1}$ is an approximation to the *inverse* of the Hessian and satisfies the "quasi-Newton condition" for k pairs of vectors $\{\vec{p}_k, \vec{q}_k\}$:

$$\vec{p}_j = \mathbf{H}_{k+1}\vec{q}_j, \quad j = 1, 2, \ldots, k. \tag{C.17}$$

where $\vec{p}_j = \vec{X}_{j+1} - \vec{X}_j, \vec{q}_j = \vec{g}_{j+1} - \vec{g}_j$. Since $\mathbf{G}\vec{p}_j = \vec{q}_j$,

$$\vec{p}_j = \mathbf{H}_{k+1}\mathbf{G}\vec{p}_j, \quad j = 1, 2, \ldots, k. \tag{C.18}$$

and the matrix $\mathbf{H}_{k+1}\mathbf{G}$ has k unit eigenvalues with eigenvectors $\{\vec{p}_j\}$. Therefore, this QN $\mathbf{H}_{k+1}$ may be used as the preconditioning matrix $\mathbf{M}^{-1}$. In this case, the vector $-\mathbf{H}_{k+1}\vec{g}_k$ will be the first nontrivial member of the sequence $\left\{\vec{d}_q\right\}$ and this search direction may be used to obtain a better reduction in the function than $-\vec{g}_k$. Thus the search direction will interpolate between the Newton direction and a quasi-Newton direction.

Since the matrix $\mathbf{H}$ is never stored explicitly, Nash [134] uses an approximation to the diagonal of the Hessian matrix.

APPENDIX D

THE CONTINUOUS FORM OF THE ADJOINT EQUATIONS

For the parameter estimation of nudging coefficients, the cost function J can be defined as an integral

$$J = \frac{1}{2}\int_{t_0}^{t_f} \{< W(\vec{X} - \vec{X}^o), \vec{X} - \vec{X}^o > + < K(G - \hat{G}), G - \hat{G} > dt. \tag{D.1}$$

The dynamical model Eq. (6.11) viewed as strong constraints can be enforced by introducing a set of undetermined Lagrange multipliers. This leads to the following Lagrange function

$$L(\vec{X}, G, P) = J + \int_{t_0}^{t_f} < P, -\frac{\partial \vec{X}}{\partial t} + F(\vec{X}) + G(\vec{X}^o - \vec{X}) > dt, \tag{D.2}$$

where P is a vector of Lagrange multiplies to be determined.

The constrained optimization problem is then replaced by a series of unconstrained minimization problems with respect to the variables U and G. By doing so, the problem of minimizing the cost function, subject to the model equations, becomes a problem of finding the stationary points of the Lagrange function. This in turn is equivalent to the determination of U and G subject to the condition that the gradient of the Lagrange function vanishes. This leads to the following first-order condition:

$$\frac{\partial L}{\partial X} = 0, \tag{D.3}$$

$$\frac{\partial L}{\partial P} = 0, \tag{D.4}$$

$$\frac{\partial L}{\partial G} = 0, \tag{D.5}$$

The solution of Eqs. (D.3-D.5) is called the stationary point of L. Eq. (D.5) results in

$$\nabla_G J + \int_{t_0}^{t_f} < (\vec{X}^o - \vec{X}), P > dt = 0, \tag{D.6}$$

which yields the gradient of the cost function with respect to G

$$\nabla_G J = -\int_{t_0}^{t_f} < (\vec{X}^o - \vec{X}), P > dt. \tag{D.7}$$

Eq. (D.4) recovers the original model Eq. (6.11), while (D.3) yields the adjoint equations

$$\frac{\partial J}{\partial X} - \frac{\partial}{\partial X}\int_{t_0}^{t_f} < P, \frac{\partial \vec{X}}{\partial t} - F(\vec{X}) - G(\vec{X}^o - \vec{X}) > dt = 0, \tag{D.8}$$

Substituting (D.1) into (D.8) and using integration by parts one obtains

$$\begin{aligned} \int_{t_0}^{t_f} W(\vec{X} - \vec{X}^o)dt \quad &- \quad \frac{\partial}{\partial X}\{P(t_f)\vec{X}(t_f) - P(t_0)\vec{X}(t_0) - \int_{t_0}^{t_f} < \vec{X}, \frac{\partial P}{\partial t} >\} \\ &+ \quad \int_{t_0}^{t_f} < \frac{\partial F}{\partial \vec{X}} - G, P > dt = 0. \end{aligned} \tag{D.9}$$

If we notice that

$$\frac{\partial}{\partial X}\{P(t_f)\vec{X}(t_f) - P(t_0)\vec{X}(t_0)\} = 0, \tag{D.10}$$

Eq. (D.9) becomes

$$\int_{t_0}^{t_f} \{\frac{\partial P}{\partial t} + \big(\frac{\partial F}{\partial \vec{X}}\big)^* P - G^* P + W(\vec{X} - \vec{X}^o)\}dt = 0, \tag{D.11}$$

for any length of assimilation window $t_f - t_0$. Thus the integrand should be zero, which results in the first order adjoint equation, given by the following equations

$$-\frac{\partial P}{\partial t} = \big(\frac{\partial F}{\partial \vec{X}}\big)^* P - G^* P + W(\vec{X} - \vec{X}^o), \tag{D.12}$$

$$P(t_f) = 0, \tag{D.13}$$

where P is the adjoint variable in the first order adjoint model and is identical to the vector of Lagrange multipliers.

APPENDIX E

KALMAN FILTERING, VARIATIONAL DATA ASSIMILATION AND VARIATIONAL NUDGING DATA ASSIMILATION

In this section we give a brief description of the Kalman filtering [68] and the connection among the KF, nudging, VNDA and VDA.

Starting from the forecast model Eqs. (1.4) and (1.5) which are advanced in discrete time step Δt, $\vec{X}_n = \vec{X}(t_n)$, $t_n = n\Delta t$, i.e.

$$\vec{X}_n^f = M_{n-1}\vec{X}_{n-1}^a, \tag{E.1}$$

where the superscript f denotes the forecast and the superscript a stands for analysis, and M is the discretized form of the system matrix, describing the model dynamics.

A linear unbiased data assimilation scheme for the analysed model state can be written as

$$\vec{X}_n^a = \vec{X}_n^f + G_n(\vec{X}_n^o - H_n\vec{X}_{n-1}^f), \tag{E.2}$$

where H represents the fact that only certain variables or combinations are observed as a set of points smaller than the total number of model grid points. The weight matrix G_n is often called the gain matrix, and the KF approach uses an optimal G_n to carry out such a linear unbiased data assimilation. The optimality is defined in the context of the following assumptions.

First assume the true evolution of the atmosphere, $\vec{X}_n^t$, is governed by

$$\vec{X}_n^t = M_{n-1}\vec{X}_{n-1}^t + b_{n-1}^t, \tag{E.3}$$

where b_{n-1}^t is a Gaussian white-noise sequence, i.e.

$$Eb_n^t = 0, \qquad Eb_n^t(b_l^t)^* = Q_n\delta_{nl}, \tag{E.4}$$

with E being the expectation operator and δ_{nl} being the Kronecker delta function. The second assumption used in optimizing the weight matrix G_n concerns the error model for the observations

$$\vec{X}_n^o = H_n \vec{X}_n^t + b_n^o, \tag{E.5}$$

where b_n^o, the observational noise, satisfies

$$Eb_n^o = 0, \qquad Eb_n^o (b_l^o)^* = R_n \delta_{nl}. \tag{E.6}$$

The third assumption is that system noise and observational noise are uncorrelated to each other:

$$Eb_n^t (b_n^o)^* = 0. \tag{E.7}$$

Using Eqs. (E.1)-(E.7), one can derive the time evolution of the error covariance matrix

$$\begin{aligned} W_n^a &\equiv E(\vec{X}_n^a - \vec{X}_k^t)^* (\vec{X}_n^a - \vec{X}_k^t) \\ &= (I - G_n H_n) W_n^f (I - G_n H_n)^* + G_n R_n G_n^*, \end{aligned} \tag{E.8}$$

where

$$W_n^f \equiv E(\vec{X}_n^f - \vec{X}_k^t)^* (\vec{X}_n^f - \vec{X}_k^t) = M_{k-1} W_{n-1}^a M_{k-1}^* + Q_{k-1}. \tag{E.9}$$

Hence, by advancing W_n^f and W_n^a along with $\vec{X}_n^f$ and $\vec{X}_n^a$, one can know how well the true state $\vec{X}_n^t$ is estimated by $\vec{X}_n^a$ for any weight matrix G_n. This in turn permits one to determine the optimal G_n by minimizing

$$J_{KF}(G_n) = tr(W_n^a), \tag{E.10}$$

where tr denotes the trace of the matrix.

The optimal weight matrix G_n at the n-th time step is obtained by using Eq. (E.8) for the matrix W_n^a and setting the derivative of $J_{KF}(G_n)$ with respect to each element of G_n equal to zero. The minimum is obtained at

$$G_n^* \equiv W_n^f H_n^* (H_n W_n^f H_n^* + R_n)^*. \tag{E.11}$$

The above linear unbiased data assimilation scheme given by Eqs. (E.1) and (E.2) with optimal gain matrix G_n^* is called the KF [90].

There are two problems which arise in the KF. The first is the computational complexity of advancing in time the error covariance matrices. While Eqs. (E.1) and (E.2) represent $O(N)$ computations per time step, Eqs. (E.8) and (E.9) represent, at face value, $O(N^2)$ computations where N is the dimension of the control variable. Second, the noise covariance matrices Q_n and R_n are assumed to be known in the subsequent derivation of the optimal G_n. This is not so in practice, and finding the actual magnitude of system and observational errors is an important function of the data assimilation process.

