FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

ADAPTIVE OBSERVATIONS IN A 4D-VAR FRAMEWORK APPLIED TO THE NONLINEAR BURGERS EQUATION MODEL

By

MD. JAKIR HOSSEN

A Thesis submitted to the
Department of Scientific Computing
in partial fulfillment of the
requirements for the degree of
Master of Science

Degree Awarded:
Fall Semester, 2008

The members of the Committee approve the Thesis of Md. Jakir Hossen defended on October 31, 2008.

Ionel Michael Navon
Professor Directing Thesis

Janet Peterson
Committee Member

Gordon Erlebacher
Committee Member

Approved:

Max Gunzburger, Chair
Department of Scientific Computing

Joseph Travis, Dean
College of Arts and Sciences

The Office of Graduate Studies has verified and approved the above named committee members.

ACKNOWLEDGEMENTS

In the first place I would like to record my gratitude to Professor I.M. Navon for his supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences through out the work. Above all and the most needed, he provided me unflinching encouragement and support in various ways. His truly scientist intuition has made him as a constant oasis of ideas and passions in science, which exceptionally inspire and enrich my growth as a student, a researcher and an aspiring scientist. I am indebted to him more than he knows.

I would like to thank Dr. Dacian N. Daescu for providing valuable comments, remarks and insights into current topics in data assimilation of my research. His constructive comments helped me to understand the topic deeply and finished the thesis successfully.

Also, I would like to thank Professor Gordon Erlebacher for his help to understand the computational tools and techniques. His inspiration and cheerful comments helped me a lot for doing better in course work as well as research. I would like to give special thanks for having agreed to be on my committee.

I would also like to thank Professor Janet Peterson for having provided me an opportunity to be able to learn from her experiences through various courses I signed up with her, and also for having agreed to be on my committee. I am very much grateful to her for her support and advice in so many issues.

I would like to thank all my friends in FSU for their help and moral support. I would like to give special thanks to Evan and Srujan.

I really appreciate the help of all the staff of system administration during all of these years. Also I would like to acknowledge the NASA, which supported me during the course of this thesis work, via research grant No. NNG06GC67G.

Finally, I would like to thank my parents for their inseparable support and prayers. My Father, Md. Hasmat Ali, in the first place is the person who put the fundament my learning

character, showing me the joy of intellectual pursuit ever since I was a child. My Mother, Safarennesa, is the one who sincerely raised me with her caring and gently love.

I owe my loving thanks my wife Zobaida Khanam and my daughter Ebadis Saba. They have missed me a lot during my study in FSU. The inspiration and support of my wife helped me to join in FSU and carry out research work. Without her encouragement and understanding it would have been impossible for me to finish this work.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

In 4D-Var data assimilation for geophysical models, the goal is to reduce the lack of fit between model and observations (strong constraint approach assuming perfect model). In the last two decades four dimensional variational technique has been extensively used in the numerical weather prediction due to the fact that time distributed observations are assimilated to obtain a better initial condition thus leading to more accurate forecasts using the above 4D-Var approach. The use of large-scale unconstrained minimization routines to minimize a cost functional measuring lack of fit between observations and model forecast requires availability of the gradient of the cost functional with respect to the control variables. Nonlinear Burgers equation model is used as numerical forecast model. First order adjoint model can be used to find the gradient of the cost functional. The use of targeted observations supplementing routine observations contributes to the reduction of the forecast analysis error and can provide improved forecast of weather events of critical societal impact, for instance, hurricanes, tornadoes, sharp fronts etc. The optimal space and time locations of the adaptive observations can be determined by using a singular vector approach. In our work we use both adjoint sensitivity and sensitivity to observation approaches to identify the optimal space and time locations for targeted observations at future time aimed at providing an improved forecast. Both approaches are compared in this work and some conclusions are outlined.

CHAPTER 1

INTRODUCTION

The numerical modeling of geophysical flows has advanced tremendously due to the availability of supercomputing facilities and the development of network of remote or in-situ observations. The domain of the modeling has been extended to complex flow such as atmospheric flow and oceanic flow. Geophysical fluids, for instance, air, atmosphere, ocean, surface or underground water are governed by the general equations of fluid dynamics.

Numerical Weather Prediction (NWP) is based on the the integration of a dynamic system of partial differential equations modeling the behavior of the atmosphere. Therefore discrete initial conditions describing the state of the atmosphere have to be provided prior to the integration, since they, along with the model equations, control the evolution of the solution trajectory in space and time. To find better initial condition we use data assimilation techniques. Data assimilation is a methodology to estimate the state of a time-evolving complex system like the atmosphere as accurately as possible from observational data and known physical laws. The physical laws are primarily represented by a numerical model of the system, which provides a short range forecast as the first guess for estimation. There are many techniques used in data assimilation such as described in early reviews of LeDimet and Navon 1988 [1]; Ghil et al. 1981 [2]; Ghil and Malanotte-Rizzoli 1991 [3]. Recently, considerable attention has been focused on variational methods for four dimensional data assimilation (LeDimet and Talagrand 1986 [4]; Derber 1987 [5]; Lewis and Derber 1985 [6]; Talagrand and Courtier 1987 [7]).

The operational implementation of NWP was started by Charney et al. 1950 [8]. The procedure for preparing initial conditions from observational data was called objective analysis in contrast to subjective analysis, because it was conducted numerically by the aid of a computer. The numerical schemes of the first generation of objective analysis were

the function fitting (Panofsky 1949 [9]; Gilchrist and Cressman 1954 [10]) and the optimal interpolation (Gandin 1963 [11]). The Optimal Interpolation (OI) is derived in terms of probability and statistics, while variational methods are based on combining model dynamics with data with the relative weighting defined in an ad hoc manner. As it is illustrated in a review of analysis methods by Lorenc 1986 [12], variational and statistical analysis methods have a common basis and can be made equivalent by proper definition of weights.

Using a Bayesian approach, Lorenc 1986 [12] derived a general objective function that can be used as the starting point for both existing OI procedure and variational analysis schemes. The derived objective function or cost function is a combination of deviations of the desired analysis from a forecast and observations weighted by the inverse of the corresponding forecast and observation-error covariance matrices. The optimal interpolation is based on the statistical estimation theory in contrast to the other two methods and gained popularity in the 1970s with sophistication to multivariate three dimensional interpolation.

New advances in data assimilation methods emerged in the 1980s, when considerable attention began to be paid to the four dimensional variational data assimilation (4D-Var) and the Kalman filter. Those assimilation methods used a flow dependent background error covariance for estimating the atmospheric state and assimilated indirect observational data such as satellite radiances without transforming them into analysis variables. Those two points are major advantages over OI, in which a statistically estimated background error covariance is used and observational data that can be assimilated are limited to observations of analysis variables. Furthermore, 4D-Var and the Kalman filter are continuous data assimilation methods based on the statistical estimation theory.

The application of calculus of variations to meteorological analysis was first studied by Sasaki 1958 [13] and it was extended to include the temporal dimension by Sasaki 1969 [14], 1970 [15]. This extension may be regarded as a prototype of 4D-Var. The objective of variational data assimilation is to determine the optimal solution of an NWP model by fitting the model dynamics to data over an interval of time, where the optimality is measured by a cost function that expresses the degree of discrepancy between the model and observational data. A direct approach for finding the optimal solution, suggested by Hoffman 1986 [16] was found to be impractical for operational NWP models implemented on present day computers. The expense of the variational assimilation can be reduced by using the adjoint of the numerical model to calculate all of the components of the gradient of the

cost function with respect to the initial conditions by one integration. The adjoint model arises from the theory of optimization and optimal control of partial differential equations (Lions 1971 [17]; Glowinski 1984 [18]). Its theoretical aspects were presented by LeDimet and Talagrand 1986 [4]; Talagrand and Courtier 1987 [7] and LeDimet 1997 [19].

The first operational implementation of variational methods to NWP was realized in June 1991, when the National Meteorological Center (NMC) of the United States adopted three dimensional variational assimilation (3D-Var) for global analysis (Parrish and Derber 1992 [20]). Although 3D-Var does not include the temporal dimension, it has an advantage over OI. Results from 4D-Var experiments with large scale numerical model were published in the early 1990s (Thepaut et al. 1991 [21]; Navon et al. 1992 [22]; M. Zupanski 1993 [23]). Thepaut et al. 1993 [24] demonstrated the ability of 4D-Var to generate flow dependent and baroclinic structure functions in Meteorological Analysis.

The forecast impact of targeting is determined by the distribution and types of routine and targeted observations, the quality of the background or first guess, and the ability of the data assimilation procedure to combine information from the background and observations. To deploy targeted observations we need to define a target area. Typically, an objective procedure (often based on adjoint or ensemble techniques) is used a day or more in advance to identify a target region for the spatial observations. It can also be determined on the basis of high probability for a large or a fast growing initial condition error.

Several techniques have been put forward to identify optimal sites for additional observations. Adjoint based techniques such as sensitivity to initial conditions and singular vectors (SV) have been tested for such tasks by many groups of researchers. Daescu and Navon 2004 [25] proposed a new adjoint sensitivity approach where they considered the interaction between adaptive observations and routine observations. The SV approach provides a possibility of searching for directions in phase space where the errors in the initial condition will amplify rapidly. The specification of the initial and final norms plays a crucial role. In the European Center for Medium-Range Weather Forecasts (ECMWF) operational EPS, SVs are computed with the so called total energy norm at initial and final time. It can be shown that among simple norms, the total energy norm provides SVs which agree best with analysis error statistics (Palmer et al. 1998 [26]). Berkmeijer et al. 1998 [27], 1999 [28] have shown that the Hessian of the cost function in a variational data assimilation scheme can be used to compute SVs that incorporate an estimate of the full analysis error covariance

at initial time and total energy norm at final time. This type of singular vector is called Hessian singular vector. Ehrendorfer and Tribbia 1997 [29] state that such an approach to determine SVs provides an efficient way to describe the forecast error covariance matrix when only a limited number of linear integrations are possible. Though finding the Hessian matrix explicitly involves a computationally intensive effort, we can calculate Hessian vector product by using second order adjoint (see LeDimet et al. [30]). This also requires an efficient generalized eigenvalue problem solver to compute SVs.

Baker and Daley 2000 [31] have shown that the adjoint of a data assimilation system provides an efficient way to estimate the sensitivities of an analysis of forecast measure with respect to observations. The sensitivities may be computed with respect to any or all of the observations simultaneously based on a single execution of the adjoint system see Zou et al 1993 [32]; Cacuci 1981 [33] etc. By using sensitivity value one can identify location of the adaptive observations which will have an impact on forecast measure at a given future forecast model time. Sensitivity to observations techniques have also been used to diagnose the effectiveness of specific targeted observations (Doerenbecher and Bergot 2001 [34]) and assess the impact of satellite data (Fourrié et al. 2002 [35]). This technique of sensitivity to observations was first presented by LeDimet et al. 1997 [19] and then discussed for 3D-Var method by Baker and Daley 2000 [31]; Zhu et al 2008 [36]; Trémolet 2008 [37] and comprehensively extended to the 4D-Var method by Daescu 2008 [38].

The key to sensitivity to observation techniques is to compute the transpose of the gain matrix that determines the weights given to the observation-minus-background residuals either explicitly or through a sequence of available operators. Daescu 2008 [38] shows that even without knowing the full gain matrix it is possible to compute observation sensitivity vector. This requires the availability of the Hessian of the cost function. There are two approaches to calculate Hessian matrix. If the size of the model is small (for example, 10^2 or less) then the above-mentioned Hessian matrix can be calculated explicitly by using the second order adjoint. If the size is larger then the information regarding Hessian matrix can be obtained from the first order adjoint by using an approximate finite difference method (See Wang et al 1995 [39]).

1.1 Thesis Structure

This thesis is organized as follows: Chapter 2 provides a brief introduction to various variational data assimilation methods and their mathematical formulation. Chapter 3 provides the derivation of tangent linear and adjoint models and a brief description of two targeting methods:– adjoint sensitivity and observation sensitivity methods. The methods to find the optimal location of adaptive observations based on targeting methods are also discussed in this chapter. Chapter 4 details the model used in our experiment and provides details of the algorithm used for calculating observation sensitivity vector. Numerical tests are then carried out and their results are assessed and discussed. Finally conclusions and future work are presented in chapter 5.

CHAPTER 2

DISCUSSION OF VARIATIONAL METHOD

In the original variational method proposed by Sasaki 1958 [13], the equations governing the flow are considered as constraints. The optimal estimate of the analysis variables was obtained by minimizing the cost function defined as a weighted sum of squared differences between analysis variables and observational data while the model acts as a strong constraint. A major advantage of the variational data assimilation methods is that it is easy to take into account dynamical balanced and smoothness by adding appropriate constraint terms. Lorenc 1986 [12] and others noted that the variational data assimilation method is based on the Bayesian probability theory. The use of a quadratic cost function is justified when the probability density function (PDFs) of background and observational errors are Gaussian.

2.1 Description of the various terms of variational method

2.1.1 State vector

The independent variables that are needed to represent the atmospheric state of the model are collectively called the state vector, denoted by x. The state vector depends on time and the model solution. If we use the nonlinear Burgers equation model, the state vector would consist of only the model solution at each time step. If model is discretized with n grid points along the x-direction, the size of state vector will be $n \times 1$ at each time step. If we use the shallow water model, the state vector would consist of three vector variables: the components of wind along the longitude and latitudes direction u and v respectively, geopotential height ϕ. Usually atmospheric studies are conducted using a discrete representation of the spherical earth. If the model is discretized with n_x grid points along the longitude , n_y along the latitude, then the number of the variables will be equal to $n = 3 \times n_x \times n_y$. Therefore, the

size of the state vector x will be $n \times 1$ which is a column vector.

2.1.2 Model

A model describes the evolution of a system represented by a set of nonlinear partial differential equation. For a very succinct presentation of the model, let us consider the following form

$$\begin{aligned} \frac{\partial x}{\partial t} &= F(x, \alpha) \\ x(t_0) &= \beta \end{aligned} \tag{2.1}$$

where x is the state vector and α and β are the parameters.
A discretized representation of the above model used for evolving the state of geophysical system from time t_k to time t_{k+1} can be represented in the form

$$x_{k+1} = \mathbf{M}x_k + \eta_k \tag{2.2}$$

The variable x_k is the state vector, the variable η_k is the model error at time t_k. Due to our inability to solve the nonlinear PDEs evolved in physical system analytically, NWP models are usually obtained after discretization using numerical methods such as finite difference, finite volume, finite element or spectral methods of the full PDEs. They are assumed to govern the flow of the atmosphere. Thus these nonlinear models provide only an approximation of the true evolution of the atmosphere, since the true evolution of the system may differ from (2.2) by unknown random and systematic errors. The mean of the model error is denoted by $E\{\eta_k\}$ where $E\{.\}$ denotes the mathematical expectation operator and the covariance of the model error at t_k is denoted by $Q_k = E\{\eta_k \eta_k^T\}$. If the numerical grid comprises of N grid points then the state vector x and model error vector η have dimension N and the model error covariance is an $N \times N$ matrix. The diagonal of the matrix contains variances for each variable of the model and off-diagonal terms are covariances between each pair of variables of the model. The error covariance matrix is given below

$$Q = \begin{pmatrix} E\{\eta_1\eta_1\} & E\{\eta_1\eta_2\} & \dots & \dots & \dots & E\{\eta_1\eta_N\} \\ E\{\eta_2\eta_1\} & E\{\eta_2\eta_2\} & \dots & \dots & \dots & E\{\eta_2\eta_N\} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ E\{\eta_N\eta_1\} & E\{\eta_N\eta_2\} & \dots & \dots & \dots & E\{\eta_N\eta_N\} \end{pmatrix}$$

It is evident from (2.1) and (2.2) that NWP models solve an initial value problem, where given an estimate of the present state of the atmosphere, say $x(t_0)$, the model can be used

to forecast the state $x(t_i)$ for a future time t_i. The quality of the forecast depends on the model predictability.

2.1.3 Observations

Observations provide an important source of information of the system for the data assimilation. They are denoted by the vector y. There are many ways to collect the observational data of the atmospheric state. Weather stations use different types of devices, for instance, balloons, buoys, ships, air-crafts, radiosondes, rawinsondes etc to provide direct observations. They also use various satellites to obtain indirect observation which measure radiances, various types of imagery (visible, infrared and water vapor images) which do not explicitly enter into the state vector representation but are functionals of the state vector. In practice, the observations are inhomogeneous and non-uniformly distributed (in both space and time). So to compare observations with the state vector it is important to use a function from model state space to observation space called an Observation operator, denoted by H. In the case of direct observations, H is a linear mapping of observations collected at various irregularly spaced locations to a regularly spaced numerical model grid. For indirect observations, H is a complex operator, and usually leads to an inverse problem.

A primary source of error associated with the observations is instrumental error. Another source of error is through the numerical operations involved in the observation operator. The observations y at time t_k are defined as

$$y(t_k) = H_k[x^t(t_k)] + \epsilon_o(t_k) \tag{2.3}$$

where x^t is the true state of the atmosphere and ϵ_o represents errors in the observations. The observations are assumed to be uncorrelated with the model errors, i.e,

$$E\{\eta_k \epsilon_{ok}^T\} = E\{\epsilon_{ok}^T \eta_k\} = 0$$

and the observational error covariance is denoted is $\mathbf{R}_k = E\{\epsilon_{ok}\epsilon_{ok}^T\}$.

There are two types of observational data used in data assimilation. One is routine observations, a set of time distributed observations, provided by the conventional observing network in the analysis time interval $[t_0, t_N]$. Besides the routine observations, a number of n_k additional observational resources may be deployed at target instant $\tau_k,\ t_0 < \tau_k < t_N,\ k = 1, 2, .., I$. They are known as targeted observations or adaptive observations.

2.1.4 Background vector

Besides the observations, it is necessary to have a complete first guess estimate of the state of the atmosphere at all the grid points in order to generate the initial conditions for the forecasts. The first guess is known as background field, denoted by x_b. In true sense, it should be the best estimate of the state of the atmosphere prior to the use of the observations. Usually a short range forecast(typically 6 hours) is used to generate x_b. The errors in the background field ϵ_b are calculated by $\epsilon_b = x_b - x^t$ and the background error covariance matrix , denoted by $\mathbf{B}$ is given as $\mathbf{B} = E\{\epsilon_b \epsilon_b^T\}$. The background error covariance representation is crucial for a good forecast mainly in data sparse regions.

2.2 Brief description of variational data assimilation method

One of the first significant data assimilation (DA) algorithms that was based on the empirical approach is called the Successive Corrections Method (SCM), introduced by Bergthorsson and Döös 1955 [40] and Cressman 1959 [41]. SCM, an iterative procedure, estimates the value of a variable on a regular grid using the background value of the field as the initial value at the zeroth iteration. Nudging is another empirical method used for DA of small scale observations, and is not generally used for large scale assimilation. It is based on a simple idea of nudging, or dynamically relaxing the solution of the numerical model towards the observations (interpolated to the model grid) by adding a suitable forcing term to the governing differential equation. Both SCM and nudging methods use the availability of observations, background field and numerical model to estimate the state of the atmosphere. Though they use all the available information, they do not provide a guarantee of obtaining an optimal estimation. Furthermore, they do not take into account the errors associated with these fields (such as the observation, background and model error covariances). Optimal interpolation was one of the first operationally implemented methods which was formulated for optimally estimating the state of the atmosphere.

2.2.1 Optimal interpolation

For finding the optimum analysis of a field of model variables x_a, given a background field x_b available at grid points and observations y, the following equations need to be used.

$$x_a = x_b + W(y - H[x_b]) \tag{2.4}$$

where W is the weighting matrix and H is the observational operator used to interpolate the background field on the model grid to the observation locations. We seek W such that the estimate x_a^*, among all possible x_a is optimal according to some criteria. An obvious optimality criteria is that x_a^* should minimize the estimation error with respect to the true state of the atmosphere.

Let us define the errors and their corresponding covariance matrices for the analysis, background and observations respectively.

$$\begin{aligned} \epsilon_a = x_a - x^t \; ; \quad P_a = \mathbf{A} = E\{\epsilon_a \epsilon_a^T\} \\ \epsilon_b = x_b - x^t \; ; \quad P_b = \mathbf{B} = E\{\epsilon_b \epsilon_b^T\} \\ \epsilon_o = y - H[x^t] \; ; \quad P_o = \mathbf{R} = E\{\epsilon_o \epsilon_o^T\} \end{aligned} \tag{2.5}$$

The minimization of the analysis error covariance matrix P_a provides us the optimal x_a^*. By subtracting x^t from both sides of (2.4) we have

$$\begin{aligned} x_a - x^t &= x_b - x^t + W(y - H[x^t + x_b - x^t]) \\ \Rightarrow \epsilon_a &= \epsilon_b + W(y - H[x^t + \epsilon_b]) \\ \Rightarrow \epsilon_a &= \epsilon_b + W(y - H[x^t] - \mathbf{H}[\epsilon_b]) \\ \Rightarrow \epsilon_a &= \epsilon_b + W(\epsilon_o - \mathbf{H}[\epsilon_b]) \end{aligned} \tag{2.6}$$

where $\mathbf{H} = \frac{\partial H}{\partial x}$ is a linear operator.

Therefore the analysis error covariance matrix P_a can be derived by using (2.6)

$$\begin{aligned} P_a &= E\{\epsilon_a \epsilon_a^T\} \\ &= E\{[\epsilon_b + W(\epsilon_o - \mathbf{H}[\epsilon_b])][\epsilon_b + W(\epsilon_o - \mathbf{H}[\epsilon_b])]^T\} \\ &= E\{\epsilon_b \epsilon_b^T + \epsilon_b(\epsilon_o - \mathbf{H}[\epsilon_b])^T W^T \\ &+ W(\epsilon_o - \mathbf{H}[\epsilon_b])\epsilon_b^T + W(\epsilon_o - \mathbf{H}[\epsilon_b])(\epsilon_o - \mathbf{H}[\epsilon_b])^T W^T\} \\ &= \mathbf{B} - W\mathbf{H}\mathbf{B} - \mathbf{B}\mathbf{H}^T W^T + W\mathbf{R}W^T + W\mathbf{H}\mathbf{B}\mathbf{H}^T W^T \\ &= \mathbf{B} - W\mathbf{H}\mathbf{B} - \mathbf{B}\mathbf{H}^T W^T + W(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)W^T \end{aligned} \tag{2.7}$$

where we used $\mathbf{B} = E(\epsilon_b \epsilon_b^T)$, $\mathbf{R} = E(\epsilon_o \epsilon_o^T)$ and $E(\epsilon_b \epsilon_o^T) = E(\epsilon_o \epsilon_b^T) = 0$.
The optimal weight matrix W that minimizes the covariance matrix P_a can be obtained by setting $\frac{\partial P}{\partial W} = 0$. That is,

$$\frac{\partial P}{\partial W} = -2\mathbf{B}\mathbf{H}^T + 2W(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T) = 0 \, . \tag{2.8}$$

We assumed that $\mathbf{B}$ is symmetric matrix. Therefore, the optimal W is given by,

$$W = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \, . \tag{2.9}$$

Therefore the optimal estimate is

$$x_a^* = x_b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(y - H[x_b]) \tag{2.10}$$

Note that using (2.9) in equation (2.7) we find an expression for P_a in terms of W

$$P_a = \mathbf{B} - W\mathbf{H}\mathbf{B} = (I - W\mathbf{H})\mathbf{B} \tag{2.11}$$

For more details see E. Kalnay 2003 [42].

2.2.2 3D-Variational data assimilation method

The objective of variation data assimilation (VDA) methods is to find an estimate of the state that fits simultaneously the background field and the observations, given their respective degree of uncertainty, i.e, the inverse of their covariances, $\mathbf{B}^{-1}$ and $\mathbf{R}^{-1}$ respectively. Such an estimate is obtained efficiently by minimizing the least squares cost functional. The variational method gives an optimal estimate of the state variable x_a as the maximum a posteriori (MAP) estimate:

$$\begin{aligned} x_a &= \arg\max_{\mathrm{x}}[p(x|x_b, y)] \\ &= \arg\max_{\mathrm{x}}[p(x_b|x)p(y|x)p(x)] \\ &= \arg\max_{\mathrm{x}} J(x) \end{aligned} \tag{2.12}$$

where

$$J(x) = p(x_b|x)p(y|x)p(x) \tag{2.13}$$

Assume that probability density functions of background and observational errors are Gaussian distributed. Therefore, they can be written in the following form,

$$p(x_b|x) = \frac{1}{(2\pi)^{n/2}|\mathbf{B}|^{1/2}} exp[-\frac{1}{2}(x-x_b)^T\mathbf{B}^{-1}(x-x_b)] \tag{2.14}$$

$$p(y|x) = \frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}} exp[-\frac{1}{2}(y-H[x])^T\mathbf{R}^{-1}(y-H[x])] \tag{2.15}$$

By using the above formula we can write $J(x)$ as

$$\begin{aligned} J(x) &= \frac{1}{(2\pi)^n|\mathbf{B}|^{1/2}|\mathbf{R}|^{1/2}} exp[-\{\frac{1}{2}(x-x_b)^T\mathbf{B}^{-1}(x-x_b) \\ &+ \frac{1}{2}(y-H[x])^T\mathbf{R}^{-1}(y-H[x])\}]p(x) \end{aligned} \tag{2.16}$$

For x we assume that the priori information is unknown. For convenience, we assume that the probability density function is Gaussian distributed with mean μ and covariance P.

Therefore

$$p(x) = \frac{1}{(2\pi)^{n/2}|P|^{1/2}} exp[-\frac{1}{2}(x-\mu)^T P^{-1}(x-\mu)] \tag{2.17}$$

$$\frac{\partial \ln p(x)}{\partial x} = -P^{-1}(x-\mu) \tag{2.18}$$

The lack of information about x implies that infinite variance, $P \to \infty$, or $P^{-1} \to 0$. Thus, without a priori knowledge on x we have

$$\frac{\partial \ln p(x)}{\partial x} = 0 \tag{2.19}$$

Therefore, maximizing $J(x)$ is equivalent to maximizing its logarithm or minimizing its negative logarithm. Since $\frac{\partial \ln p(x)}{\partial x} = 0$, we can neglect $p(x)$ from the equation (2.16). We can rewrite the equation (2.12) as

$$x_a = \arg\min_{\mathrm{x}} \mathcal{J}(x) \tag{2.20}$$

where

$$\mathcal{J}(x) = \underbrace{\frac{1}{2}(x-x_b)^T\mathbf{B}^{-1}(x-x_b)}_{\mathcal{J}_b} + \underbrace{\frac{1}{2}(y-H[x])^T\mathbf{R}^{-1}(y-H[x])}_{\mathcal{J}_o} \tag{2.21}$$

In order to minimize the cost functional (2.21) we need to have its gradient and Hessian with respect to the state vector. At the optimal point x^* of $\mathcal{J}(x)$, gradient $\nabla_x\mathcal{J}(x^*) = 0$ and the Hessian $\frac{\partial^2 J}{\partial x^2}|_{x^*}$ is positive definite.

The gradient $\nabla_x \mathcal{J}(x)$ of the cost function $\mathcal{J}(x)$ can be derived in the following way.

$$\nabla_x \mathcal{J}(x) = \mathbf{B}^{-1}(x - x_b) - \mathbf{H}^T\mathbf{R}^{-1}(y - H[x]) \tag{2.22}$$

At the optimal point x^* the gradient $\nabla_x \mathcal{J}(x)$ is zero , i.e.,

$$\begin{aligned}
&\nabla_x \mathcal{J}(x^*) = 0 \\
\Rightarrow &\mathbf{B}^{-1}(x^* - x_b) - \mathbf{H}^T\mathbf{R}^{-1}(y - H[x^*]) = 0 \\
\Rightarrow &\mathbf{B}^{-1}(x^* - x_b) - \mathbf{H}^T\mathbf{R}^{-1}(y - H[x^* - x_b + x_b]) = 0 \\
\Rightarrow &\mathbf{B}^{-1}(x^* - x_b) - \mathbf{H}^T\mathbf{R}^{-1}(y - H[x_b] - \mathbf{H}(x^* - x_b)) = 0 \\
\Rightarrow &\mathbf{B}^{-1}(x^* - x_b) - \mathbf{H}^T\mathbf{R}^{-1}(y - H[x_b]) + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}(x^* - x_b) = 0 \\
\Rightarrow &(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})(x^* - x_b) = \mathbf{H}^T\mathbf{R}^{-1}(y - H[x_b]) \\
\Rightarrow &x^* - x_b = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(y - H[x_b]) \\
\Rightarrow &x^* = x_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(y - H[x_b])
\end{aligned}$$

Therefore the optimal solution can be obtained from the formula

$$x^* = x_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(y - H[x_b]) \tag{2.23}$$

By using the formula (2.22) we can find the expression for the Hessian.

$$\nabla^2_{xx} \mathcal{J}(x^*) = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \tag{2.24}$$

Since $\mathbf{B}$ and $\mathbf{R}$ are symmetric positive definite, the Hessian evaluated at x^* is positive definite, which implies that x^* furnishes a minimum of the cost functional.