The nudging scheme is carried out in the following procedure

$$\vec{X}_n^a = M_{n-1}\vec{X}_{n-1}^a + G_n(\vec{X}_n^o - H_n\vec{X}_n^a). \tag{E.12}$$

The optimal nudging coefficients G_n are obtained by minimizing a cost function measuring the distance between the analysis and the observations.

$$J(U) = \frac{1}{2}\sum_{n=0}^{M} < W(C\vec{X}_n^a - \vec{X}_n^o), (C\vec{X}_n^a - \vec{X}_n^o) >, \tag{E.13}$$

From Eqs. (E.1) and (E.10) we see that the core of the KF is the optimal merging of observation and forecast information in the sense that the expected mean-square estimation error is minimized at every time step. The VNDA, on the other hand, is the optimal merging of observations and analysis in the sense that the total differences between them in a certain window of assimilation is minimized (see Eqs. (1.3) and (E.12)). The main differences between the KF and the VNDA are:

1. observation errors at different times are assumed to be uncorrelated;

2. the weight matrix G_n at each time step is determined sequentially in the KF, while the nudging coefficients at each time step in the window of assimilation

are obtained simultaneously. However, the two problems of the KF described above disappear in the VNDA. The computational cost is reduced by using an adjoint model integration in the parameter-estimation mode of VDA. Moreover, the VDNA does not require any knowledge of the noise covariance matrices.

Therefore, the estimate NDA, the VNDA and the KF differ from each other mainly in the choice of the weight matrix G_n. The VDA, on the other hand, takes both the model forecasts and the observations as perfect, i.e. $b_n^f = 0$ and $n_n^o = 0$ when $n \neq 0$. It attempts to obtain an optimal initial condition which minimizes the cost function

$$J(U) = \frac{1}{2} \sum_{n=0}^{M} < W(C\vec{X}_n^f - \vec{X}_n^o), (C\vec{X}_n^f - \vec{X}_n^o) >, \tag{E.14}$$

The theoretical framework of estimation and control theory provides the foundation of data assimilation techniques. The estimated NDA and the KF are closer to the estimation theory, the VDA to the optimal control aspect, while the VNDA is a combination of both [119].

In the nonlinear case, an extended Kalman filter (EKF) [163, 66] can be used. The EKF differs from the KF in that is uses a nonlinear model to make a forecast

$$\vec{X}_n^f = N_{n-1}\vec{X}_{n-1}^f, \tag{E.15}$$

where N represents the discretized form of the nonlinear system matrix; and the forecast error covariances are integrated with the tangent linear model, i.e. Eq. (E.9) is replaced by

$$W_n^f \equiv E(\vec{X}_n^f - \vec{X}_k^t)^*(\vec{X}_n^f - \vec{X}_k^t) = A_{k-1}W_{n-1}^a A_{k-1}^* + Q_{k-1}. \tag{E.16}$$

where A is the Jacobian matrix $\partial N/\partial \vec{X}$. The EKF brings new features concerning the time evolution of the forecast error covariances since the nonlinear model is linearized around a time and space dependent flow configuration to yield the tangent linear model.

For the sake of completeness, the entire extended KF algorithm is included here. It reads as follows:

$$\vec{X}_n^f = A_{n-1}\vec{X}_{n-1}^a, \tag{E.17}$$

$$W_n^f \equiv E(\vec{X}_n^f - \vec{X}_k^t)^*(\vec{X}_n^f - \vec{X}_k^t) = A_{k-1}W_{n-1}^a A_{k-1}^* + Q_{k-1}. \tag{E.18}$$

$$G_n^* \equiv W_n^f H_n^*(H_n W_n^f H_n^* + R_n)^*. \tag{E.19}$$

$$W_n^a = (I - G_n^* H_n)W_n^f, \tag{E.20}$$

$$\vec{X}_n^a = \vec{X}_n^f + G_n(\vec{X}_n^o - H_n\vec{X}_{n-1}^f). \tag{E.21}$$

where Eq. (E.20) is obtained from Eq. (E.8) under the assumption $G_n = G_n^*$.

APPENDIX F

PHYSICAL PROCESSES

The FSUGSM contains parameterizations of a variety of physical effects; these attempt to simulate diabatic heating, water vapor sources and frictional forces. Diabatic heating, due to long wave and short wave radiation, deep cumulus and dry convection, large scale condensation and sensible heat flux from the (land and sea) surface are all incorporated into this model. The model contains water vapor sources, due to deep cumulus convection, large scale condensation and (land and sea) surface flux. Frictional forces are included through calculations of the surface stress and a vertical diffusion of momentum and water vapor are also included in the FSUGSM, although not entirely for physical reasons.

F.0.1 Radiative processes

Forcing due to long wave and short wave radiation has been found to be important for tropical NWP [126]. The method used for the calculation of the radiative processes in the FSUGSM is essentially the same as that of Chang [25]. It was originally implemented in a limited area, multi-level grid point primitive equation model at FSU. The highlights of this scheme are a specification of the diurnal change (in the incoming short wave radiation) and a parameterizations of the effects of clouds (on both the long wave and short wave radiation).

Diurnal change in the incoming solar radiation S_o (the solar constant: here we use a value of 1393 Wm^{-2}) is effected by computing the zenith angle (Z) of the sun

as a function of local time of day (as well as its latitudinal and seasonal dependent). The total insolation, $S_o \cos Z$, is assumed to consist of two components, one which is scattered by the earth's atmosphere and another which is absorbed by it. The scattered and absorbed amounts are assigned to be 65.1% and 34.9% of the total, respectively. For the absorbed part of the short wave radiation (i.e., that part which is capable of being absorbed) an absorptivity function, dependent on water vapor content, temperature, pressure and zenith angle, is defined. This function determined the amount of depletion the solar radiation undergoes before reaching a given reference level. Both direct and reflected solar radiation are considered in the model's radiation package. The albedo of the earth's surface for summer season was tabulated from the works of Posey and Clapp [166] and Kondratyev [94] and interpolated to the Gaussian grid.

The presence (or lack) of single or multiple layers of clouds is empirically determined for the radiative computations. Three basic cloud types are permitted: low (centered at 850 mb), middle (centered at 700 mb) and high (centered at 500 mb). Clouds are specified from the relative humidity distribution, with minimum threshold values of 66%, 35% and 26%, respectively, for the presence of the above 3 cloud types. The fractional coverage of each of the types is then deduced from linear empirical functions of the relative humidity. The presence of clouds increases the absorptivity and causes partial reflection of the short wave radiation (the cloud albedos are empirically specified).

The scattered part of the total insolation is assumed to be unimportant in the atmospheric heating. However, in the heat balance of the earth's surface, to be discussed below, the scattered component is included in the total short wave radiation reaching the ground. Thus, the scattered short wave radiation does play a role in determining the ground temperature.

To compute the net long wave radiation (emitted by the earth's surface and the atmosphere) which reaches some reference level, the emissivity method is utilized. The emissivity of a given atmospheric layer is a function of optical depth (which in turn is taken to be a function of water vapor content, temperature and pressure). The long wave radiation emitted from the earth's surface is determined from the computed (or, in the case of oceanic points, specified) surface (ground or water) temperature. The net infrared radiation reaching some reference level is obtained from the emission of the atmospheric layers above and below that level, plus the terrestrial radiation (modified by the emissivity between the reference level and the surface).

Clouds are treated as black bodies for long wave radiation calculations. Thus, clouds modulate the long wave radiation in the model's atmosphere by producing enhanced long wave irradiance above and below the cloud layers. If a model level lies within a cloud, then the long wave irradiance at that level is set to zero. Future details on long wave (as well as short wave) radiation scheme are found in Chang [25].

The diabatic heating due to short wave and long wave radiation is determined from the convergence of the flux of radiant energy. Hence, the total (long wave plus short wave) radiative heating is given by

$$[\frac{\partial T}{\partial t}]_{rad} = \frac{-g}{C_p p_s}(\frac{\partial F_S}{\partial \sigma} + \frac{\partial F_L}{\partial \sigma}), \tag{F.1}$$

where F_S and F_L are the net short wave and long wave irradiances, respectively. Equation (F.1) is applied at the $\tilde{\sigma}$-levels of the FSUGSM; thus, the net irradiances are computed at the σ-levels. Finite differences in the vertical then yield the diabatic heating at the $\tilde{\sigma}$-levels.

F.0.2 Cumulus convection and large scale condensation

The diabatic heating caused by deep cumulus convection is recognized as a major energy source in the tropics. Deep convection also acts as a prominent sink of the

moisture in the large scale water vapor budget. Therefore, it is essential to include the effects of cumulus convection in a global prediction model, even if only in a parametric sense. An implicit treatment of the life cycles of the individual cumulus clouds and their interaction with the large scale fields is not practical for models such as the FSUGSM due to limitations of horizontal resolution. There are a number of existing schemes for the parameterization of cumulus convection in large scale prognostic models. The schemes which are most widely used are those of Arakawa and Schubert [6] and Kuo [101]. At FSU, a slightly modified version of Kuo's [101] parameterization method has been used in grid point, tropical prediction models with a fair degree of success ([92, 96]). The tractability of Kuo's method also makes it appealing for use in NWP. This method has therefore been adapted for the FSUGSM.