Theorem 2.2.1. *The Hessian of the cost function of the variational analysis is equal to the inverse of the analysis error covariance matrix*

$$\mathbf{A} = (J'')^{-1} \tag{2.25}$$

Proof. The Hessian of the cost function given in (2.21) is obtained by differentiating twice with respect to the control variable x.

$$\begin{aligned}
\mathcal{J}(x) &= \frac{1}{2}(x - x_b)^T\mathbf{B}^{-1}(x - x_b) + \frac{1}{2}(y - H[x])^T\mathbf{R}^{-1}(y - H[x]) \\
\nabla_x \mathcal{J}(x) &= \mathbf{B}^{-1}(x - x_b) - \mathbf{H}^T\mathbf{R}^{-1}(y - H[x]) \\
J'' &= \nabla^2_{xx} \mathcal{J}(x) = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}
\end{aligned} \tag{2.26}$$

By setting, $\nabla_x \mathcal{J}(x_a) = 0$ we have

$$\begin{aligned}
&\mathbf{B}^{-1}(x_a - x_b) - \mathbf{H}^T\mathbf{R}^{-1}(y - H[x_a]) = 0 \\
&\mathbf{B}^{-1}(x_a - x^t - (x_b - x^t)) - \mathbf{H}^T\mathbf{R}^{-1}(y - H[x^t + x_a - x^t]) = 0 \\
&\mathbf{B}^{-1}(\epsilon_a - \epsilon_b) = \mathbf{H}^T\mathbf{R}^{-1}(y - H[x^t] - \mathbf{H}[x_a - x^t]) = 0 \\
&\mathbf{B}^{-1}\epsilon_a - \mathbf{B}^{-1}\epsilon_b = \mathbf{H}^T\mathbf{R}^{-1}(\epsilon_o - \mathbf{H}[\epsilon_a])) \\
&(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\epsilon_a = \mathbf{B}^{-1}\epsilon_b + \mathbf{H}^T\mathbf{R}^{-1}\epsilon_o
\end{aligned}$$

Now,

$$\begin{aligned}
(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\epsilon_a\epsilon_a^T(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) &= (\mathbf{B}^{-1}\epsilon_b + \mathbf{H}^T\mathbf{R}^{-1}\epsilon_o)(\mathbf{B}^{-1}\epsilon_b + \mathbf{H}^T\mathbf{R}^{-1}\epsilon_o)^T \\
&= \mathbf{B}^{-1}\epsilon_b\epsilon_b^T\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\epsilon_o\epsilon_b^T\mathbf{B}^{-1} \\
&+ \mathbf{B}^{-1}\epsilon_b\epsilon_o^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{H}^T\mathbf{R}^{-1}\epsilon_o\epsilon_o^T\mathbf{R}^{-1}\mathbf{H}
\end{aligned}$$

By taking expectation of the result we have,

$$\begin{aligned}
(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})E(\epsilon_a\epsilon_a^T)(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) &= \mathbf{B}^{-1}E(\epsilon_b\epsilon_b^T)\mathbf{B}^{-1} \\
&+ \mathbf{H}^T\mathbf{R}^{-1}E(\epsilon_o\epsilon_b^T)\mathbf{B}^{-1} \\
&+ \mathbf{B}^{-1}E(\epsilon_b\epsilon_o^T)\mathbf{R}^{-1}\mathbf{H} \\
&+ \mathbf{H}^T\mathbf{R}^{-1}E(\epsilon_o\epsilon_o^T)\mathbf{R}^{-1}\mathbf{H} \\
(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{A}(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) &= \mathbf{B}^{-1}\mathbf{B}\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{R}\mathbf{R}^{-1}\mathbf{H} \\
(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{A}(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) &= \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \qquad (2.27)
\end{aligned}$$

Where we used $\mathbf{A} = E(\epsilon_a\epsilon_a^T)$, $\mathbf{B} = E(\epsilon_b\epsilon_b^T)$, $\mathbf{R} = E(\epsilon_o\epsilon_o^T)$
and $E(\epsilon_b\epsilon_o^T) = E(\epsilon_o\epsilon_b^T) = 0$ since we assume background and observation errors are uncorrelated.

Therefore,

$$A = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1} \qquad (2.28)$$

From the equation (2.26) and (2.28) we can write

$$\mathbf{A} = (J'')^{-1} \qquad (2.29)$$

□ □

See details in F.Bouttier and P.Courtier 1999 [43]

2.2.3 Physical space analysis system (PSAS)

The PSAS is another Variational Data Assimilation method which is introduced by Da Silva et. al 1995 [44]. The optimal estimate of the state is obtained by minimizing a cost functional which is defined in the physical space of the observations. If the size of observation vector y is much smaller than that of the state vector x, then the dimension in which the minimization is carried out to solve the PSAS problem is significantly smaller than that for the 3D problem. Let δx denote the increment to the background field x_b to obtain the optimal estimate of the state vector x such that

$$x = x_b + \delta x \tag{2.30}$$

where δx is obtained by solving the following equation

$$\delta x = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\delta y \tag{2.31}$$

and let δy denote the residual in the observational space, defined by $\delta y = y - H[x_b]$. The equation (2.31) can be solved by splitting into two steps:

- First step: $w = (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\delta y$
- Second step: $\delta x = \mathbf{B}\mathbf{H}^T w$

The first step requires calculating the inverse of a matrix that is the most computer intensive. To avoid this, it can be solved by minimizing the cost function

$$J(w) = \frac{1}{2}w^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)w - w^T\delta y \tag{2.32}$$

This is an efficient method if the dimension of the observational space is much smaller than that of the model space.

2.2.4 4D-Variational data assimilation

Four dimensional data assimilation(4D-Var) seeks the MAP estimate for the state variables of a numerical model during a certain period that is called an assimilation window. Although the formulation of 4D-Var is obtained from a simple extension of 3D-Var, there is a distinguishing advantage of 4D-Var that the full model dynamics is used for analysis. In four dimensional variational assimilation, an initial condition is sought so that the forecast best

fits the observations within an assimilation interval. The method requires a time dependent nonlinear model to obtain the optimal estimate $x_0^a \in \mathbb{R}^n$ of the initial condition by minimizing the cost function defined as

$$\mathcal{J}(x_0) = \underbrace{\frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b)}_{\mathcal{J}_b} + \underbrace{\frac{1}{2}\sum_{i=0}^{N}(y_i - H_i[x_i])^T \mathbf{R}_i^{-1}(y_i - H_i[x_i])}_{\mathcal{J}_o} \tag{2.33}$$

where x_0 denotes the initial state at the beginning of the assimilation window, x_b is a prior(background) estimate to the initial state, $y_i \in \mathbb{R}^{k_i}$, $i = 0, 1, 2, .., N$ is the set of observations made at time t_i and $x_i = \mathbf{M}(t_0, t_i)(x_0)$. The model $\mathbf{M}$ is nonlinear and we assume that the model is perfect. $H_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ is the observation operator that maps the state space into the observation space at time t_i. $\mathbf{B}$ is the background error covariance matrix and $\mathbf{R}_i$ is the observational error covariance matrix at time t_i. We assume that background errors and observation errors are uncorrelated with each other. In our case, we take the error covariance matrices $\mathbf{B}$ and $\mathbf{R}_i$ to be diagonal. The control variable or the variable with respect to which the cost function (2.33) is minimized is the initial state of the model $x(t_0)$ within the assimilation window.

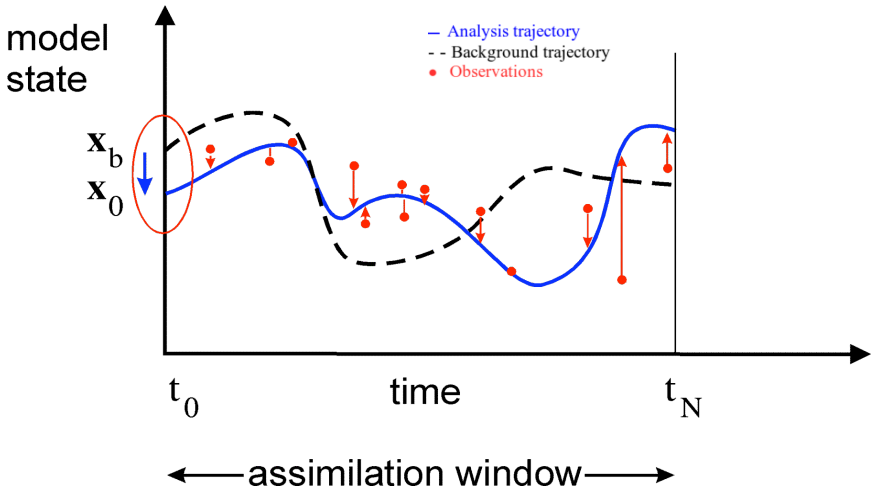

Figure 2.1: Schematic representation of 4D-Var. Using information from the previous forecast (x_b) and available observations within a time interval, a new optimal x_0 is obtained via 4D-Var so that the corrected forecast (analysis trajectory) closely fits the observations.

Note that the second component of the cost functional, $\mathcal{J}_o$, describes the sum of least squared differences between each of the observations and the output from the forecast model over the entire assimilation window. The first term $\mathcal{J}_b$, measured the least squares distance between the background field and model initial state $x(t_0)$ remains the same as the 3D-Var cost functional.

In order to minimize the cost functional in (2.33) with respect to $x_0 = x(t_0)$, we need to calculate the gradient of the cost functional with respect to the control variable i.e. $\nabla_{x_0}\mathcal{J}$ and ensure that the Hessian of the cost functional evaluated at the minimum is positive definite. Since the cost functional is convex by construction and the background and observation error covariance, $\mathbf{B}$ and $\mathbf{R}_i$ are symmetric positive definite(by definition) the Hessian of the cost functional is positive definite.

CHAPTER 3

TARGETING METHODS FOR ADAPTIVE OBSERVATIONS

3.1 The adjoint sensitivity (AS) approach

The minimum point of the cost function is numerically obtained using large scale unconstrained minimization algorithms such as quasi-Newton methods, conjugate gradient method etc. Those methods are iterative methods and require the gradient of the cost function as well as its value at each iteration. The adjoint method provides an efficient algorithm to calculate the gradient of the cost function with respect to control variables. Let us consider a first order variation of the cost functional $\mathcal{J}$ by introducing a small change $\delta x(t_0)$ in the control variable $x(t_0)$

$$\begin{aligned} \delta\mathcal{J} &= \mathcal{J}[x(t_0)+\delta x(t_0)] - \mathcal{J}[x(t_0)] \\ &\approx \frac{\partial \mathcal{J}}{\partial x(t_0)}\delta x(t_0) \\ &= [\nabla_{x(t_0)}\mathcal{J}]^T \delta x(t_0) \end{aligned} \tag{3.1}$$

where $[\nabla_{x(t_0)}\mathcal{J}]^T\delta x(t_0)$ is the directional derivative which is equivalent to Gateaux derivative in finite dimension. The gradient of background component $\mathcal{J}_b$ of the cost function in (2.33) with respect to $x(t_0)$ is given by

$$\nabla_{x_0}\mathcal{J}_b = \mathbf{B}^{-1}(x_0 - x_b) \tag{3.2}$$

The gradient of observation term $\mathcal{J}_o$ of the cost function in (2.33) with respect to $x(t_0)$ is given by

$$\nabla_{x_0}\mathcal{J}_o = \sum_{i=0}^{N}\left[\frac{\partial(y_i - H_i[x_i])}{\partial x_0}\right]^T \mathbf{R}_i^{-1}(y_i - H_i[x_i]) \tag{3.3}$$

Here $y_i = y(t_i)$ is the time distributed observations, $x_0 = x(t_0)$, model solution at each t_i, $x_i = x(t_i) = \mathbf{M}_{0,i}(x_0)$, $\mathbf{M}_{0,i} = \mathbf{M}(t_0, t_i)$. Let us consider a small perturbation δx_0 to the initial state of the model, then we have

$$\bar{x}_i = \mathbf{M}_{0,i}(x_0 + \delta x_0) \tag{3.4}$$

Expanding the right hand side by Taylor series we have

$$\begin{aligned} \bar{x}_i &= \mathbf{M}_{0,i}(x_0) + \frac{\partial \mathbf{M}_{0,i}}{\partial x_0}\delta x_0 + O(\delta x_0 \delta x_0^T) \\ &= \mathbf{M}_{0,i}(x_0) + \mathcal{M}(t_0, t_i)\delta x_0 + O(\delta x_0 \delta x_0^T) \end{aligned} \tag{3.5}$$

Here $\mathcal{M}(t_0, t_i) = \frac{\partial \mathbf{M}_{0,i}}{\partial x_0}$

$$\delta x_i = \bar{x}_i - x_i \approx \mathcal{M}(t_0, t_i)\delta x_0 \tag{3.6}$$

By using the chain rule in calculus we can write

$$\begin{aligned} \frac{\partial (y_i - H_i[x_i])}{\partial x_0} &= -\frac{\partial H_i}{\partial x_i}\frac{\partial x_i}{\partial x_0} \\ &= -\mathbf{H}_i \mathcal{M}(t_0, t_i) \end{aligned} \tag{3.7}$$

where

$$\begin{aligned} \mathcal{M}(t_0, t_i) &= \mathcal{M}(t_0, t_1)\mathcal{M}(t_1, t_2).....\mathcal{M}(t_{i-1}, t_i) \\ &= \prod_{j=0}^{j=i-1} \mathcal{M}(t_j, t_{j+1}) \end{aligned} \tag{3.8}$$

Therefore, by using the above results, the gradient of the observational cost in (3.3) can be written as

$$\begin{aligned} \nabla_{x_0}\mathcal{J}_o &= \sum_{i=0}^{N} \left[\frac{\partial (y_i - H_i[x_i])}{\partial x_0}\right]^T \mathbf{R}_i^{-1}(y_i - H_i[x_i]) \\ &= -\sum_{i=0}^{N} [\mathbf{H}_i \mathcal{M}(t_0, t_i)]^T \mathbf{R}_i^{-1}(y_i - H_i[x_i]) \\ &= -\sum_{i=0}^{N} [\mathcal{M}(t_0, t_i)]^T \mathbf{H}_i^T \mathbf{R}_i^{-1}(y_i - H_i[x_i]) \\ &= -\sum_{i=0}^{N} \mathcal{M}^T(t_0, t_i) \mathbf{H}_i^T \mathbf{R}_i^{-1}(y_i - H_i[x_i]) \end{aligned} \tag{3.9}$$

where $\mathcal{M}^T(t_0,t_i)=\mathcal{M}^T(t_{i-1},t_i)\mathcal{M}^T(t_{i-2},t_{i-1})\ldots\ldots\mathcal{M}^T(t_0,t_1)$. The linear operator $\mathcal{M}(t_0,t_i)$ in (3.5) is called Tangent Linear Model (TLM) and its transpose $\mathcal{M}^T(t_0,t_i)$ in equation (3.9) is called adjoint model (ADM) and it is computationally obtained by transposing each and every linear statement in the TLM, integrated backward in time. The full gradient of the cost functional (2.33) is

$$\begin{aligned}\nabla_{x_0}\mathcal{J} &= \nabla_{x_0}\mathcal{J}_b+\nabla_{x_0}\mathcal{J}_o \\ &= \mathbf{B}^{-1}(x_0-x_b)-\sum_{i=0}^{N}\mathcal{M}^T(t_0,t_i)\mathbf{H}_i^T\mathbf{R}_i^{-1}(y_i-H_i[x_i]) \\ &= \mathbf{B}^{-1}(x_0-x_b)-\sum_{i=0}^{N}\mathcal{M}_{0,i}^T\mathbf{H}_i^T\mathbf{R}_i^{-1}(y_i-H_i[x_i]) \qquad (3.10)\end{aligned}$$

where

$$\mathcal{M}_{0,i}=\mathcal{M}(t_0,t_i)$$

The equation (3.10) shows that every iteration of the 4D-Var minimization requires the computation of the gradient i.e. computing the increment $y_i-H_i[x_i]$ at the observation time t_i during a forward integration, multiplying them by $\mathbf{H}_i^T\mathbf{R}_i^{-1}$ and integrating these weighted increments back to the initial time using ADM. Therefore implementation of the 4D-Var algorithm involves a sequence of calls to a minimization algorithm, which uses the functional and gradient information in an unconstrained minimization iterative mechanism, which yields the optimum value of initial condition $x^*=x_0$ thus providing the minimum value of the cost function $\mathcal{J}(x^*)$.

Usually, the forecast model is nonlinear. So the development of the tangent linear and adjoint models is extremely complicated and time consuming. To make the task easier and save the time several automatic differentiation software packages have been developed, for instance, TAMC, ADIFOR etc. that provide forward and reverse sweeps.