In the current model, the basic variables which are required for cumulus parameterizations are temperature, specific humidity and vertical motion ($\dot{\sigma}$, from which we can evaluate the vertical p-velocity, ω). These variables are available on the Gaussian grid points (at the $\tilde{\sigma}$-levels) at every time step. The large scale moisture supply, I, available for convective heating (or, equivalently, rainfall) is defined by

$$I = -\frac{1}{g}\int_{p_t}^{p_{cb}} \omega \frac{\partial}{\partial p} dp, \tag{F.2}$$

where p_t is the cloud top pressure (i.e., the pressure at which the moist adiabat, originating from p_{cb}, intersects the model sounding) and p_{cb} is the pressure of the cloud base. The cloud base is defined by the first model level (proceeding upward from the bottom) where the relative humidity equals or exceeds 70%. At any time step (for each Gaussian grid point), parameterized cumulus convection will be invoked if all of the following conditions are met: (a) the moisture supply, I, is positive; (b) the cloud base relative humidity criterion is satisfied; and (c) the atmosphere is conditionally unstable. Conditional instability exits if

$$\frac{\partial \theta}{\partial p} + \frac{\theta}{T}\frac{L(T_d)}{C_p}\frac{\partial r_S}{\partial p} \geq 0, \tag{F.3}$$

where θ is the potential temperature (although the same notation is used for latitude and potential temperature, there is no confusion if read consistently) and r_S is the saturation specific humidity.

Following Kuo [102], Kanamitsu [92] and Krishnamurti *et al.* [96] if the above condition are satisfied, the convective heating rate is given by

$$\frac{\partial T}{\partial t}|_{conv} = a_T\{\frac{T_S - T}{\Delta\tau} + \frac{T}{\theta}\omega\frac{\partial\theta}{\partial p}\}, \tag{F.4}$$

and the convective moistening rate is

$$\frac{\partial r}{\partial t}|_{conv} = a_r\{\frac{r_S - r}{\Delta\tau}\}. \tag{F.5}$$

In (F.4) and (F.5), T_S and r_S correspond to the saturation values of T and r (water-vapor mixing ratio), respectively, on the moist adiabat, which passes through p_{cb}, and $\Delta\tau$ is the cloud time scale (here, following earlier studies using Kuo's parameterization scheme with a grid-point model, $\Delta\tau$ is set to 1800 seconds). The coefficients a_T and a_r are determined by means of the following relations:

$$a_T = \frac{(1-b)I}{\frac{C_p}{L(T_d)g}\int_{p_t}^{p_{cb}}\{\frac{T_S-T}{\Delta\tau} + \frac{T}{\theta}\omega\frac{\partial\theta}{\partial p}\}dp}, \tag{F.6}$$

$$a_r = \frac{bI}{\frac{1}{g}\int_{p_t}^{p_{cb}}\{\frac{r_S-r}{\Delta\tau}\}dp}, \tag{F.7}$$

where the parameter b specifies the partitioning of the total moisture supply (I) into heating and moisture. The determination of b has been a subject of great interest to those scientists working with Kuo's method of convective parameterization. Kuo [102] presented observational evidence which suggests that b should be quite close to zero. Kanamitsu [92] proposed an *ad hoc* specification of b for use in a prognostic model. Anthes [5] gave a formula for b which is based on the large scale relative humidity distribution (in the vertical). However, none of the above formulations were entirely satisfactory from a physical point of view.

Krishnamurti *et al.* [97] utilized Kuo's method with the GATE phase III data (Sept. 1 through 18, 1974 over the eastern tropical Atlantic) to compute convective rainfall rates. In that study, the available moisture supply (I) was defined as in (F.2) and Kuo's [102] scheme was employed in a semi-prognostic manner (i.e., a series of single time step predictions). It was found that a very close agreement between observed and computed rainfall rates could be obtained by setting b to be a constant value of zero. In view of this and the inadequate physical justification for other types of closure, it was decided to set b to zero for present study. It should be noted that this implies that the convective moisture is zero in this case (by virtue of (F.5) and (F.7)). In other words, all of the large scale moisture supply defined by I is condensed out as rainfall. However, horizontal advection and diffusion of moisture (discussed below) can still contribute to moistening when the convective parameterization is invoked.

Large scale condensation heating (and rainfall) is simulated in the present model by the removal of supersaturation whenever it is observed in the prediction. This phenomenon typically occurs in regions of large scale ascent of air having absolutely stable (with respect to moist adiabatic) lapse rates, although other processes (such as radiative cooling) can also cause supersaturation. The numerical procedure for calculation of the supersaturation heating in the FSUGSM is the same as that described by Daley *et al.* [40]. At each time step, the model soundings of temperature and moisture are adjusted (if supersaturation is found) before any of the other physical forcing effects are evaluated. Any excess oversaturation is assumed to be condensed out as rain.

F.0.3 Boundary layer processes

A fairly simply treatment of the planetary boundary layer is used for the FSUGSM. Over the oceanic Gaussian grid points, the surface fluxes of sensible heat and water vapor and the surface stresses are calculated from bulk aerodynamics principles. Hence,

the sensible heat flux is given by

$$F_{sen} = 1.6 \times 10^{-3} C_p \rho_a (T_o - T_a) |\vec{V}_a|, \tag{F.8}$$

while the water vapor flux is specified by

$$F_{vap} = 1.4 \times 10^{-3} \rho_a (r_S T_o - r_a) |\vec{V}_a|, \tag{F.9}$$

and the surface stress is determined by

$$F_{mom} = |\vec{\tau}| = 1.1 \times 10^{-3} \rho_a |\vec{V}_a|^2, \tag{F.10}$$

where T_o is monthly climatological sea surface temperature (in this study, for June) taken from the work of Alexander and Mobley [1] while ρ_a, T_a, $\vec{V}_a$ and r_a are the surface (anemometer level) air density, temperature, wind and specific humidity, respectively. These quantities are obtained by extrapolation from the lowest model level, assuming a well mixed surface layer (in terms of potential temperature, wind and relative humidity). The numerical values of the surface exchange coefficients indicated in (F.8) through (F.10) are those suggested by Businger and Seguin [18] from GATE data.

Over land, the surface fluxes of sensible and latent heat are computed in conjunction with a determination of heat balance of the earth's surface. Daytime and nighttime are treated as separate cases in this formulation. If the net radiation (i.e., short wave plus long wave radiation) at the surface is negative (i.e. upward), then nighttime conditions are presumed. In this case, an empirically determined fraction of net radiation is used to lower the ground temperature. The remaining radiative deficit is assumed to be a downward flux of sensible heat which cools the surface air. The flux of latent heat is set to zero over land areas at night. The calculation of nighttime surface stress essentially follows the stable surface layer formulation of Businger *et al.* [17], which makes use of similarity theory.

For the daytime situation, the sensible and latent heat fluxes and the ground temperature are determined via the iterative solution of an explicit heat balance equation for the earth's surface, with the storage term neglected:

$$0 = R_{net} - \sigma_B T_o^4 - F_{sen} - L(T_d)F_{vap}, \tag{F.11}$$

where R_{net} is the net surface radiance (excluding the upward long wave irradiance) and σ_B is the Stefan-Boltzmann constant. The fluxes of sensible and latent heat, F_{sen} and $L(T_d)F_{vap}$, are given by relations similar to (F.8) and (F.9). However, rather than using the saturation specific humidity r_S at the ground temperature T_o in the calculation of F_{vap}, a ground wetness (saturation ratio) is used to modify r_S. This ground wetness, g_W, is an empirical function of α_S, the surface albedo,

$$g_W = 0.85[1 - e^{-200(0.25-\alpha_S)^2}]. \tag{F.12}$$

For a low albedo (less than 10%) the ground wetness approaches 0.85. For albedo values greater than 25%, g_W is set to zero. This formulation has been extensively tested and was found to avoid excessively high water vapor fluxes over tropical land areas. Eq. (F.11) is solved iteratively for the ground temperature T_o by Newton-Raphson technique. The daytime surface stress calculation essentially follows Businger *et al.* [17].

Above the surface level (over both land and ocean), the fluxes of heat, moisture and momentum are specified by the mixing length theory, as in Smagorinsky *et al.* [191]. That is, the fluxes of sensible heat, water vapor and momentum are respectively given by

$$F_{sen} = -\rho \ell_m^2 |\frac{\partial \vec{V}}{\partial z}| \frac{\partial \theta}{\partial z}, \tag{F.13}$$

$$F_{vap} = -\rho \ell_m^2 |\frac{\partial \vec{V}}{\partial z}| \frac{\partial r}{\partial z}, \tag{F.14}$$

$$F_{mom} = -\rho \ell_m^2 |\frac{\partial \vec{V}}{\partial z}| |\frac{\partial \vec{V}}{\partial z}|, \tag{F.15}$$

where ℓ_m^2 is the mixing length, which is a linear function of height. Above 3 km (from the surface topography), ℓ_m^2 vanishes and the fluxes are therefore zero. However, in the moisture prediction an additional flux, over that implied by (F.14), is used. This added flux is defined by an explicit vertical diffusion of water vapor. It is expressed by

$$\frac{\partial r}{\partial t}|_{diff} = K_r(\sigma)\frac{\partial^2 r}{\partial \sigma^2}, \tag{F.16}$$

where the diffusion coefficient $K_r(\sigma)$ is given by

$$K_r(\sigma) = (5\sigma - 1.25) \times 10^{-7} s^{-1}, \tag{F.17}$$

Above the $\sigma = 0.25$ level, $K_r(\sigma)$ is set to zero. Thus, the diffusion of moisture is stronger near the surface level. As noted earlier, in the discussion of the convective parameterization, all of the large (model) scale vertical advection of moisture is consumed for the convective rainfall. Here the inclusion of the vertical diffusion of moisture in vertical column, prevents excessive drying due to the cumulus convection.