3.1.1 Location of adaptive observation by AS

The first approach to identify the adaptive observations is adjoint sensitivity method. In practice it is of interest to assess the observation impact on the forecast measure $\mathcal{J}^v(x_v)$ on verification domain at verification time t_v. The functional $\mathcal{J}^v$ is defined (see Daescu and Navon 2004 [25]) as a scalar measure of the forecast error over the verification domain $\mathcal{D}_v$

$$\mathcal{J}^v=\frac{1}{2}(x_v^f-x_v^t)^TP^TEP(x_v^f-x_v^t) \qquad (3.11)$$

where $x_v^f = \mathbf{M}_{0,v}(x_0^a)$ and $x_v^t = \mathbf{M}_{0,v}(x_0^t)$. P is a projection operator on $\mathcal{D}_v$ satisfying $P^*P = P^2 = P$ and E is a diagonal matrix of the total energy norm.

To select the adaptive observations locations, the gradient of cost functional $\mathcal{J}^v$ defined in equation (3.11) is used. The gradient of the function (3.11) at t_i is defined as

$$\nabla \mathcal{J}^v(x_i) = \mathcal{M}_{i,v}^T P^T E P(x_v^f - x_v^t) \tag{3.12}$$

where

$$x_i = x(t_i)$$

We use gradient of the function defined in (3.12) to evaluate the sensitivity of the forecast error with respect to the model state at each targeting instant τ_k. A large sensitivity value indicates that small variations in the model state will have a significant impact on the forecast at the verification time. The adjoint sensitivity field with respect to total energy metric is defined as

$$F_v(\tau_k) = ||\nabla \mathcal{J}^v(x(\tau_k))||_E \in \mathbb{R}^n \tag{3.13}$$

where total energy metric is defined as

$$||x||_E = \frac{1}{2}x^2 \tag{3.14}$$

Adaptive observation at τ_k are deployed at the first n_k locations x_k where $F_v(\tau_k)$ attains maximum values.

3.2 Observation sensitivity (OS) approach

The Hessian of the cost function $\mathcal{J}$ is obtained by differentiating equation (3.10) with respect to x_0^a

$$\nabla^2_{x_0^a x_0^a} \mathcal{J} = \mathbf{B}^{-1} + \sum_{i=0}^{N} \mathcal{M}_{0,i}^T \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i \mathcal{M}_{0,i} \tag{3.15}$$

By differentiating (3.10) with respect to y_i

$$\nabla^2_{y_i x_0^a} \mathcal{J} = -\mathbf{R}_i^{-1} \mathbf{H}_i \mathcal{M}_{0,i} \tag{3.16}$$

The theory discussed in Daescu 2008 [38] shows that if the cost function $\mathcal{J}(x, u)$ is a twice continuously differentiable function involving parameter vector (input data) $u \in \mathbb{R}^p$, then the optimal solution $\bar{x}$ corresponding to the parameter $\bar{u}$ that minimizes $\mathcal{J}$ is obtained by

satisfying the condition $\nabla_x \mathcal{J}(\bar{x}, \bar{u}) = 0$ and $\nabla^2_{xx} \mathcal{J}(\bar{x}, \bar{u})$ is positive definite. The implicit function theorem applied to the first order optimality condition

$$\nabla_x \mathcal{J}(\bar{x}, \bar{u}) = 0 \in \mathbb{R}^n$$

guarantees the existence of a vicinity of $\bar{u}$ where the optimal solution is a function of data $x = x(u)$ and the gradient matrix

$$\nabla_u x = (\nabla_u x_1, \nabla_u x_2, \dots \nabla_u x_n) \in \mathbb{R}^{p \times n} \tag{3.17}$$

is expressed as

$$\nabla_u x(u) = -\nabla^2_{ux} \mathcal{J}[x(u), x] \{\nabla^2_{xx} \mathcal{J}[x(u), x]\}^{-1} \tag{3.18}$$

By theorem 2.2.1 we can write

$$\mathbf{A} \approx \{\nabla^2_{x_0^a x_0^a} \mathcal{J}\}^{-1} \in \mathbb{R}^{n \times n} \tag{3.19}$$

So from the equations (3.16) , (3.18), (3.19) we obtain

$$\begin{aligned} \nabla_{y_i} x_0^a &= -\nabla^2_{y_i x_0^a} \mathcal{J} \mathbf{A} \\ &= \mathbf{R}_i^{-1} \mathbf{H}_i \mathcal{M}_{0,i} \mathbf{A} \in \mathbb{R}^{k_i \times n} \end{aligned} \tag{3.20}$$

The gradient of $\mathcal{J}^v$ defined in (3.11) at verification time t_v is

$$\nabla_{x_v} \mathcal{J}^v = P^T E P (x_v^f - x_v^t) \tag{3.21}$$

Using the chain rule we obtain

$$\nabla_{y_i} \mathcal{J}^v = \nabla_{y_i} x_0^a \nabla_{x_0^a} \mathcal{J}^v \in \mathbb{R}^{k_i} \tag{3.22}$$

where $\nabla_{x_0^a} \mathcal{J}^v$ is the gradient of $\mathcal{J}^v$ at initial time obtained by using adjoint model. So, the equation for $\nabla_{x_0^a} \mathcal{J}^v$ can be written as

$$\nabla_{x_0^a} \mathcal{J}^v = \mathcal{M}^T_{0,v} \nabla_{x_v} \mathcal{J}^v \in \mathbb{R}^n \tag{3.23}$$

$\nabla_{x_0^a} \mathcal{J}^v$ represents forecast sensitivity to analysis. By using (3.20), equation (3.22) can be written as

$$\nabla_{y_i} \mathcal{J}^v = \mathbf{R}_i^{-1} \mathbf{H}_i \mathcal{M}_{0,i} \mathbf{A} \nabla_{x_0^a} \mathcal{J}^v \in \mathbb{R}^{k_i} \tag{3.24}$$

which provides the forecast sensitivity to the observations.

3.2.1 Location of adaptive observation by OS

The data assimilation system allows fusing model forecast with distributed observations. There are several sophisticated data assimilation methods that have already been developed and implemented. Amongst them, 4D-Var is considered as one of the best data assimilation methods since it uses concepts of optimal control of partial differential equations. If routine observations for different time steps are used then the observation sensitivity vector at those time steps can be calculated. The observation sensitivity vectors can then be used to determine the location for additional observations. Since it is costly to take additional observations at all the time steps where routine observations are available, it is important to determine time and space locations at which to take adaptive observations that will reduce forecast error significantly. To do that we first calculate L_∞ norm of the sensitivity vector $\nabla_{y_i}\mathcal{J}^v$ at each time step where observations are available.

$$F_v(t_i) = ||\nabla_{y_i}\mathcal{J}^v||_\infty \in \mathbb{R} \tag{3.25}$$

We then choose the expected number of time steps $\tau_k, k = 1, .., I$ where $F_v(\tau_k)$ attains the maximum value.

We define the observation sensitivity field with respect to total energy metric as

$$O_v(\tau_k) = ||\nabla_{y_{\tau_k}}\mathcal{J}^v||_E \in \mathbb{R}^{k_i} \tag{3.26}$$

We then choose n_k adaptive observations at each target instant $\tau_k,\ k = 1, 2, ..., I$ where $O_v(\tau_k)$ attains the maximum value.

CHAPTER 4

DESIGN OF NUMERICAL EXPERIMENTS

The four dimensional data assimilation method requires a significant computational effort to evaluate the forecast sensitivity to observations where a large scale model is used. This task is a little easier if a simple nonlinear model is used such as the Burgers equation model to be described below. The algorithm to compute the observation sensitivity are given below:–

- Calculate model solution x_v^t at t_v with initial condition x_0^t by

$$x_v^t = \mathbf{M}_{0,v}(x_0^t) \tag{4.1}$$

- Obtain optimal initial condition x_0^a by minimizing the cost functional $\mathcal{J}$ defined in (2.33). Calculate model forecast

$$x_v^f = \mathbf{M}_{0,v}(x_0^a) \tag{4.2}$$

- Compute $\nabla_{x_v}\mathcal{J}^v = P^T EP(x_v^t - x_v^f)$ and use it as initial condition for adjoint model.
- Integrate Adjoint model backward from t_0 to t_v : $\nabla_{x_0^a}\mathcal{J}^v = \mathcal{M}_{0,v}^T \nabla_{x_v}\mathcal{J}^v$
- Solve the linear system for z_0 : $\mathbf{A}^{-1}z_0 = \nabla_{x_0^a}\mathcal{J}^v$
  $\mathbf{A}^{-1}$ is calculated by using second order adjoint model.
- Integrate tangent linear model t_0 to t_i: $z_i = \mathcal{M}_{0,i}z_0$
- Mapping on observation space, weighting: $\nabla_{y_i}\mathcal{J}^v = \mathbf{R}_i^{-1}\mathbf{H}_i z_i$

4.1 Experimental setup

Numerical Experiments are set up with a one-dimensional nonlinear Burgers equation (Burgers 1948). The equation is given below

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}\,, \qquad \begin{matrix} -3 \le x \le 3 \\ 0 < t \le T \end{matrix} \tag{4.3}$$

with the initial conditions

$$u(x,0) = \begin{cases} 1.0 & -3 \le x \le 0 \\ 0.0 & 0 < x \le 3 \end{cases}$$

and boundary condition is

$$u(-3.0,t) = 1.0 \quad \& \quad u(3.0,t) = 0.0$$

For this combination of initial and boundary conditions (4.3) has an exact solution derived by using Cole-Hopf transformation (see J.D. Cole 1951 [45], E. Hopf 1950 [46]). The exact solution is given by

$$u = \frac{\int_{-\infty}^{\infty}[\frac{(x-\xi)}{t}]e^{-0.5RG}d\xi}{\int_{-\infty}^{\infty}e^{-0.5RG}d\xi} \tag{4.4}$$

where

$$G(\xi;x,t) = \int_0^{\xi} u_0(\xi')d\xi' + \frac{(x-\xi)^2}{2t}$$

and

$$R = \frac{1}{\nu}$$

To discretize the equation we used forward in time and centered in space (FTCS) finite-difference scheme. The numerical grid comprises of 101 mesh points in space and 300 steps in time, say, $n_x = 101$ and $n_t = 300$. The time interval is $\Delta t = 0.01$. The Reynolds numbers that we used are $R = 100$, $R = 200$ and $R = 300$.

The resulting discrete equation is

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{4\Delta x}\{(u_{j+1}^n)^2 - (u_{j-1}^n)^2\} + \frac{\Delta t}{R\Delta x^2}\{u_{j+1}^n - 2u_j^n + u_{j-1}^n\} \tag{4.5}$$

where Δx is space interval and the solution at each time step is $u_j \in \mathbb{R}^{n_x}$.

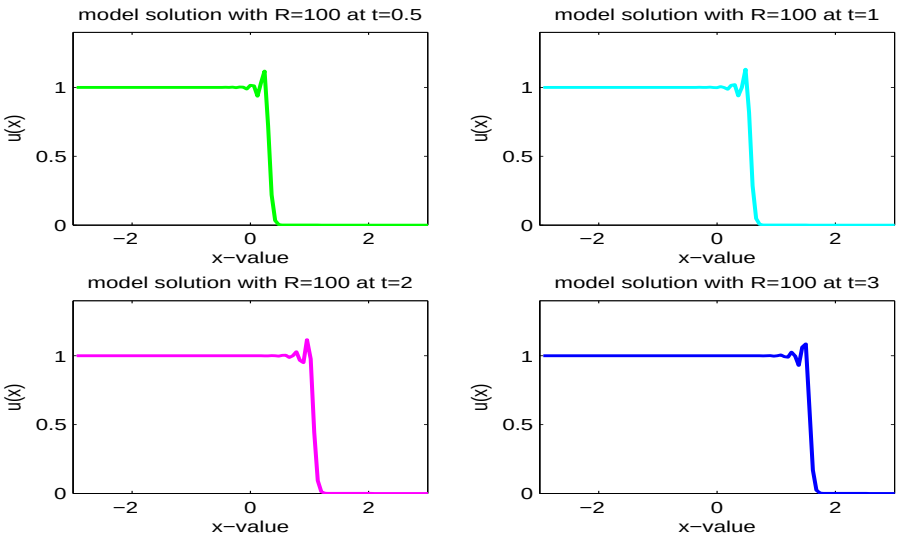

Figure 4.1: Burgers model solution for $R = 100$.

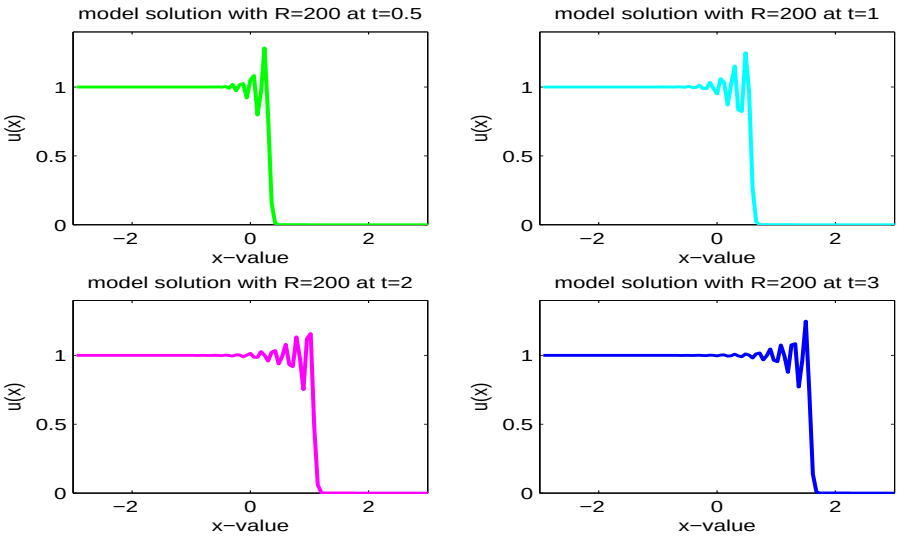

Figure 4.2: Burgers model solution for $R = 200$.

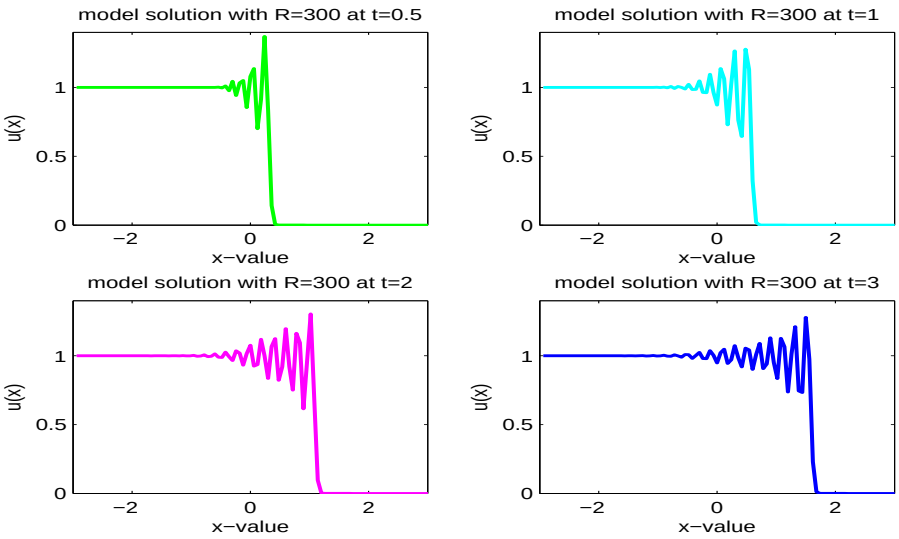

Figure 4.3: Burgers model solution for $R = 300$.

We displayed the model solutions for $R = 100$, $R = 200$ and $R = 300$ in Figures 4.1, 4.2 and 4.3 only at several time steps. We see that the model solution exhibits some oscillations for all cases. The unphysical oscillation originates due to high Reynolds numbers. We experienced that when the Reynolds numbers get higher, the numerical solution exhibits more oscillations.

We define our verification domain so that it contains the oscillations in numerical solution with larger amplitude at the verification time t_v. We obtain that such oscillations are contained roughly in the interval $[1.1, 1.6]$ at verification time $t_v = 3$. So the spatial domain $\mathcal{D}_v = [1.1, 1.6]$ is our verification domain. In Figures 4.4, 4.5 and 4.6 we display the verification domain by drawing a rectangular box.

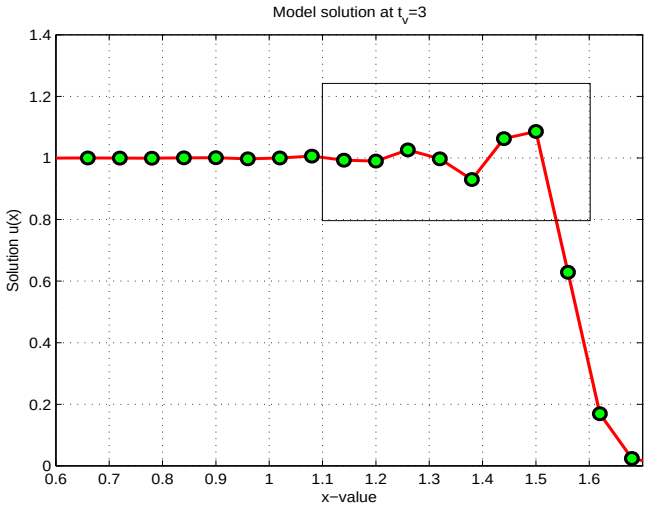

Figure 4.4: Verification domain which contains oscillations with larger amplitude at verification time t_v is shown in rectangular box. Figure displays the model solution for $R = 100$.

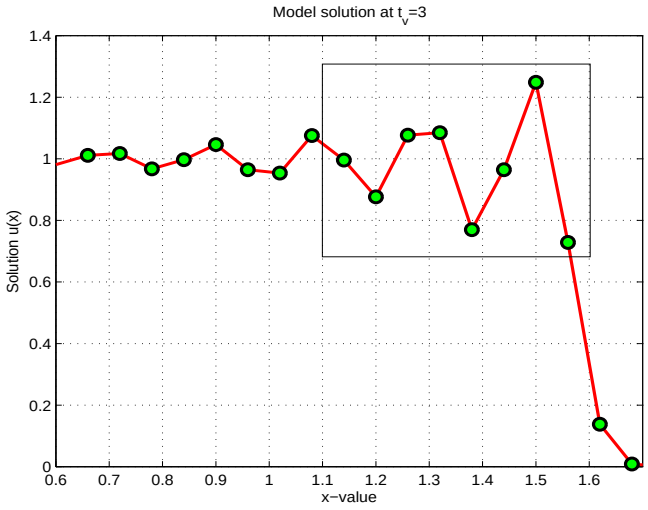

Figure 4.5: Verification domain which contains oscillations with larger amplitude at verification time t_v is shown in rectangular box. Figure displays the model solution for $R = 200$.

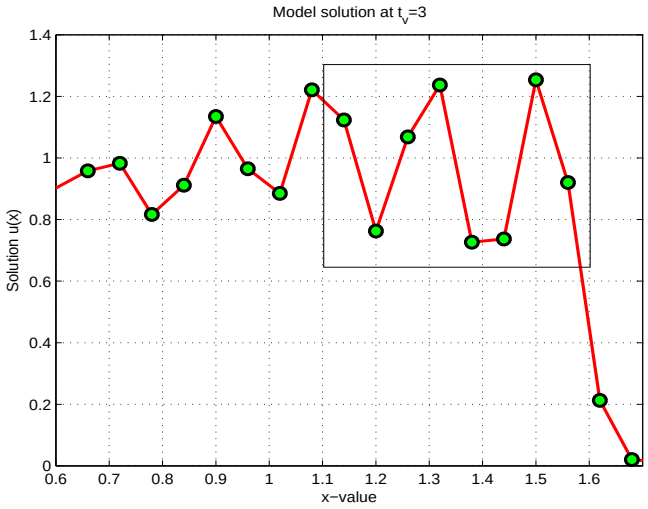

Figure 4.6: Verification domain is shown in rectangular box. Figure displays the model solution for $R = 300$.

In the equation (2.33) we see that the cost function requires background information as well as observations. Since we have used Burgers equation model, we assume that the background term is zero. For the observation term we generated the observation vector from the model solution initialized with x_0 by perturbing with normally distributed random numbers $N(0, \sigma^2)$. The standard deviation is chosen as $\sigma = .05$. The observation errors are assumed uncorrelated, so the observation error covariance matrices $\mathbf{R}_i$ are diagonal. In our case, all the matrices are the same and the values of diagonal entries are σ^2. The observational data are available at all grid points in spatial dimension. We took the routine observations at every 5th time step between $t_0 = 1$ and $t_N = 100$, where $[t_0, t_N]$ is our window of data assimilation. We calculated the gradient of the cost function $\mathcal{J}$ with respect to control variable which is required for minimization routines. For our experiment we used the L-BFGS (limited memory Quasi-Newton) minimization algorithm (Liu and Nocedal 1989 [47]) to solve the nonlinear 4D Var minimization problem with convergence criteria $||\nabla_{x_0}\mathcal{J}|| \leq 10^{-5}$ where $||\cdot||$ denotes the Euclidean norm in $\mathbb{R}^n$. To calculate the gradient of the cost functional with respect to control variable we used the adjoint model. We verified the correctness of the gradient according to the procedure discussed in Navon et al. 1992[22].

The formula is written as a function of α

$$\phi(\alpha) = \frac{J(x + \alpha \nabla J) - J(x)}{\alpha \nabla J^{\mathrm{T}} \nabla J} = 1 + O(\alpha)$$

Therefore the gradient provided by the adjoint model is assumed to be accurate up to machine accuracy if $\lim_{\alpha \to 0} \phi(\alpha) = 1$. The graph of the value obtained from the **alpha test** is displayed in Figure 4.7. The figure shows that the value of the $\phi(\alpha) \approx 1$ with increasing number of significant digits as α decreases.

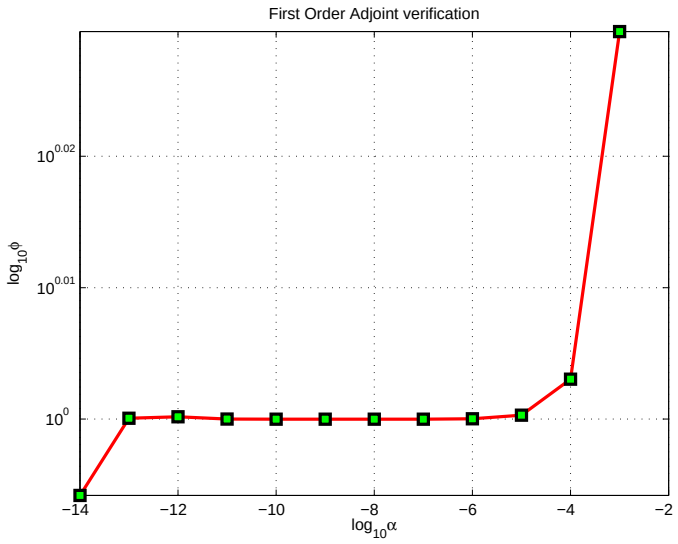

Figure 4.7: Verification of the first order adjoint.

4.2 Results

Our main objective is to reduce the forecast error on the verification domain $\mathcal{D}_v$ at verification time t_v. To fulfill our objective we defined the cost function in (3.11) over the verification domain at verification time and studied two different approaches:– adjoint sensitivity and observation sensitivity and compared them. Adjoint method is rather simple and straightforward. It just requires the value of the gradient of the cost functional $\mathcal{J}_v$ which is defined in (3.12). On the other hand, calculation of observation sensitivity is not trivial. It requires the information of the Hessian of the cost function defined in (3.15). To calculate the Hessian requires the availability of the second order adjoint model which can provide each column of the Hessian.

4.2.1 Adjoint sensitivity

The adjoint sensitivity vector $F_v(x_i) \in \mathbb{R}^{n_x}$ is calculated by using the formula (3.13) which requires the gradient of the cost functional $\mathcal{J}^v$. The gradient $\nabla\mathcal{J}^v(x_i) \in \mathbb{R}^{n_x}$ is calculated by the equation (3.12). As the formula requires, we first calculated x_v^f and x_v^t. We used numerical model and the given initial condition x_0^t in (4.1) in order to calculate x_v^t at t_v. We then calculated x_v^f at t_v by using (4.2) with optimal initial condition x_0^a obtained by minimizing the cost function (2.33). We took the difference $x_v^f - x_v^t$ and used it as the initial condition to the adjoint model $\mathcal{M}^T$. We then calculate the gradient$\nabla\mathcal{J}^v(x_i) \in \mathbb{R}^{n_x}$ at each $t_i \in [t_0, t_N]$. Once we obtained the gradient, we chose the target instant $\tau_k,\ k = 1, 2, .., I$. In our experiment we found that L_∞ norm of the gradient vector at each t_i is increasing as t_i is increasing. Thus, we chose every 10th time step starting from initial time step t_0 as the target instant τ_k to place the adaptive observations. We then calculated the optimal locations for adaptive observation according to the method discussed in section 3.1.1. The adjoint sensitivity vectors for few time steps are displayed in Figures 4.8, 4.9 and 4.10. The locations of adaptive observations are shown in Figures 4.11, 4.12 and 4.13. We decided to choose 5 adaptive observations at each target instant τ_k.

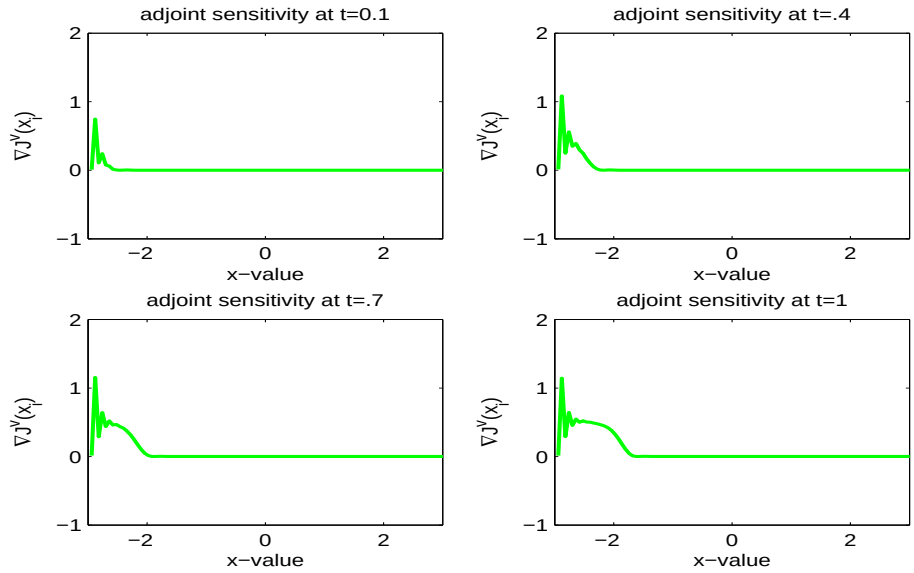

Figure 4.8: Figure of the adjoint sensitivity vector at different time steps for $R = 100$.

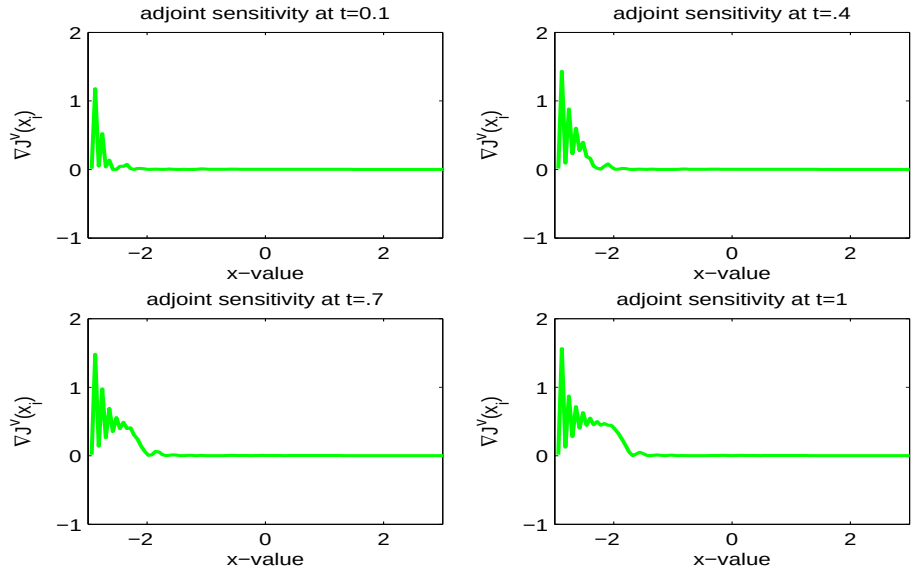

Figure 4.9: Figure of the adjoint sensitivity vector at different time steps for $R = 200$.

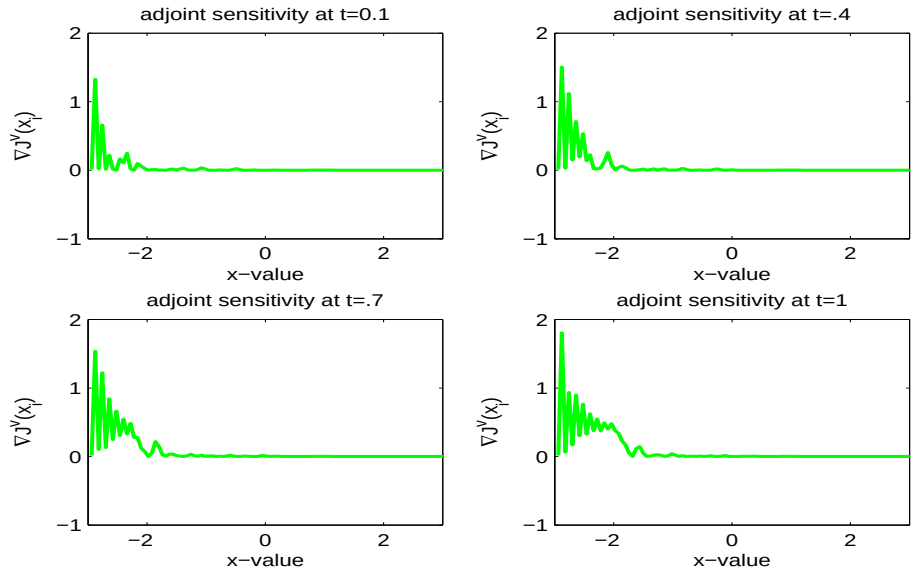

Figure 4.10: Figure of the adjoint sensitivity vector at different time steps for $R = 300$.