F.0.4 Dry convective adjustment

Whenever the lapse rate of potential temperature is found to be dry unstable (superadiabatic) over a model grid point, a dry convective adjustment is carried out. This situation typically occurs at the lower levels, over land areas possessing relatively high ground temperatures (i.e., deserts). The adjustment procedure removes heat to the top of the layer, while conserving dry static energy ($gz + C_pT$). The end result of dry convective adjustment is a dry adiabatic lapse rate in the layer. In practice, this is done iteratively, checking the model's atmospheric column from top to bottom for the occurrence of superadiabatic layers. This is because the adjustment of a given layer may produce dry instability in the layer below.

F.0.5 Surface topography

Since the FSUGSM is a σ-coordinate model in the vertical direction, it allows for the realistic inclusion of the surface topography (since the earth's surface is a coordinate surface). The effects of orography in the prognostic equations appear mainly through the horizontal variations of q (i.e., $\ln p_s$, which is strongly dependent upon topography). Φ_S (the earth's topography) is spectrally truncated as the other variables in the model.

F.0.6 Computational requirements

The implementation of the transform technique in the FSUGSM requires that (in the computer code) memory storage be allocated to both grid-point and spectral coefficient representations of the dependent variables. All the spectral coefficients reside in the core simultaneously; however, only one latitude of the grid-point values (for all model levels) is stored at any instant. Contributions from each latitude to the spectral transforms are accumulated successively in the computer program, thus negating the need to store the entire grid in memory. Physical effects computations make use of column models, and therefore only require storage for the vertical levels.

REFERENCES

[1] Alexander, R. C. and Mobley, 1974: Monthly average sea surface temperatures and ice-pack limits on a 1° global grid. *Rand Corporation Report*, Santa Monica, CA, 90406, 30 pp.

[2] Andersson, E., A. Hollingsworth, G. Kelly, P. Lönnberg, J. Pailleux, and Z. Zhang, 1991: Global observing system experiments on operational statistical retrievals of satellite sounding data. *Mon. Wea. Rev.*, **199**, 1851-1864.

[3] Andersson, E., J. Pailleux, J. N. Thépaut, J. R. Eyre, A. P. McNally, G. A. Kelly and P. Courtier, 1992: Use of radiance in 3D/4D variational data assimilation. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 123-156.

[4] Anthes, R. A., 1974: Data assimilation and initialization of hurricane prediction models. *J. Atmos. Sci.*, **31**, 702-719.

[5] Anthes, R. A., 1977: A cumulus parameterization scheme utilizing an one dimensional cloud model. *Mon. Wea. Rev.*, **105**, 270-286.

[6] Arakawa, A. and Schubert W. H., 1974: Interaction of a cumulus cloud ensemble with the large scale environment, Part I. *J. Atmos. Sci.*, **31**, 674-701.

[7] Asselin, R., 1972: Frequency filter for time integrations. *Mon. Wea. Rev.*, **100**, 487-490.

[8] Banks, H. T. and K. Kunisch, 1989: *Estimation techniques for distributed parameter systems*. Birkhauser, Boston (Systems & Control: Formulations & Applications), Vol. 11, 315 pp.

[9] Bao, J. W. and T. T. Warner, 1993: Treatment of On/Off switches in the adjoint method: FDDA experiments with a simple model. *Tellus*, in press.

[10] Bertsekas, D. P. 1982: *Constrained optimization and Lagrange multiplier methods.* Addison-Wesley, 491 pp.

[11] Bernardet, P., S. Farges and K. Yassine, 1992: Adjoint of non-hydrostatic model. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 339-374.

[12] Bourke, W. 1972: An efficient one level primitive equation spectral model. *Mon. Wea. Rev.*, **100**, 683-689.

[13] Bourke, W. 1974: A multi-level spectral model. I –Formulation and hemisphere integrations. *Mon. Wea. Rev.*, **102**, 687-710.

[14] Buckley, A. G. and A. Lenir, 1983: QN-like variable storage conjugate gradients. *Math. Prog.*, **27**, 155-175.

[15] Burger, J., J. L. Brizaut and M. Pogu, 1992: Comparison of two methods for the calculation of the gradient and of the Hessian of the cost functions associated with differential systems. *Mathematics and Computers in Simulation*, **34**, 551-562.

[16] Bus, J. C. P., 1977: Convergence of Newton-like methods for solving systems of nonlinear equations. *Numerische Mathematik*, **27**, 271-281.

[17] Businger, J. A.,J. C. Wyngaard, Y. Izumi and E. F. Bradley, 1971: Flux-profile relationship in the atmospheric surface layer. *J. Atmos. Sci.*, **28**, 181-189.

[18] Businger, J. A. and W. Seguin, 1977: Transport across the air-sea interface: air-sea surface fluxes of latent and sensible heat and momentum. In Proc. of the GATE Workshop, published by the National Center for Atmospheric Research, Boulder, Colorado, 80303, 441-453.

[19] Cacuci, Dan G., 1981: Sensitivity theory for nonlinear systems. I: Nonlinear functional analysis approach. *J. Math. Phy.* **22** (12),2794-2803.

[20] Cacuci, Dan G., 1981: Sensitivity theory for nonlinear systems. II. Extensions to additional classes of responses. *J. Math. Phy.* **22** (12), 2803-2812.

[21] Cai, Y. and I. M. Navon, 1993: Iterative domain decomposition algorithms: Theory and application. High performance computing in the geoscience, edited by F.X. Le Dimet, Klumer Academic Publishers B.V. 000-000.

[22] Carrera, J. and P. S. Neuman, 1986: estimation of aquifer parameters under transient and steady state conditions, 1: Maximum likelihood method method incorporating prior information. *Water Resources Research*, **22** (2), 199-210.

[23] Carrera, J. and P. S. Neuman, 1986: estimation of aquifer parameters under transient and steady state conditions, 2: Uniqueness, stability and solution algorithms. *Water Resources Research*, **22** (2), 211-227.

[24] Carrera, J. and P. S. Neuman, 1986: estimation of aquifer parameters under transient and steady state conditions, 3: Application to synthetic and field data. *Water Resources Research*, **22** (2), 228-242.

[25] Chang, C. B., 1979: On the influences of solar radiation and diurnal variation of surface temperatures on African disturbances. *Rept. No.* 79-3, Dept. of Meteorology, Florida State University, Tallahassee, FL 32306, 157 pp.

[26] Chao W. C. and L. Chang, 1991: Development of a 4-dimensional analysis system using the adjoint method at GLA. part 1: dynamics. *Mon. Wea. Rev.*, **120**, 1661-1673.

[27] Charney, J. G., R. FjØrtoft and J. Von Neumann, 1950: Numerical integration of the barotropic vorticity equation. *Tellus,* **2**, 237-257.

[28] Charney, J. G., M. Halem and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.,* **2**, 1160-1163.

[29] Cohn, S. E., 1982: *Methods of sequential estimation for determining initial data in numerical weather prediction.* Ph.D Thesis, New York University, 183 pp.

[30] Cohn, S. E. and D. F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119**, 1757-1785.

[31] Cohn, S., 1993: The Kalman filtering versus variational analysis debate. Personal communication.

[32] Cohn, S, 1993: Dynamics of short-term univariate forecast error covariances. *Mon. Wea. Rev.*, **121** (11), 3123-3149.

[33] Courtier, P.,1985: Experiments in data assimilation using the adjoint model technique. Workshop on High-Resolution Analysis ECMWF (UK) June 1985.

[34] Courtier, P. and O.Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint equations Part 2. Numerical results.*Q. J. R. Meteorol. Soc.*, **113**, 1329-1347.

[35] Courtier, P., 1987: *Application du contrôle optimal à la prévision numérique en météorologie.* Thése de Doctorat de l'Université-Pierre-et-Marie, Paris.

[36] Courtier, P. and O.Talagrand, 1990: Variational assimilation of meteorological observations with direct and adjoint shallow-water equations. *Tellus.*, **42A**, 531-549.

[37] Courtier, P. and J. N. Thépaut, 1992: A strategy for operational implementation of 4D-VAR. In Proc. of ECMWF Workshop on variational assimilation, with special emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 437-463.

[38] Courtier, P., P. Undén, W. Heckley, D. Vasijevic, E. Andersson, J. N. Thépaut, F. Rabier and R. Hoffman, 1993: Progress in variational data assimilation at ECMWF. *Research activities in atmospheric and oceanic modeling,* G. J. Boer (ed.), *Report No.*, **18**, 1.42.

[39] Cressman, G., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367-374.

[40] Daley, R., Girard C., Henderson and J. Simmonds, 1976: Short-term forecasting with a multi-level spectral primitive equation model. Part I – Model Formulation. *Atmosphere*, **14** (12), 98-116.

[41] Daley, R., 1991: *Atmospheric data analysis.* Cambridge University Press, Cambridge.

[42] Daley, R, 1993: Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophysicae*, **11**, 634-647.

[43] Daley. R, and R. Menard, 1993: Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev,* **121,** (5), 1554-1565.