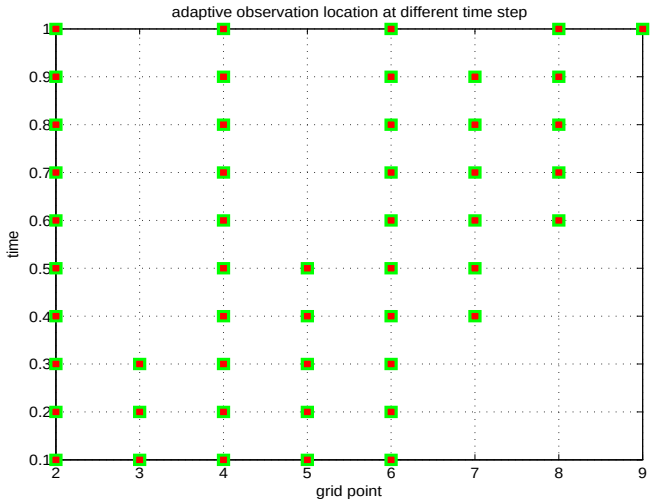

Figure 4.11: Locations of adaptive observations based on adjoint sensitivity vector at target instant τ_k. Figure shows the result for $R = 100$.

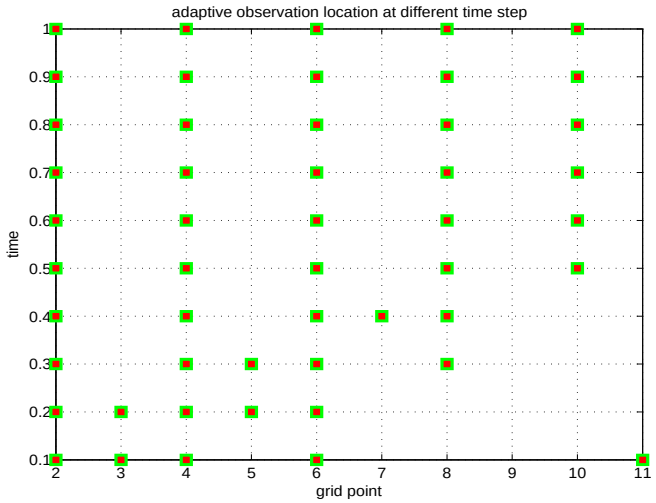

Figure 4.12: Locations of adaptive observations based on adjoint sensitivity vector at target instant τ_k. Figure shows the result for $R = 200$.

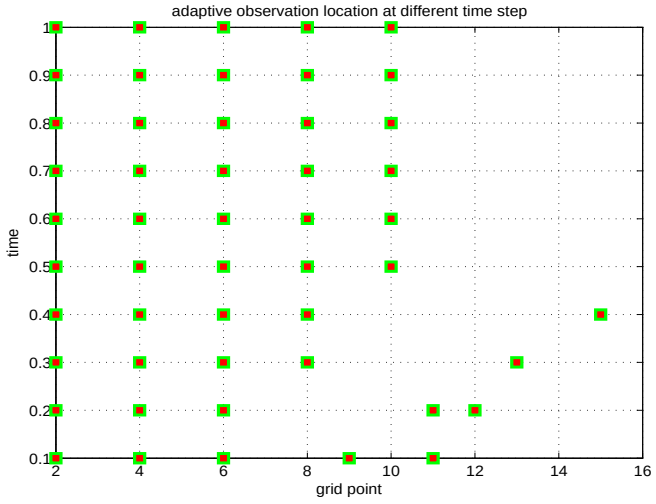

Figure 4.13: Locations of adaptive observations based on adjoint sensitivity vector at target instant τ_k. Figure shows the result for $R = 300$.

4.2.2 Observation sensitivity

We calculate the observation sensitivity by using the algorithm mentioned at the beginning of this chapter. In the first stage we initialize $\nabla_{x_v}\mathcal{J}^v$ given in equation (3.21). This value is used as initial condition for adjoint model and integrate the model backward to obtain the gradient $\nabla_{x_0^a}\mathcal{J}^v$ at optimal initial analysis vector x_0^a. The minimization process of the cost function $\mathcal{J}$ in 4D Var method provides the optimal analysis initial condition x_0^a. Once we have the gradient at t_0 we solve the linear system $\mathbf{A}^{-1}z_0 = \nabla_{x_0^a}$ where $\mathbf{A}^{-1}$ is the inverse covariance matrix at optimal solution which is equivalent to Hessian matrix. It is computationally very expensive to calculate Hessian matrix explicitly for a large scale model but much cheaper for a small scale model. Since our model is one dimensional with 10^2 variables, we easily calculated the Hessian explicitly . We used a second order adjoint model to obtain the full Hessian. To obtain each column of the Hessian of the cost function we used the unit vectors. For obtaining the entries of j th column we used unit vector e_j whose jth entry is 1 and all the other entries are zeros. We obtain that the matrix is symmetric and positive definite for the three cases tested $R = 100, 200, 300$.

To prove that the Hessian is positive definite we calculated the eigenvalues of it by using MATLAB. The Hessian eigenvalues are shown for all the cases in Figures 4.14, 4.15 and 4.16. These figures show that all the eigenvalues are positive so the matrix is positive definite for each case tested.

To show the Hessian is symmetric for each case we compute $d(k) = |h(i,j) - h(j,i)|$, where $k = 1, .., n_x^2$ and $h(i,j), i, j = 1, .., n_x$ are the elements of the Hessian. Then we calculate L_∞ norm of d and obtained that $||d||_\infty = 0$. It means $h(i,j) = h(j,i)$ which is the condition for symmetry.

Since the Hessian is symmetric and positive definite, we solved the linear system by employing Cholesky Decomposition method to obtain upper triangular matrix and then used backward substitution method to obtain the solution of the system. We then used tangent linear model $\mathcal{M}_{o,i}$, linearized observational operator $\mathbf{H}_i$ and observation error covariance matrix $\mathbf{R}_i^{-1}$ in order to obtain the observation sensitivity vector $\nabla_{y_i}\mathcal{J}^v \in \mathbb{R}^{n_x}$ at t_i. In our case we assumed that $\mathbf{H}_i$ at t_i are unit matrix where routine observations are available (at every 5th time step between $t_0 = 1$ and $t_N = 100$). As mentioned before, the observation error covariance matrices $\mathbf{R}_i$ are diagonal matrices whose diagonal entries are σ^2. So $\mathbf{R}_i^{-1}$

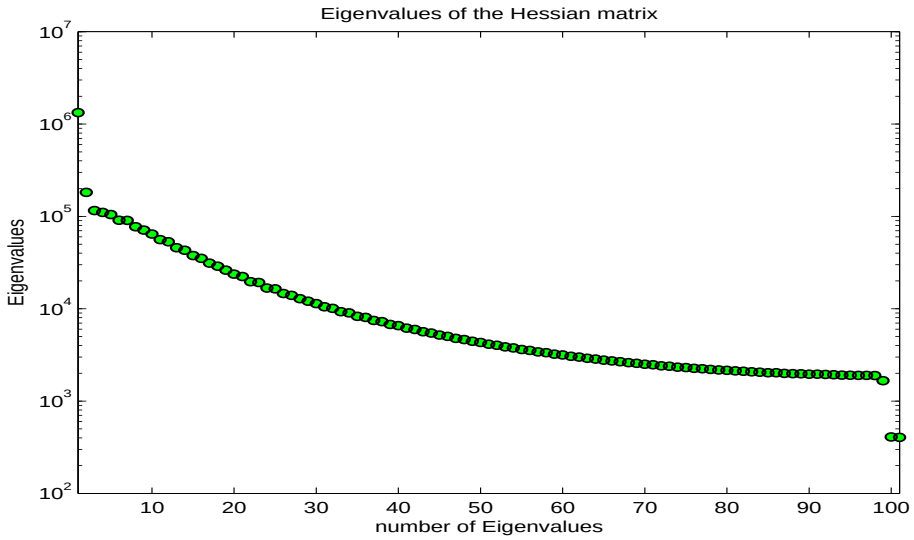

Figure 4.14: Eigenvalues of the Hessian calculated with Reynolds number $R = 100$.

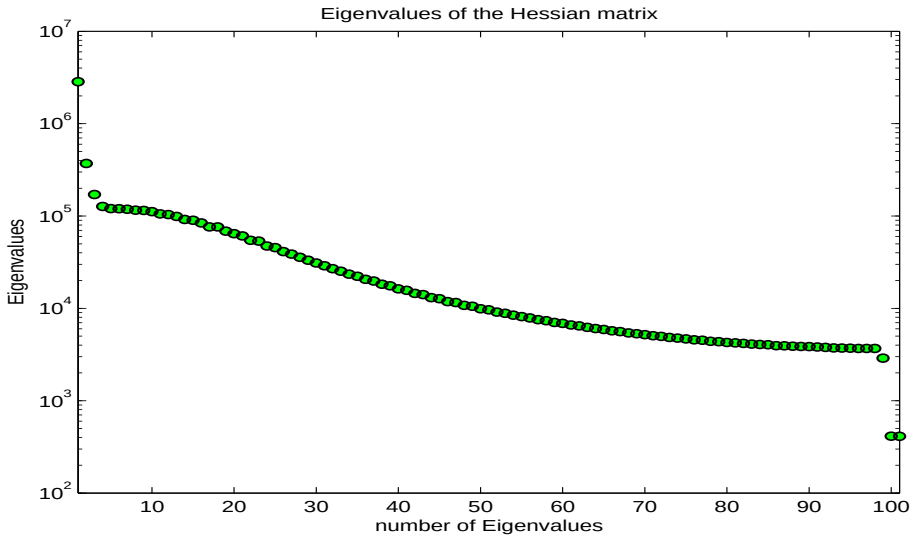

Figure 4.15: Eigenvalues of the Hessian calculated with Reynolds number $R = 200$.

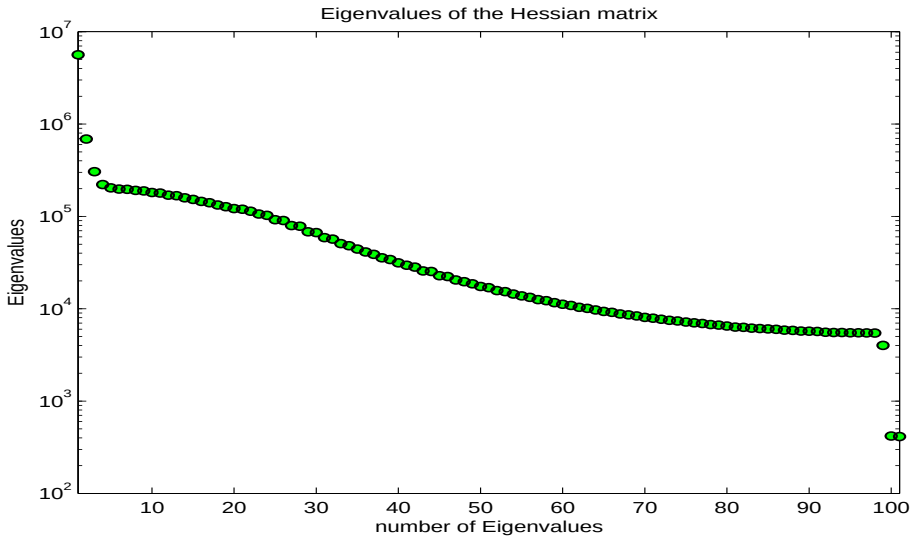

Figure 4.16: Eigenvalues of the Hessian calculated with Reynolds number $R = 300$.

are also diagonal with diagonal entries $\frac{1}{\sigma^2}$. The observation sensitivity vectors $\nabla_{y_i}\mathcal{J}^v$ are available only at time steps where routine observations are available and zeros at all the other time steps. Observation sensitivity vectors are displayed in Figures 4.17, 4.18 and 4.19.

It is well known that in general the more observations used in the data assimilation process, the better the forecast obtained. But sometimes more observations result in a worse forecast due to non-homogeneous distribution of the observations. For geophysical models in meteorology and oceanography it is known that there are more observations on the land and less observations on the ocean, especially in the Southern Hemisphere. It is expensive to deploy the observational devices over the ocean. But if we can ascertain that the observations at certain space and time locations have greater impact on the forecast at verification time, we can attempt to collect observational data by deploying observational devices only for those space and time locations. In this work we implemented this idea and we used observation sensitivity method as a guiding metric to obtain appropriate space and time locations to deploy adaptive observations so that the forecast error is reduced on the verification domain at t_v significantly. We chose the adaptive observation locations on the

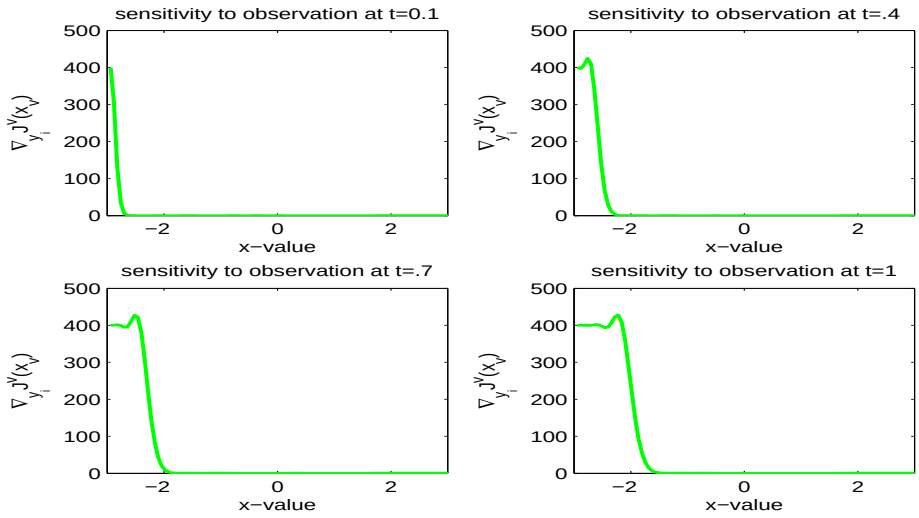

Figure 4.17: Figure displays the observation sensitivity vector for $R = 100$ at different time steps.

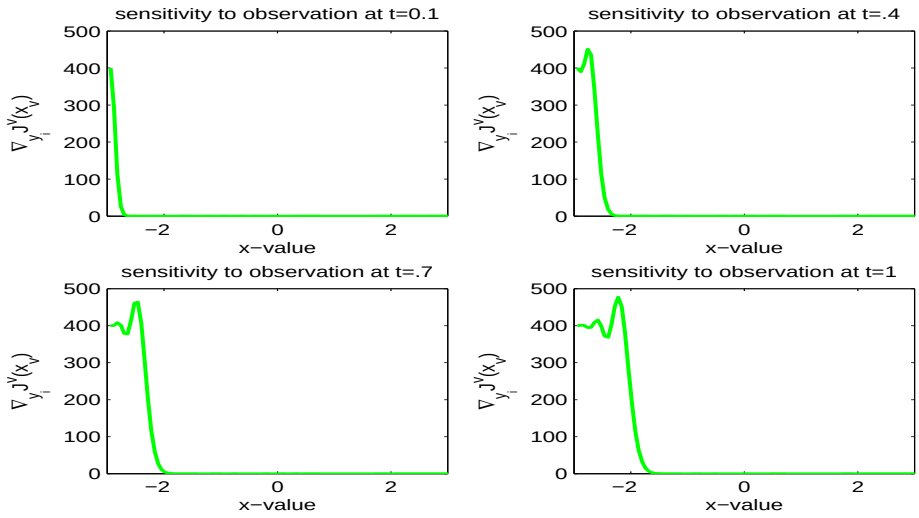

Figure 4.18: Figure displays the observation sensitivity vector for $R = 200$ at different time steps.

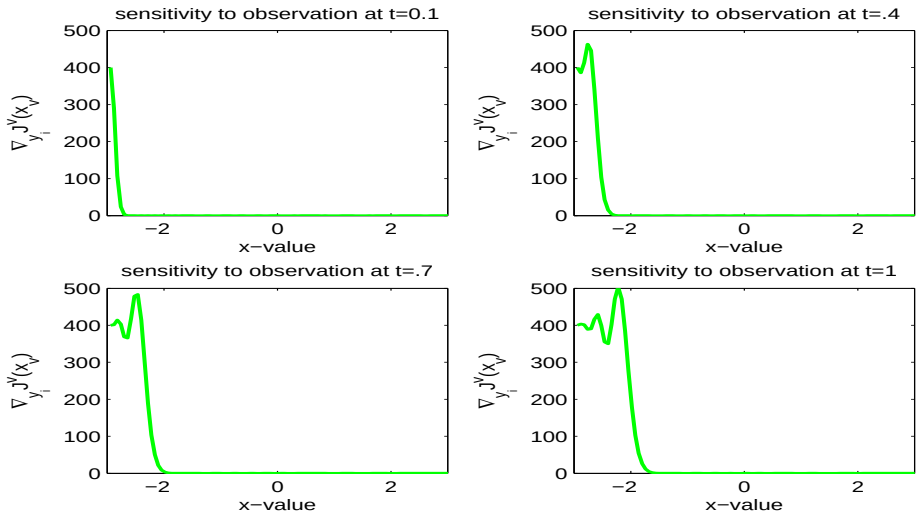

Figure 4.19: Figure displays the observation sensitivity vector for $R = 300$ at different time steps.

basis of the method discussed in section 3.2.1. In this method we first chose target instants τ_k, $k = 1, .., 10$ at which the L_∞ norm of $\nabla_{y_i}\mathcal{J}^v$ attains maximum value and then at each τ_k we took 5 adaptive observations so that $||\nabla_{y_{\tau_k}}\mathcal{J}^v||_E$ attains a larger value at each τ_k. The locations of the adaptive observations are displayed in Figures 4.20, 4.21 and 4.22.

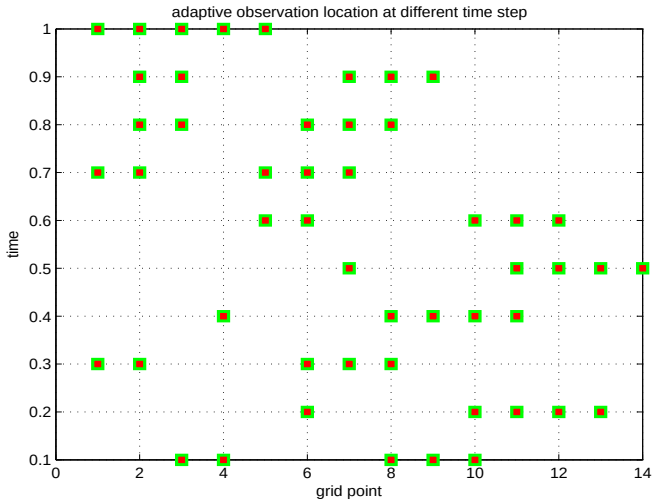

Figure 4.20: Location of adaptive observations based on observation sensitivity vector for $R = 100$.

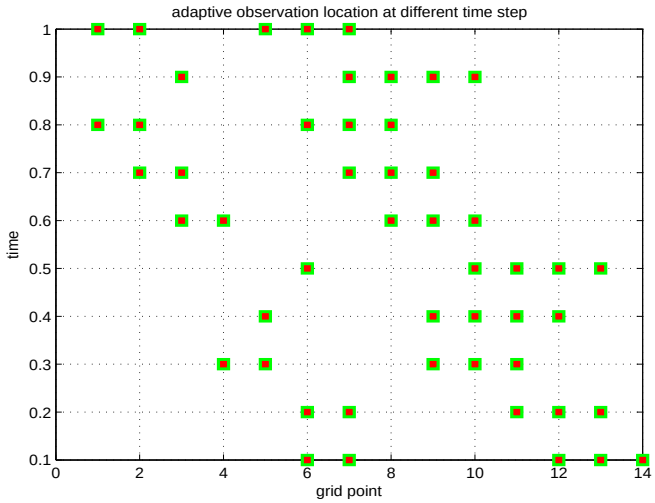

Figure 4.21: Location of adaptive observations based on observation sensitivity vector for $R = 200$.

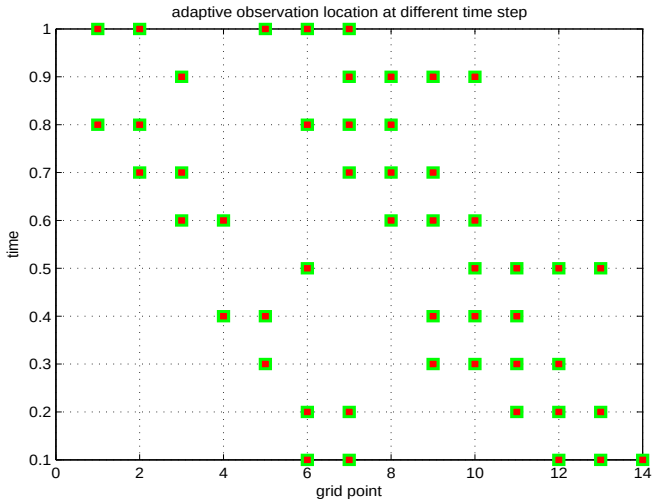

Figure 4.22: Location of adaptive observations based on observation sensitivity vector for $R = 300$.

We perform the data assimilation with routine and adaptive observations by using 4D-Var method. First, we minimize the cost functional with routine observations. The initial guess provided to the minimization routine is produced by perturbing the given initial condition x_0 with Gaussian random number by 20%. At each iteration the minimization routine provides the initial analysis x_0^a and the corresponding value of cost function $\mathcal{J}(x_0^a)$ and gradient $\nabla\mathcal{J}(x_0^a)$. We also calculated the cost function $\mathcal{J}^v$ using the equation (3.11) which requires x_v^t and x_v^f. x_v^f is calculated using Burgers equation model with x_0^a obtained by minimizing the cost function in (2.33) whereas x_v^t is fixed for all iteration and calculated with given initial condition x_0^t. We then used adaptive observations along with routine observations to perform the data assimilation with the same method, 4D-Var method. We again calculated initial analysis vector x_0^a, cost function $\mathcal{J}(x_0^a)$ as well as $\mathcal{J}_v(x_0^a)$ at each iteration. The only difference is that we provided one additional term for adaptive observations to the cost function in (2.33). The figures are displayed in Figures 4.23, 4.24 and 4.25. We found that when we used only few adaptive observations the forecast error is reduced significantly. We also displayed the evolution of cost function in Figures 4.26, 4.27 and 4.28 as well as its gradient in Figures 4.29, 4.30 and 4.31.

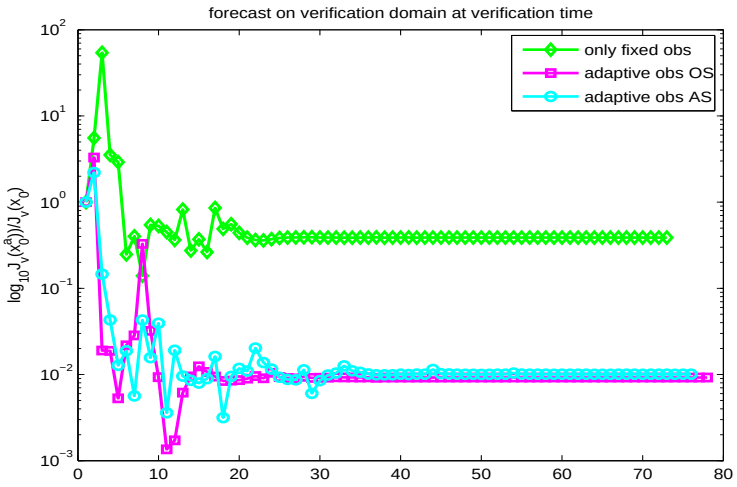

Figure 4.23: In every iteration we calculate the forecast error reduction at t_v over verification domain which is quantified by $\mathcal{J}^v(x_0^a)/\mathcal{J}^v(x_0)$. The normalized values $\mathcal{J}^v(x_0^a)/\mathcal{J}^v(x_0)$ are shown on semi-logarithmic scale. Figure shows the result for $R = 100$.

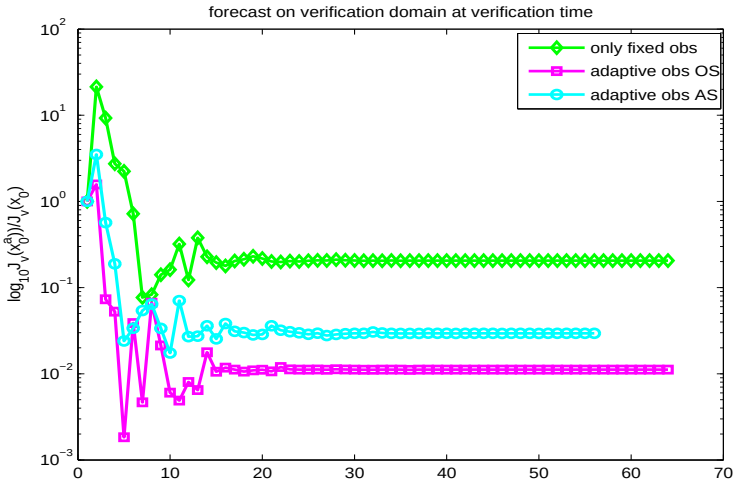

Figure 4.24: In every iteration we calculate the forecast error reduction at t_v over verification domain which is quantified by $\mathcal{J}^v(x_0^a)/\mathcal{J}^v(x_0)$. The normalized values $\mathcal{J}^v(x_0^a)/\mathcal{J}^v(x_0)$ are shown on semi-logarithmic scale. Figure shows the result for $R = 200$.

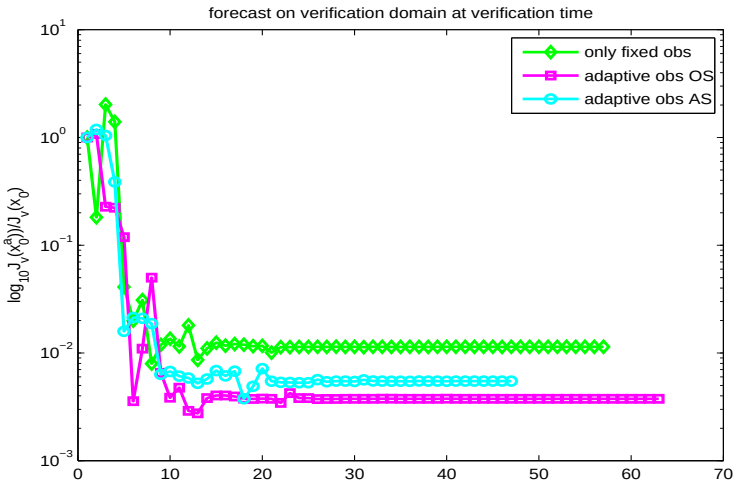

Figure 4.25: In every iteration we calculate the forecast error reduction at t_v over verification domain which is quantified by $\mathcal{J}^v(x_0^a)/\mathcal{J}^v(x_0)$. The normalized values $\mathcal{J}^v(x_0^a)/\mathcal{J}^v(x_0)$ are shown on semi-logarithmic scale. Figure shows the result for $R = 300$.

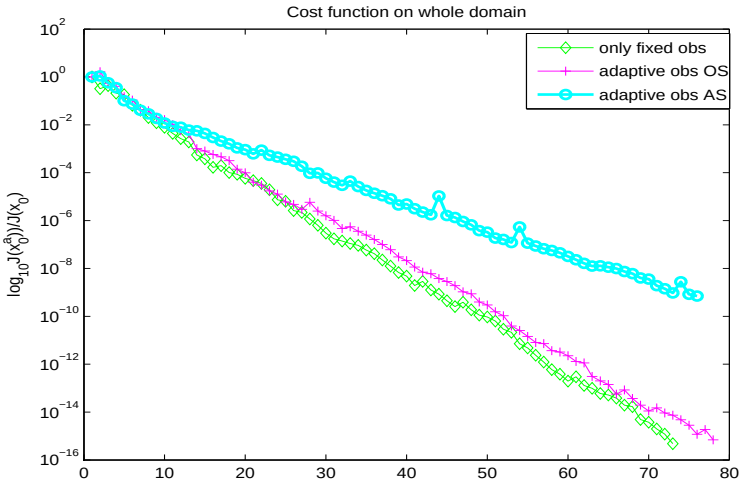

Figure 4.26: The minimization of the cost function $\mathcal{J}$ when both routine and adaptive observations are assimilated. The normalized values $\mathcal{J}(x_0^a)/\mathcal{J}(x_0)$ are shown on semi-logarithmic scale. Figure shows the cost function at every iteration for $R = 100$.

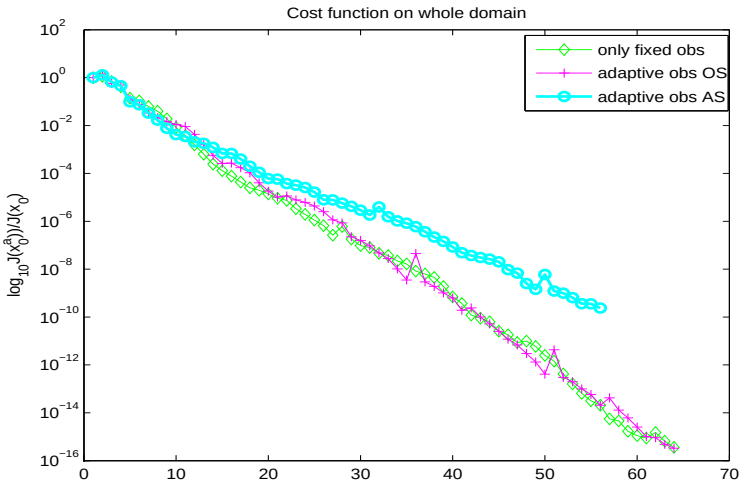

Figure 4.27: The minimization of the cost function $\mathcal{J}$ when both routine and adaptive observations are assimilated. The normalized values $\mathcal{J}(x_0^a)/\mathcal{J}(x_0)$ are shown on semi-logarithmic scale. Figure shows the cost function at every iteration for $R = 200$.