[44] Daley. R, and R. Menard, 1993: Evaluation of assimilation systems. Notes for three lectures at the international summer school on assimilation of meteorological and oceanographic observations. La Garde, France, August 2-27.

[45] Das, S. K. and R. W. Lardner, 1992: Variational parameter estimation for a two-demensional numerical tidal model. *International Journal for Numerical Methods in Fluids*, **15**, 313-327.

[46] Davidon, W. C., 1959: Variable metric method for minimization. A. E. C. *Research and Development Report*, ANL-5990 (Rev.).

[47] Davis H. E. and R. E. Turner, 1977: Updating prediction models by dynamic relaxation. An examination of the technique. *Q. J. R. Meteorol. Soc.*, **103**, 225-245

[48] Dembo, R. S., S. C. Eisenstat and T. Steihaug, 1982: Inexact Newton methods. *SIAM Journal of Numerical analysis,* **19**, 400-408.

[49] Dembo, R. S. and T. Steihaug, 1983: Truncated-Newton algorithms for large-scale unconstrained optimization. *Math. Prog.*, **26**, 190-212.

[50] Derber, J. C.,1985: *The variational four dimensional assimilation of analysis using filtered models as constraints.* Ph.D. Thesis, Univ. of Wisconsin-Madison, 141 pp.

[51] Derber, J. C.,1987: Variational four dimensional analysis using the quasi-geostrophic constraint. *Mon. Wea. Rev.*, **115**, 998-1008.

[52] Derber, J. C., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437-2446.

[53] Derber, J. C., D. F. Parrish and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Weather and Forecast,* **6**, 538-547.

[54] Derber, John C., David F. Parrish and Joseph G. Sela, 1992: The SSI analysis system and extensions to 4-D. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 15-29.

[55] Dixon, L. C. W., 1991: Use of automatic differentiation for calculating Hessians and Newton steps. *Automatic differentiation of algorithms: theory, implementation, and application.* Andreas Griewank and George F. Corliss (eds.), SIAM, Philadelphia, 115-125.

[56] Ehrendorfer, M., 1992: Four-dimensional data assimilation: comparison of variational and sequential algorithms. *Q. J. R. Meteorol. Soc.*, **118**, 673-713.

[57] Eliassen, A., 1954: Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. *Report #5*, Institute of Weather and Climate Res., Academy of Science, Oslo, 11 pp.

[58] Errico, R. M., Tomislava Vukićević, and Kevin Raeder, 1993: Examination of the accuracy of a tangent linear model. Submitted to *Tellus.*

[59] Eyre, J. R., G. A. Kelly, A. P. McNally and E. Andersson, 1993: Assimilation of TOVS radiance information through one dimensional variational analysis. *Research activities in atmospheric and oceanic modeling,* G. J. Boer (ed.), *Report No.*, **18**, 1.43.

[60] Fletcher, R., 1980: *Practical method of optimization. Vol. 1: Unconstrained optimization.* Wiley and Sons, New York, 120 pp.

[61] Fletcher, R., 1980: *Practical method of optimization. Vol. 2: Constrained optimization.* Wiley and Sons, New York, 224 pp.

[62] Fletcher, R., 1987: *Practical methods of optimization. Volume 1. 2nd ed.*, John Wiley, Chichester, New York.

[63] Gandin, L. S., 1963: *Objective analysis of meteorological fields.* Gidrometeorol. Izd. Leningrad (in Russian), English Translation by Israel Program for Scientific Translations, Jerusalem, 1965, 242 pp.

[64] Ghil, M., B. Shkoller and V. Yangarber, 1977: A balanced diagnostic system compatible with a barotropic prognostic model. *Mon. Wea. Rev.*, **105**, 1223-1238.

[65] Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Issacson, 1981: Application of estimation theory to numerical weather prediction. L. Bengtsson, M. Ghil and E. Kallen (eds.), *Dynamic Meteorology: Data assimilation methods*, Springer-Verlag, New York, 139-244.

[66] Ghil, M., S. E. Cohn, and A. Dalcher, 1982: Sequential estimation, data assimilation, and initialization. D. Williamson (ed.), The interaction between objective analysis and initialization, *Publ. Meteorol.*, 127 (Proc. 14th standard seminar), McGill University, Montreal, 83-97.

[67] Ghil, M., 1986: Sequential estimation and satellite data assimilation in meteorology and oceanography. Y. K. Sasaki, T. Gat-Chen, L. White, M. M. Zaman, C. Zieler, L. P. Chang and D. J. Rusk (ed.), *Variational Methods in Geosciences*, Elsevier, Amsterdam, 91-100.

[68] Ghil, M. and P. Malonotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**, 141-266.

[69] Gilbert, J. C., 1992: Automatic differentiation and iterative processes. *Optimization Methods and Software,* **1**, 13-21.

[70] Gill, P. E. and W. Murray, 1972: Quasi-Newton methods for unconstrained optimization. *J. Inst. Maths Applics*, **9**, 91-108.

[71] Gill, P. E. and W. Murray, 1974: *Numerical methods for unconstrained optimization.* Academic Press, London and New-York, 283 pp.

[72] Gill, P. E. and W. Murray, 1979: Newton-type methods for unconstrained and linearly constrained optimization. *Math. Prog.*, **28**, 311-350.

[73] Gill, P. E. and W. Murray, 1981: *Practical optimization.* Academic Press, 401 pp.

[74] Gill, P. E., W. Murray, M. C. Saunders and M. H. Wright, 1983: Computing forward-difference intervals for numerical optimization. *SIAM J. Sci. Stat. Comput.*, **4**, 310-321.

[75] Grammeltvedt, A., 1969: A survey of finite-difference schemes for the primitive equations for a barotropic fluid. *Mon. Wea. Rev.*, **97**, 387-404.

[76] Griewank, A. and Ph. L. Toint, 1982: Partitioned variable metric updates for large structured optimization problems. *Numer. Math.*, **39**, 119-137.

[77] Griewank, A. and George F. Corliss., 1991: *Automatic differentiation of algorithms: theory, implementation, and application,* SIAM, Philadelphia, 353 pp.

[78] Hall, M. C. G., D. G. Cacuci and M. E. Schlesinger, 1982: Sensitivity analysis of a Radiative-Convective Model by the Adjoint Method. *J. of the Atmos. Sci.*, **39**, 2038-2050.

[79] Hall, M. C. G. and D. G. Cacuci, 1983: Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models. *J. of Atmos. Sci.*, **40**, 2537-2546.

[80] Hall, M. C. G., 1986: Application of adjoint sensitivity theory to an atmospheric general circulation model. *J. Atmos. Sci.*, **43**, 2644-2657.

[81] Haltiner, G. J. and R. T. Williams, 1980: *Numerical prediction and dynamic meteorology.* John Wiley & Sons, New York, 477 pp.

[82] Hamming, R. W., 1973: *Numerical methods for scientists and engineers,* 2nd edition, McGraw-Hill and New York.

[83] Heckley, W. A., P. Courtier and E. Andersson, 1992: The ECMWF variational analysis: general formulation and use of background information. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 49-94.

[84] Hestenes M. and E. Stiefel, 1952: Methods of conjugate gradients for solving linear systems. *J. Res. National Bureau of Standards*, **45**, 409-436.

[85] Hoffmann, R. N., 1982: SASS wind ambiguity removal by direct minimization. *Mon. Wea. Rev.*, **110**, 434-445.

[86] Hoffmann, R. N., 1984: SASS wind ambiguity removal by direct minimization. Part II: Use of smoothness and dynamical constraints. *Mon. Wea. Rev.*, **112**, 1829-1852.

[87] Hoffmann, R. N., 1986: A four dimensional analysis exactly satisfying equations of motion. *Mon. Wea. Rev.*, **114**, 388-397.

[88] Hoke, J. E. and R. A. Anthes, 1976: The initialization of numerical models by a dynamic initialization technique. *Mon. Wea. Rev.*, **104**, 1551-1556.

[89] Horton, C. W., L. E. Reichl and V. G. Szebehely, (editors), 1983: *Long-time prediction in dynamics*, Wiley-Interscience, New York, 496 pp.

[90] Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *trans. ASME, J. Basic Eng.,* **82D**, 35-45.

[91] Kalman, R. E. and R. S. Bucy, 1961: New results in linear filtering and prediction theory. Trans. ASME, *J. Basic Eng.*, **82**, 35-45.

[92] Kanamitsu, M., 1975: On numerical prediction over a global tropical belt. *Rept. No.* 75-1, Dept. of Meteorology, Florida State University, Tallahassee, FL 32306, 282 pp.

[93] Kitamura, S. and S.Nakagiri, 1977: Identifiability of spatially varying and constant parameters in distributed systems of parabolic type. *SIAM, Journal on Control and Optimization*, **15**, 785-802.

[94] Kondratyev, K. Ya., 1972: *Radiation processes in the atmosphere.* World Meteorological Organization Publication No. 309, Geneva, Switzerland, 214 pp.

[95] Kontarev, G., 1980: The adjoint equation technique applied to meteorological problem. *ECMWF Tech. Rep.*, Reading, U. K., 21, 1-21 pp.

[96] Krishnamurti, T. N., Pan H. Chang C. B., Ploshay J. and Oodally W., 1979: Numerical weather prediction for GATE. *Quart. J. Roy. Met. Soc.*, **108**, 979-1010.

[97] Krishnamurti, T. N., Ramanathan Y., Pan H. Pasch R. J. and Molinari J., 1980: Cumulus parameterization and rainfall rates I. *Mon. Wea. Rev.*, **108**, 465-472.

[98] Krishnamurti, T. N., X. Jishan, H. S. Bedi, K. Ingles and D. Oosterhof, 1991: Physical initialization for numerical weather prediction over tropics, *Tellus,* **43A**, 53-81.

[99] Krishnamurti, T. N.,H. S. Bedi and K. Ingles, 1993: Physical initialization using SSM/I rain rates. *Tellus*, **45A**, 247-269.

[100] Krylov, V. I., 1962: *Approximate calculation of integrals*. ACM Monograph Series, the MacMillan Company, New York, NY.

[101] Kuo, H. L., 1965: On formulation and intensification of tropical cyclones through latent heat release by cumulus convection. *J. Atmos. Sci.*, **22**, 40-63.

[102] Kuo, H. L., 1974: Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *J. of Atmos. Sci.*, **31**, 1232-1240.

[103] Lacarra, J. F. and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. *Tellus,* **40A**, 81-95.

[104] Lardner, R. W., A. H. Al-Rabeh and N. Gunay, 1993: optimal estimation of parameters for a two-dimensional hydrodynamical model of the Arabian Gulf. Personal communication. Water Resource and Environment Division, Research Institute, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

[105] Lardner, R. W., 1993: Optimal control of open boundary conditions for a numerical tidal model. *Computer Methods in Applied Mechanics and Engineering*, **102**, 367-387.

[106] Le Dimet, F. X., 1982: A general formalism of variational analysis. *CIMMS Report*, Norman, OK 73091, 22, 1-34 pp.

[107] Le Dimet, F. X. and A. Nouailler,1986: *Assimilation of Dynamics in Geosciences*. Y. K. Sasaki, Editor, Elsevier, Amsterdam, 181-198.

[108] Le Dimet, F. X. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97-110.

[109] Le Dimet, F. X. and I. M. Navon, 1988: Variational and optimization methods in meteorology: a review. *Tech. Rep. FSU-SCRI-88-144,* Florida State University, Tallahassee, Florida, 83 pp.

[110] Legler, D. M., I. M. Navon and J. J. O'Brien, 1989: Objective analysis and assimilation of meteorological observations: theoretical aspects. *Mon. Wea. Rev.*, **38A**, 97-110.

[111] Legler, D. M. and I. M. Navon, 1991: VARIATM–A Fortran code for objective analysis of pseudo-stress with large-scale conjugate-gradient minimization. *Computers and Geosciences,* **17** (1), 1-21.

[112] Lewis, J. M. and J. C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 309-322.

[113] Li, Yong, I. M. Navon, P. Courtier and P. Gauthier, 1993: Variational data assimilation with a semi-implicit semi-Lagrangian global shallow-water equation model and its adjoint. *Mon. Wea. Rev.*, **121**, 159-169.

[114] Li, Y. and K. K. Droegemeier, 1993: The influence of diffusion and associated errors on the adjoint data assimilation technique. *Tellus*, **45A**, 1-14.

[115] Li, Y., I. M. Navon, W. Y. Yang, X. L. Zou, J. R. Bates, S. Moorthi and R. W. Higgins, 1993: 4-D assimilation experiments with a multilevel semi-Lagrangian semi-implicit GCM. *Mon. Wea. Rev.*, in press.

[116] Lions J. L., 1971: *Optimal control of systems governed by partial differential equations*. Translated by S. K. Mitter, Springer-Verlag, Berlin-Heidelberg, 404 pp.

[117] Liu, D. C and Jorge Nocedal, 1989: On the limited memory BFGS method for large scale minimization. *Math. Prog.*, **45**, 503-528.

[118] Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701-721.

[119] Lorenc, A. C.,1986: Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177-1194.

[120] Lorenc, A. C., 1988 a: Optimal nonlinear objective analysis. *Q. J. R. Meteorol. Soc.*, **114**, 205 -240.

[121] Lorenc, A. C., 1988 b: A practical approximation to optimal four dimensional objective analysis. *Mon. Wea. Rev.*, **116**, 730-745.

[122] Lorenc, A. C., R. S. Bell and B. Macpherson, 1991: The meteorological office analysis correction data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **117**, 59-89.

[123] Luenberger, David G.,1984: *Linear and nonlinear programming*. 2nd edition, Addison-Wesley Inc., Reading, MA, 491 pp.

[124] Lyne, W. H., R. Swinbank and N. T. Birch, 1982: A data assimilation experiment and the global circulation during the FGGE special observing periods. *Q. J. R. Meteorol. Soc.*, **108**, 575-594.

[125] Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model, with application to nonlinear normal mode initialization. *Beitr. Phys. Atmos.*, **50**, 253-271.

[126] Manabe, S., D. G.Hahn, and Jr. J. L. Holloway, 1974: The seasonal variation of the tropical circulation as simulated by a global model of the atmosphere. *J. Atmos. Sci.*, **31**, 43-83.

[127] Marchuk, G. I., 1974: *Numerical solution of the problem of dynamics of atmosphere and ocean* (in Russian). Leningrad, Gidrometeoizdat, 303 pp.

[128] Moore, A. M, and B. F. Farrell, 1993: Rapid perturbation growth on spatially and temporally varying oceanic flows determined using adjoint method: application to the gulf stream. *J. Phys. Oceanogr.*, **23**, 1682-1702.

[129] Nash, S. G., 1982: *Truncated-Newton methods.* Ph.D. Thesis, Computer Science Department, Stanford University, CA.

[130] Nash, S. G., 1984: Newton-type minimization via the Lanczos method. *SIAM J. Numer. Anal.*, **21** (4), 770-788.

[131] Nash, S. G., 1984: Truncated-Newton methods for large-scale function minimization. *Applications of Nonlinear Programming to Optimization and Control,* H.E. Rauch (ed.), Pergamon Press, Oxford, 91-100.

[132] Nash, S. G., 1984: User's guide for TN/TNBC: Fortran routines for nonlinear optimization. *Tech. Rep. No.*, **397**, 17 pp.

[133] Nash, S. G., 1984: Solving nonlinear programming problems using truncated-Newton techniques. *Numerical optimization,* P. T. Boggs, R. H. Byrd and R. B. Schnabel (ed.), SIAM, Philadelphia, 119-136.

[134] Nash, S. G., 1985: Preconditioning of truncated-Newton methods. *SIAM J. Sci. Stat. Comput.*, **6** (3), 599-616.

[135] Nash, S. G. and A. Sofer, 1989: Block truncated-Newton methods for parallel optimization. *Math. Prog.*, **45**, 529-546.

[136] Nash, S. G. and A. Sofer, 1989: A parallel line search for Newton type methods in computer science and statistics. In *Proc. 21-st Symposium on the Interface*, K. Berk and L. Malone (ed.), ASA, 134-137.

[137] Nash, S. G. and Jorge Nocedal, 1989: A numerical study of the limited memory BFGS method and the truncated-Newton method for large scale optimization. *Tech. Rep. NAM*, **02**, Dept. of Electrical Engineering and Computer Science, Northwestern University, Dec., 19 pp.

[138] Nash, S. G. and A. Sofer, 1990: Assessing a search direction within a truncated-Newton method. *Operations Research Letters,* **9** (4), 219-221.

[139] Navon, I. M., 1981: Implementation of 'a posteriori' method for enforcing conservation of potential enstrophy and mass in discretized shallow water equations models. *Mon. Wea. Rev.*, **109**, 946-958.

[140] Navon, I. M., 1982: 'A posteriori' numerical techniques for enforcing simultaneous conservation of the integral invariants. *Notes on numerical fluid mechanics*, Henri Viviand (ed.), GAMM, *Vol. 2*, 230-241.

[141] Navon, I. M. and R. de Villiers, 1983: Combined penalty multiplier optimization methods to enforce integral invariants conservation. *Mon. Wea. Rev.*, **111**, 1228-1243.

[142] Navon, I. M., 1985: A review of variational optimization methods in meteorology. Proc. of the International Symposium on Variational Methods in Geosciences, Y. K. Sasaki (ed.), University of Oklahoma, Norman, OK 1–5, october 15–17.

[143] Navon, I. M., 1986: A review of variational and optimization methods in meteorology. Festive Volume of the International Symposium on Variational Methods in Geosciences (Y.K. Sasaki, Ed.), Elsevier Science Pub. Co. Developments in Geo-mathematics, *Vol. 5*, 29-34.

[144] Navon, I. M. and D. M. Legler, 1987: Conjugate gradient methods for large scale minimization in meteorology. *Mon. Wea. Rev.*, **115**, 1479-1502.

[145] Navon, I. M., 1987: The Bayliss-Isaacson algorithm and the constraint restoration method are equivalent. *Meteorology and Atmospheric Physics*, **13**, 143-152.

[146] Navon, I. M., P. K. H. Phua and M. Ramamurthy, 1988: Vectorization of conjugate-gradient methods for large-scale minimization. In Proceeding Supercomputing'88, IEEE computer society and ACM SIGARCH, WASHINGTON, DC, (Reviewed by the IEEE Computer Society Scientific Review Board), 410-419.

[147] Navon, I. M., P. K. H. Phua and M. Ramamurthy, 1990: Vectorization of conjugate-gradient methods for large-scale minimization. *J. O. T. A.,* **66** (1), 71-94.

[148] Navon, I. M. and X. Zou, 1991: Application of the adjoint model in meteorology. In Proceedings of the international conference on automatic differentiation of algorithms: Theory, Implementation and Application, SIAM Publications, A. Griewank and G. Corliss (Eds.), SIAM, Philadelphia, PA.