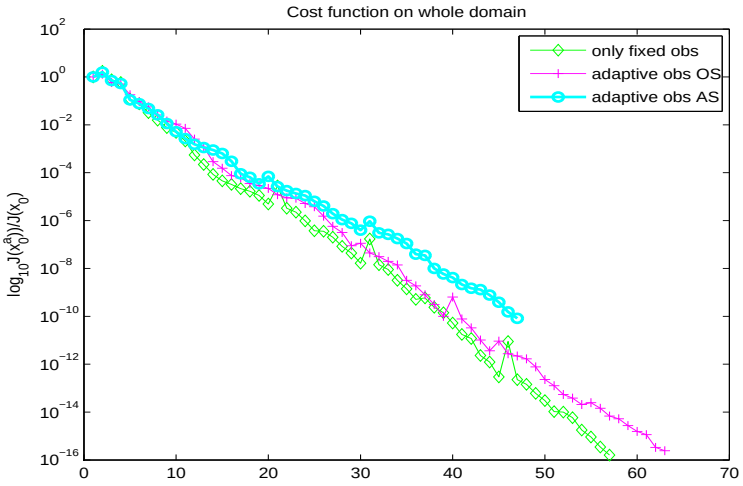

Figure 4.28: The minimization of the cost function $\mathcal{J}$ when both routine and adaptive observations are assimilated. The normalized values $\mathcal{J}(x_0^a)/\mathcal{J}(x_0)$ are shown on semi-logarithmic scale. Figure shows the cost function at every iteration for $R = 300$.

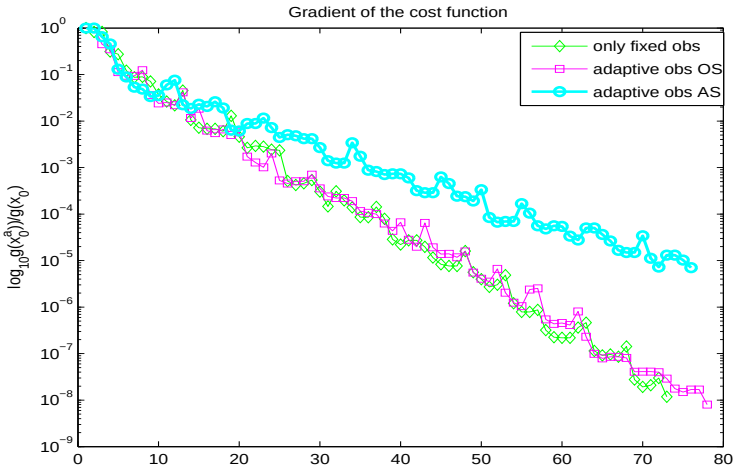

Figure 4.29: The gradient of the cost function $\nabla_x \mathcal{J}(x_0^a)$ are shown when both routine and adaptive observations corresponding to AS and OS are assimilated. The normalized values are shown on semi-logarithmic scale. The figure shows the gradient of the cost function for $R = 100$.

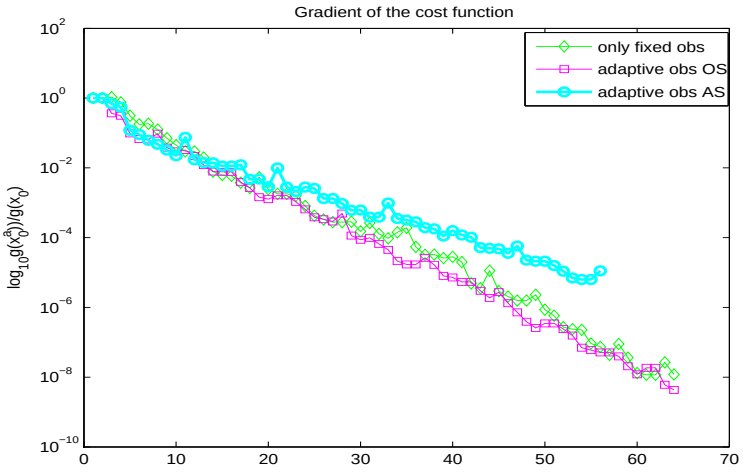

Figure 4.30: The gradient of the cost function $\nabla_x \mathcal{J}(x_0^a)$ are shown when both routine and adaptive observations corresponding to AS and OS are assimilated. The normalized values are shown on semi-logarithmic scale. The figure shows the gradient of the cost function for $R = 200$.

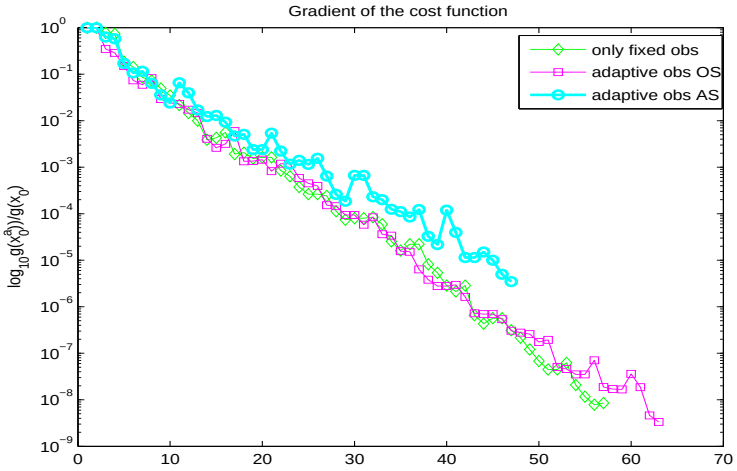

Figure 4.31: The gradient of the cost function $\nabla_x \mathcal{J}(x_0^a)$ are shown when both routine and adaptive observations corresponding to AS and OS are assimilated. The normalized values are shown on semi-logarithmic scale. The figure shows the gradient of the cost function for $R = 300$.

4.2.3 Comparison of two targeting methods

We begin our analysis by presenting the results obtained from our experiment when routine observations as well as adaptive observations are being assimilated. During the iterative process we monitored the distribution of the forecast error over the verification domain at the verification time after the data assimilation has taken place. First we used only routine observations for data assimilation and then used adaptive observations with routine observations. We used both the targeting methods:– adjoint and observation sensitivity methods for finding optimal sites for adaptive observations. The results for forecast errors are presented in the Figures 4.23, 4.24 and 4.25. These figures show that the forecast errors improved significantly when adaptive observations are used. Again we noticed that observation sensitivity method performs better than the adjoint sensitivity method for all the three cases tested i.e $R = 100, 200, 300$. The performance of minimization routine depends on the Reynolds number. This is evident from the Figures 4.23 4.24 and 4.25 that the OS method performs superbly for $R = 200$.

Table 4.1: Performance of minimization routine with different Reynolds numbers when routine and adaptive observations corresponding to AS and OS are used.

Reynolds number	Observations	number of iterations
100	Routine	76
	Adaptive AS	76
	Adaptive OS	78
200	Routine	63
	Adaptive AS	56
	Adaptive OS	64
300	Routine	57
	Adaptive AS	47
	Adaptive OS	63

CHAPTER 5

SUMMARY AND CONCLUSIONS

In this work we applied two different targeting methods– adjoint sensitivity and observation sensitivity to find the optimal sites of observational data that have greater impact on the forecast error reduction. On the basis of the information provided by these procedures we chose the adaptive observations space and time locations. These methods have been previously tested only for targeting adaptive observations distributed at space locations. We realized that since 4D-var is time dependent and uses time distributed observations, it is important to find the time location for adaptive observations. We chose the time location for the case of observation sensitivity on the basis of the largest value of the norm of OS vector at each targeting instant τ_k. On the other hand we chose the time location for the case of adjoint sensitivity at every 10th time step in the assimilation window proceeding backward from the final target time $t_N = 100$.

In our experiment we found that observation sensitivity approach performed better than the adjoint sensitivity approach. But there was a computational cost for obtaining a better forecast i.e., we needed to provide Hessian matrix information. To calculate the Hessian of the cost function, 4D-Var requires use of second order adjoint. For large scale modeling it is impossible to compute this matrix explicitly due to the limitations of computer memory. But the evolved system of linear equations can be solved by using iterative procedure. This may require matrix vector multiplication which can be provided by using second order adjoint. In our experiment we saw that a few adaptive observations added to the large number of existing routine observations can improve the forecast significantly. In future work we want to use 2D model, for instance, shallow water model to implement the idea of observation sensitivity method. We may also consider the techniques of reduced order modeling to reduce the cost for obtaining Hessian information.

REFERENCES

[1] F.X. LeDimet and I.M. Navon. Variational and optimization methods in meteorology: SCRI report. Technical Report 88-144, 1988.

[2] M. Ghil, S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson. *Dynamic Meteorology: Data Assimilation Methods, chapter Applications of estimation theory to numerical weather prediction*. Springer-Verlag, New York, 1981.

[3] M. Ghil and P. Malanotte-Rizzoli. Data assimilation in meteorology and oceanography. *Adv. Geophys.*, 33:141–266, 1991.

[4] F.X. LeDimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, 38 A:97–110, 1986.

[5] J. Derber. Variational four-dimensional analysis using quasi-geostrophic constraints. *Mon. Wea. Rev.*, 115:998–1008, 1987.

[6] J.M. Lewis and J.C. Derber. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, 37A:309–322, 1985.

[7] O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: Theory. *Q. J. Roy. Meteor. Soc.*, 113:1311–1328, 1987.

[8] J. G. Charney, R. Fjörtoft, and John Von Neumann. Numerical integration of the barotropic vorticity equation. *Tellus*, 2:237–254, 1950.

[9] H. Panofsky. Objective weather-map analysis. *J. Appl. Meteor.*, 6:386–392, 1949.

[10] B. Gilchrist and G. Cressman. An experiment in objective analysis. *Tellus*, 6:309–318, 1954.

[11] L. S. Gandin. Objective analysis of meteorological fields. Israel Program for Scientic Translations, 1963.

[12] A. Lorenc. Analysis methods for numerical weather prediction. *Q. J. Roy. Meteor. Soc.*, 112:1177–1194, 1986.

[13] Y.K. Sasaki. An objective analysis based on the variational method. *J. Met. Soc. Jap.*, II(36):77–88, 1958.

[14] Y.K. Sasaki. Proposed inclusion of time variation terms, observational and theoretical , in numerical variational objective analysis. *J. Met. Soc. Jap.*, 47:115–124, 1969.

[15] Y.K. Sasaki. Some basis formalisms in numerical variational analysis. *Mon. Wea. Rev.*, 98:875–883, 1970.

[16] R. N. Hoffmann. A four-dimensional analysis exactly satisfying equations of motion. *Mon. Wea. Rev.*, 114:388–397, 1986.

[17] J.L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer-Verlag, Berlin Heidelberg New York, 1971.

[18] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems*. Springer-Verlag, New York, 1984.

[19] F.X. LeDimet, H.E. Ngodock, B. Luong, and J. Verron. Sensitivity analysis in data assimilation. *J. Meteor. Soc. Japan*, 75:245–255, 1997.

[20] D.F. Parrish and J.D. Derber. The national meteorological center spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, 120:1747–1763, 1992.

[21] J.N. Thepaut and P. Courtier. Four-dimensional variational assimilation using the adjoint of a multilevel primitive-equation model. *Q. J. Roy. Meteor. Soc.*, 117:1225–1254, 1991.

[22] I. M. Navon, X. Zou, J. Derber, and J. Sela. Variational Data Assimilation with an Adiabatic Version of the NMC Spectral Model. *Mon. Wea. Rev.*, 120:1433–1446, 1992.

[23] M. Zupanski. Regional 4-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.*, 121:2396–2408, 1993.

[24] J.N. Thepaut, R.N. Hoffman, and P. Courtier. Interactions of dynamics and observations in a 4-dimensional variational assimilation. *Mon. Wea. Rev.*, 121:3393–3414, 1993.

[25] D.N. Daescu and I.M. Navon. Adaptive observations in the context of 4D-Var data assimilation. *Meteorol. Atmos. Phys.*, 85:205–226, 2004.

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

[27] J. Barkmeijer, M. Van Gijzen, and F. Bouttier. Singular vectors and estimates of the analysis-error covariance metric. *Q. J. Roy. Meteor. Soc.*, 124:1695–1713, 1998.

[28] J. Barkmeijer, R. Buizza, and T.N. Palmer. 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Q. J. Roy. Meteor. Soc.*, 125:2333–2351, 1999.

[29] M. Ehrendorfer and J. J. Tribbia. Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, 54:286–313, 1997.

[30] F.X. LeDimet, I.M. Navon, and D.N. Daescu. Second order information in data assimilation. *Mon. Wea. Rev.*, 130(3):629–648, 2002.

[31] N. L. Baker and R. Daley. Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Q. J. Roy. Meteor. Soc.*, 126:1431–1454, 2000.

[32] X. Zou, I.M. Navon, and J.G. Sela. Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus*, 45A:370–387, 1993.

[33] D.G. Cacuci. Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach. *J. Math. Phys.*, 22:2794–2802, 1981.

[34] A. Doerenbecher and T. Bergot. Sensitivity to observations applied to FASTEX cases. *Nonlinear Processes in Geophys.*, 8:467–481, 2001.

[35] N. Fourrié, A. Doerenbecher, T.Bergot, and A. Joly. Adjoint sensitivity of the forecast to TOVS observations. *Q. J. Roy. Meteor. Soc.*, 128:2759–2777, 2002.

[36] Y. Zhu and R. Gelaro. Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Mon. Wea. Rev.*, 136:335–351, 2008.

[37] Yannick Trémolet. Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus*, 60A:964–978, 2008.

[38] Dacian N. Daescu. On the sensitivity equations of four-dimensional variational (4D-Var) data assimilation. *Mon. Wea. Rev.*, 136:3050–3065, 2008.

[39] Zhi Wang, I.M. Navon, X. Zou, and F.X. LeDimet. A truncated newton optimization algorithm in meteorology applications with analytic hessian/vector products. *Computational Optimization and Applications*, 4:241–262, 1995.

[40] P. Bergthorsson and B. Döös. Numerical weather map analysis. *Tellus*, 7:329–340, 1955.

[41] G.P. Cressman. An operational objective analysis system. *Mon. Wea. Rev.*, 87:367–374, 1959.

[42] E. Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge Univ. Press, New York, 2003.

[43] F. Bouttier and P. Courtier. Data assimilation concepts and methods, 2002.

[44] A. Da Silva, J. Pfaendtneer, J. Guo, M. Sienkiewicz, and S. Cohn. Assessing the effects of data selection with DAO's physical-space statistical analysis system. In *Proceedings of the second international symposium on the assimilation of observations in meteorology and oceanography, Tokyo, Japan*. World Meteorological Organization and Japan Meteorological Agency, Tokyo, Japan, 1995.

[45] J.D. Cole. On a quasi-linear parabolic equation occurring in aerodynamics. *Quart. Appl. Math.*, 9:225–236, 1951.

[46] E. Hopf. The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Comm. Pure Appl. Math.*, 3:201–230, 1950.

[47] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large-scale minimization. *Math. Prog*, 45:503–528, 1989.

BIOGRAPHICAL SKETCH

MD. JAKIR HOSSEN

Md. Jakir Hossen was born on May 5, 1975 in a village (called Changahata) in Comilla, Bangladesh. He graduated from the University of Dhaka in 1999 with a degree in Mathematics. He completed his M.Sc. degree in Applied Mathematics from the same university in 2001. He had been working as a faculty member in the Department of Mathematics and Natural Science, BRAC University from August 2002 to August 2006. In the fall of 2006, he started his graduate studies in Computational Science at Florida State University. His research interests include scientific computing, probabilistic data assimilation, risk and decision analysis, computational fluid dynamics etc.