[149] Navon, I. M., X. L. Zou, J. Derber and J. Sela, 1992: Variational data assimilation with an adiabatic version of the NMC Spectral Model. *Mon. Wea. Rev.*,**122**, 1433-1446.

[150] Navon, I. M. and Y. Cai, 1993: Domain decomposition and parallel processing of a finite-element model of the shallow-water equations. *Computer Methods in Applied Mechanics and Engineering*, **106**, No. 1-2, 179-212.

[151] Navon, I. M., 1993: Application of variational methodology in the context of a diabatic version of NASA/GLA SLSI model. Proposal to NASA, 77 pp.

[152] Navon, I. M., N. J. Zabusky and X. Zou, 1993: Variational data assimilation with a coarse-resolution shallow water equations model using high-resolution shallow water equations model. *Research activities in atmospheric and oceanic modeling,* G. J. Boer (ed.), *Report No.*, **18**, 1.46-1.47.

[153] Neuman, S. P., 1980: A statistical approach to the inverse problem of aquifer hydrology, 3. Improved solution method and added perspective. *Water Resources Research,* **16** (2), 331-346.

[154] Nocedal, J., 1980: Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, **35**, 773-782.

[155] O'Leary, D. P., 1983: A discrete Newton algorithm for minimizing a function of many variables. *Math. Prog.*, **23**, 20-23.

[156] Orszag, Steven A., 1970: Transform method for calculation of vector coupled sums: application to the spectral form of the vorticity equation. *J. Atmos. Sci.*, **27**, 890-895.

[157] Pailleux, J., W. Heckly, D. Vasiljevic, J. N. Thépaut, F. Rabier, C. Cardinali and E. Andersson, 1991: Development of a variational assimilation system. ECMWF Research Department Technical Memorandum 179, Reading, U.K..

[158] Pailleux, J., 1992: Organisation of 3D variational analysis within the "IFS/ARPEGE" project future plans at meteo France. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 37-46.

[159] Panchang, V. G. and J. J. O'Brien, 1989: On the determination of hydraulic model parameters using the strong constraint formulation. in *Modeling Marine Systems,* **I**. Ed. A. M. Davies. CRC Press, Inc, 5-18 pp.

[160] Panofsky, H., 1949: Objective weather map analysis. *J. Meteorol.*, **6**, 386-392.

[161] Parrish, D. and J. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

[162] Pasch, P. J., 1983: *On the onset of the planetary scale monsoon.* Ph.D. Thesis, Department of Meteorology, FSU, Tallahassee, FL.

[163] Gauthier, P. and P. Courtier, 1992: Data assimilation with an extended Kalman filter. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 171-190.

[164] Penenko, V. and N. N. Obraztsov, 1976: A variational initialization method for the fields of the meteorological elements. *Meteorol. Gidrol.* (English translation), **11**, 1-11.

[165] Phillips, N. A., 1976: The impact of synoptic observing and analysis systems on flow pattern forecasts. *Bull. Am. Meteorol. Soc.,*, **57**, 1225-1328.

[166] Posey, J. W. and P. F. Clapp, 1964: Global distribution of normal surface albedo. *Geofisica Internacional (Mexico),* **4**, 33-48.

[167] Powell, M. J. D., 1982: *Nonlinear optimization 1981.* Academic Press, London, 298 pp.

[168] Rabier, F. P. Courtier and O. Talagrand, 1992: An application of adjoint models to sensitivity analysis. *Beitr. Phys. Atmos.*, **65** (3), 177-192.

[169] Rabier, F. and P. Courtier, 1992: Four-dimensional assimilation in the presence of baroclinic instability. *Q. J. R. Meteorol. Soc.*, **118**, 649-672.

[170] Rabier, F., P. Courtier, J. Pailleux, O. Talagrand, J. N. Thepaut and D. Vasiljevic, 1992: Comparison of four-dimensional variational assimilation with simplified sequential assimilation. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 271-325.

[171] Rabier, F., P. Courtier, J. Pailleux, O. Talagrand, and D. Vasiljevic, 1993: A comparison between four-dimensional variational assimilation and simplified sequential assimilation relying on three-dimensional variational analysis. *Q. J. R. Meteorol. Soc.* **119**, 845-880.

[172] Ramamurthy, M. K. and F. H. Carr, 1987: Four-dimensional data assimilation in the monsoon region. Part I: Experiments with wind data. *Mon. Wea. Rev.*, **115**, 1678-1706.

[173] Ramamurthy, M. K. and I.M. Navon, 1990: Application of a conjugate-gradient method to variational assimilation of meteorological fields. International symposium on assimilation of observations in meteorology and oceanography, O. Talagrand and F. X. Le Dimet (eds.), 359-365.

[174] Ramamurthy, M. K. and I. M. Navon, 1992: The conjugate-gradient variational analysis and initialization method: An application to MONEX SOP-2 Data. *Mon. Wea. Rev.*, **120** (10), 2360-2377.

[175] Ramamurthy, M. K, and T. Y. Xu, 1993: Continuous data assimilation experiments with the NMC Eta model: A GALE IOP 1 Case study. *Mon. Wea. Rev.*, **121** (11), 3082-3105.

[176] Robert, A. J., J. Henderson and C. Turnbull, 1972: An implicit time integration scheme for baroclinic models of the atmosphere. *Mon. Wea. Rev.*, **100**, 329-335.

[177] Robertson, A. W, 1992: the investigation of regional climate anomalies with a linear barotropic model and an adjoint technique. *Q. J. R. Meteorol. Soc.,* **118**, 1187-1209.

[178] Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.,* **29**, 809-815.

[179] Sasaki, Y. K., 1955: A functional study of the numerical prediction based on the variational principle. *J. Meteor. Soc. Japan,* **33**, 262-275.

[180] Sasaki, Y. K., 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan,* **36**, 77-88.

[181] Sasaki, Y. K., 1969: Proposed inclusion of time-variation terms, observational and theoretical in numerical variational objective analysis. *J. Meteor. Soc. Japan,* **47**, 115-203.

[182] Sasaki, Y. K., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, **98**, 857-883.

[183] Sasaki, Y. K., 1970: Numerical variational analysis formulated under the constraints as determined by long-wave equations as a low-pass filter. *Mon. Wea. Rev.*, **98**, 884-898.

[184] Sasaki, Y. K., 1970: Numerical variational analysis with weak constrain and application to the surface analysis as a low-pass filter. *Mon. Wea. Rev.*, **98**, 899-910.

[185] Schlick, T. and A. Fogelson, 1992: TNPACK–A truncated Newton minimization Package for large-scale problems: I. Algorithm and usage. *ACMTOMS*, **18** (1), 46-70.

[186] Schlick, T. and A. Fogelson, 1992: TNPACK–A Truncated Newton minimization package for large-scale problems: II. Implementation examples. *ACMTOMS*, **18** (1), 71-111.

[187] Schlick, T., 1993: Modified Cholesky factorizations for sparse preconditioners. *SIAM J. Sci. Comput.,* **14**, 424-445.

[188] Schlitzer. R., 1993: Determining the mean, large-scale circulation of the atlantic with the adjoint method. *J. Phys. Oceanogr.* **23**, 1935-1952.

[189] Shanno, D. F., 1978: Conjugate-gradient methods with inexact searches. *Methods Oper. Res.*, **3**, 244-256.

[190] Shanno, D. F. and K. H. Phua, 1980: Remark on algorithm 500 – a variable method subroutine for unconstrained nonlinear minimization. *ACMTOMS*, **6**, 618-622.

[191] Smagorinsky, J., S. Manabe and J. L. Holloway Jr., 1965: Numerical results from a nine-level general circulation model of the atmosphere. *Mon. Wea. Rev.*, **93**, 727-768.

[192] Smagorinsky, J., K. Miyakoda and R. Strickler, 1970: The relative importance of variables in initial conditions for dynamical weather prediction. *Tellus,* **122**, 141-157.

[193] Smedstad, O. M. and J. J. O'Brien, 1991: Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model. *Prog. Oceanogr.*, **26**, 1250-1277.

[194] Staufer, D. R. and M. L. Seaman, 1990: Use of four-dimensional data assimilation on a limited area mesoscale model. Part I: Experiments with synoptic-scale data. *Mon. Wea. Rev.*, **118**, 1250-1277.

[195] Stensrud, D. J. and J. W. Bao, 1992: Behaviors of variational and nudging assimilation techniques with a chaotic low-order model. *Mon. Wea. Rev.*, **120**, 3016-3028.

[196] Stephens, J. J., 1966: *A variational approach to numerical weather analysis and prediction.* Ph.D. Thesis, Texas A&M University, College-Station, TX, 77863, Rep. 3, 243 pp.

[197] Stephens, J. J., 1968: variational initialization with the balance equation. *Mon. Wea. Rev.*, **96** 229-231.

[198] Stewart III, G. W., 1967: A modification of Davidon's minimization method to accept difference approximations of derivatives. *Journal of the Association for Computing Machinery,* **14** (1), 72-83.

[199] Strang, Gilbert, 1986: *Introduction to applied mathematics.* Wellesley-Cambridge Press, 758 pp.

[200] Sykes, J. F., J. L. Wilson and R. W. Andrews, 1985: Sensitivity analysis for steady state groundwater flow using adjoint operators. *Water Resources Research*, Vol. **21**, No. 3, 359-371.

[201] Symes, W., 1993: Personal communication. Dept. of Math., Rice University, Houston, TX.

[202] Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation-Part 1. Theory. *Q. J. R. Meteorol. Soc.*, **113**, 1311-1328.

[203] Tarantola, A., 1987: *Inverse problem theory. methods for data fitting and model parameter estimation.* Elsevier, Amsterdam, 613 pp.

[204] Thacker, W. C. and R. B. Long,1988: Fitting dynamics to data. *J. Geophys. Res.*, **93**, 1227-1240.

[205] Thacker, W. C., 1989: The role of Hessian matrix in fitting models to measurements. *J. Geophys. Res.*, **94**, 6177-6196.

[206] Thacker, W. C., 1992: Oceanographic inverse problems. *Physica D*, **60**, 16-37.

[207] Thépaut, J. N. and P. Courtier, 1991: Four-dimensional variational assimilation using the adjoint of a multilevel primitive-equation model. *Q. J. R. Meteorol. Soc.*, **117**, 1225-1254.

[208] Thépaut J.-N., D. Vasiljevic, P. Courtier and J. Pailleux, 1992: Variational assimilation of meteorological observations with a multilevel primitive-equation model. *Q. J. R. Meteorol. Soc.*, **119**, 153-186

[209] Thépaut, J. N., P. Courtier, and R. H. Hoffman, 1992: Use of dynamical information in a 4-D variational assimilation. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 237-269.

[210] Thépaut, J. D. Vasiljevic and P. Courtier, 1993: Variational assimilation of conventional meteorological observations with a multilevel primitive-equation model. *Q. J. R. Meteorol. Soc.*, **119**, 153-186.

[211] Thomas, R. Cuthbert, Jr., 1987: *Optimization using personal computers.* John Wiley & Sons, New York, 474 pp.

[212] Todling, R., 1992: *The Kalman filter for two-dimensional stable and unstable atmospheres.* Ph.D. Thesis, University of California at Los Angeles, 171 pp.

[213] Vasiljevic, D., C. Cardinali and P. Unden, 1992: ECMWF 3D variational data assimilation of conventional observations. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 389-426.

[214] Vukicevic, T., 1991: Nonlinear and linear evolution of initial forecast errors. *Mon. Wea. Rev.*, **119**, 1602-1611.

[215] Vukicevic, T. and R. M. Errico, 1992: Linearization and adjoint of parameterized moist diabatic processes. NCAR 1030/92-1.

[216] Vukicevic, T., 1993: Adjoints of discontinuous processes. *Tellus*, in press.

[217] Wang, W. and T. T. Warner, 1988: Use of four-dimensional data assimilation by Newtonian relaxation and latent heat forcing to improve a mesoscale model precipitation forecast: A case study. *Mon. Wea. Rev.*, **116**, 2593-2613.

[218] Wang, Z., I. M. Navon, F. X. Le Dimet and X. Zou, 1992: The second order adjoint analysis: Theory and application. *Meteorol. and Atmos. Phy.*, **50**, 3-20.

[219] Wang, Z., I. M. Navon and X. Zou, 1993: The adjoint truncated Newton algorithm for large-scale unconstrained optimization. *Tech. Rep. FSU-SCRI-92-170,* Florida State University, Tallahassee, Florida, 44 pp.

[220] Washington, W. M. and R. J. Duquet, 1963: *An objective analysis of stratospheric data by Sasaki's method.* Dept. of Meteorology, Penn. State University, University Park, PA, 16802, 83 pp.

[221] Wergen, W., 1992: Effect of model errors in variational assimilation (abstract). In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, 465-466.

[222] Wiener, N., 1949: *Extrapolation, interpolation and smoothing of stationary time series with engineering applications*. M.I.T. Press, Cambridge, MA, 163 pp.

[223] Wiener, N., 1956: Nonlinear prediction and dynamics. *3rd Berkeley Symposium Math. Stat. Probability,* Vol. **3**, University of California Press, 247-252.

[224] Wolfe, P., 1968: The secant method for simultaneous nonlinear equations. *Comm. Ass. Comput. Mach.*, **2**, 12-13.

[225] Zou, X., I. M. Navon, F. X. Le Dimet, A. Nouailler and T. Schlick, 1990: A comparison of efficient large-scale minimization algorithms for optimal control applications in meteorology. *Tech. Rep. FSU-SCRI-90-167,* Florida State University, Tallahassee, Florida, 44 pp.

[226] Zou, X. and I. M. Navon, 1991: On the effect of gravity oscillation, boundary control and incomplete observations in a shallow water equations model with application to variational data assimilation. *Conference on Numerical Weather Prediction, AMS, October 14-18, 1991. Denver, Colorado,* **13B.4**, 459-462.

[227] Zou, X., I. M. Navon, M. Berger, Paul K. H. Phua, T. Schlick and F. X. Le Dimet, 1992: Numerical experience with limited-memory quasi-Newton methods and truncated Newton methods. *SIAM Jour. on Numerical Optimization,* **3**, 582-608.

[228] Zou, X., I. M. Navon and F. X. Le Dimet, 1992: Incomplete observations and control of gravity waves in variational data assimilation. *Tellus*, **44A**, 273-296.

[229] Zou, X., I. M. Navon and F. X. Le Dimet, 1992: An optimal nudging data assimilation scheme using parameter estimation. *Q. J. R. Meteorol. Soc.,* **118**, 1163-1186.

[230] Zou, X.,I. M. Navon and J. Sela, 1993: Control of gravity oscillations in variational data assimilation. *Mon. Wea. Rev.*, **121**, 272-289.

[231] Zou, X., I. M. Navon and J. Sela, 1993: Variational data assimilation with threshold processes using the NMC multilevel primitive-equation model. *Tellus*, in press.

[232] Zou, X., I. M. Navon, A. Barcilon, J. Whittaker, and D. Cacuci, 1993: An adjoint sensitivity study of blocking in a two-layer isentropic model. *Mon. Wea. Rev.*, **121** (10), 2833-2857.

[233] Zou, X. and I. M. Navon, 1993: Variational data assimilation: Some aspects of theory and application. Environmental Modeling, Vol II, P. Zannetti (Editor), Failure Analysis Association Inc, California, USA, in press.

[234] Zou, J. P. and G. Holloway, 1993: Estimating model inputs from time mean sea surface height by a free surface model and its adjoint. *J. Phys. Oceanogr.,* in press.

[235] Zupanski, M., 1993: Regional four-dimensional variational data assimilation in a quasi-operational forecasting Environment. *Mon. Wea. Rev.*, **121**, 2396-2408.

[236] Zupanski, M., 1993: A new preconditioning for large scale minimization problems. *Tellus*, submitted.

[237] Zupanski, D., 1993: The effects of discontinuities in the Belts-Miller cumulus convection scheme on four-dimensional variational data assimilation. *Tellus*, in press.

BIOGRAPHICAL SKETCH

Zhi Wang

EDUCATION

5/91–11/93 Doctor of Philosophy in Applied Mathematics/Computer Science,
Florida State University, Tallahassee, FL 32306 USA.

1/88–4/91 Master of Science in Applied Mathematics,
Florida State University, Tallahassee, FL 32306 USA.

9/84–5/87 Master of Science in Computational Mathematics,
Harbin Institute of Technology, Harbin, China.

9/80–7/84 Bachelor of Science in Applied Mathematics,
Northeast Forestry University, Harbin, China.

PUBLICATIONS

1 Zhi Wang, 1993: Variational data assimilation with 2-D shallow water equations and 3-D FSU global spectral model. Ph.D. Thesis.

2 Zhi Wang, I. M. Navon, F. X. Le Dimet and X. Zou, 1992: The second order adjoint analysis: Theory and applications. *Meteorology and Atmospheric physics,* **50**, 3-20.

3 Zhi Wang, I. M. Navon and X. Zou, 1993: The adjoint truncated Newton algorithm for large-scale unconstrained optimization. *Submitted to* Optimization Methods and Software.

4 Zhi Wang, I. M. Navon and X. Zou, 1993: Variational data assimilation with FSU 3-D global spectral model. *In preparation.*

5 Zhi Wang and I. M. Navon, 1993: Variational nudging data assimilation with FSU global spectral model in its full physics operational form. *In preparation.*

6 Zhi Wang, 1987: Methods of solving the plasma equilibrium equations. Master's Degree Thesis.

NUMERICAL MODELS DEVELOPED

1 The second order adjoint model of a limited-area shallow water equations model.

2 The first order adjoint model of the FSU global multilevel spectral model.

PRESENTATIONS

Speaker at *International Workshop on Adjoint Applications in Dynamic Meteorology.* Asillomar, CA USA, 1992.

PROGRAMMING EXPERIENCE

Experienced programmer in C, C^{++} and FORTRAN. Familiarity with X Window, Pascal, BASIC and assembly. Five year experience in large programming projects in FORTRAN.

Extensive experience in numerical modeling and numerical optimization on main frame supercomputers. Familiarity with complex graphics visualization.

Experienced in using AIX, UNIX, VAX/VMS, MS-DOS, Sun-OS and Cray-YMP.

Experienced in use of typesetting tools TEX , LATEX , TROFF and document preparation environments.

TEACHING EXPERIENCE

8/90–Now *RESEARCH ASSISTANT*, Supercomputer Computations Research Institute, FSU. Working on numerical weather prediction, data assimilation and optimal control problems.

5/91–9/91 *LECTURE INSTRUCTOR*, Department of Mathematics, FSU. Responsible for lecturing, testing and grading in undergraduate Precalculus.