

The impact of background error on incomplete observations for 4-D VAR data assimilation with the FSU GSM *

Zhuo Liu and I.M. Navon

(School of Computational Science and Information Technology,
Florida State University, Florida, 32306)

ABSTRACT

In the study of 4D Var data assimilation of atmospheric models, an important issue to address is the case of incomplete observations in either the space or time dimension.

In an ideal setting 4-D Var data assimilation assumes the observation data field to be complete, and if there are gaps in the data, these are being taken into account in the process of data assimilation.

To assess the impact of incomplete observations on the 4D Var data assimilation, we carried out some assimilation experiments with a model consisting of the dynamical core of FSU GSM by reducing the number of observations in both the space and time dimensions respectively.

The impact of the J_b background error covariance term on problem of incomplete observations in either time or space direction has been investigated using the new FSU GSM consisting of a T126L14 global spectral model in a parallel environment using MPI version of its adjoint model.

Numerical experiments aimed at assessing impact of incomplete observations on 4-D Var data assimilation were carried out as follows: first, a twin experiment with observations available at every model grid point (thereafter referred to complete observations) was carried out using the dynamic core of FSU GSM and its adjoint model to ensure that the assimilation system is well

*Supported by N.S. F.Grant ATM-0201808

constructed. For such an experiment, one knows in advance the exact solution, and the minimum value of the cost function is zero. We then reduced the available observations to every 2, 4 or 8 spatial grid points, respectively. We also carried out another set of experiments with data holes where all the observations were missing, *e.g.* at ocean grid points locations. We then carried out experiments reducing the number of time instants where observations are available only every 2, 4 or 8 time steps in the window of assimilation .

The results obtained show that spatial incomplete observations lead to a slow down in the cost functional minimization. Although the decrease rate of cost function with incomplete observations where observations are available only at every 2, 4 or 8 observation grids exhibited a similar pattern, the impact(degree) of reasonable retrieval initial data strongly depends on the density of observations. The impact of incomplete observations is even more pronounced for experiments where no observations were available over oceans, in which case the lack of fit between a control run and the aforementioned could not be reduced.

In contrast to above results, experiments involving reduction of the number of time instants where observations are available in the assimilation window allowed a successful retrieval of the initial data. The results obtained were insensitive to whether observational data was available only every 2,4 or 8 time steps versus that of the full observation.

To sum-up, the lack of the observations in grid space strongly affect the results of the minimization and retrieval of initial data, while that in time dimension or some variables will have no significant affection on the results.

Impact of various scenarios of incomplete observations on ensuing forecasts, and root mean square error were investigated for 24-72h forecasts for cases when the cost functional included and/or excluded the background covariance term.

To further investigate the issue of incomplete observations, we carried out another set of

experiments by adding a background term J_b to the cost function. Background state propagates information from observations at early times into the data holes. By considering an assimilation with a single observation, it can be shown that the background error covariance matrix controls the way in which information is spread from that observation to provide statistically consistent increments at the neighboring grid points.

1 Introduction

The purpose of data assimilation is to generate an analysis that represents the present state of the atmosphere. The analysis should be as close to the truth as possible.

Data assimilation provides, for a given time period, the window of assimilation, an estimate of the system evolution which is optimally adjusted to fit both the observations and the model. It allows a historical analysis of the behavior of the system, the location of specific behavior, *etc.* It is also a relevant tool for testing physical assumptions and for sensitivity studies. The aim of 4-D Var is to estimate the current best state of the system using all past and present information. Only the current state is optimal. This approach is widely used in operational forecasting methods in meteorology. It consists in finding, using informations provided by a field of observations over a period, a model solution which is as close as possible to the observations in order to provide accurate initial conditions (or boundary conditions) for the following forecast.

4-D Var data assimilation assumes the observation data to be complete, nevertheless, if there are gaps in the data, both in the spatial or temporal dimensions, these can be taken into account in the variational data assimilation.

A way of overcoming this problem is to fill the gaps in the data using an adequate interpolation method. If the gaps are not too large, the differences in the results using the original and the interpolated data set should be small.

In the southern hemisphere, and tropics and most of the surface is covered by oceans, which lack the density of conventional wind-profile data coverage and uniform distribution available over the the Northern hemisphere.

Current forecast model, data assimilation systems and the development of space-based and other advanced observing systems have led to a slow, albeit steady reduction in short- and medium-range forecast error over the past few decades.

However, despite these advances, it is still clear that a major component of forecast error is due to analysis error in relatively poorly sampled regions such as the mid-Pacific Ocean, the tropics and the southern oceans.

The deployment of dropsondes from aircraft or the more recently available unmanned aerial vehicles(UAVs) have made it possible to sample these poorly observed regions. However, such developments are costly, and only make sense if it can be determined in advance where such observations are likely to have the most positive impact on subsequent forecasts. Fortunately, there are now several objective techniques ideally suited for this problem.

Zou *et al.*[3] studied the incomplete observation in space and time dimension with a shallow water equation model, they showed that the case of insufficient data can lead to rather large changes in the pattern of behavior of minimizing the objective function, and suggested to add an additional term of penalization in the objective function in order to improve the conditioning of the Hessian of the objective function.

Zhu *et al.*[16] used an adjoint model of a finite-element shallow-water equation model to study the impact of different resolution of the model to the minimization of the cost function, which shows the coarse-mesh resolution model exhibited a faster convergence rate than the one corresponding to the fine-mesh model.

In this paper, we use a new MPI-based parallel version of FSU global spectral model to study

the incomplete observations in space and time dimension. The results show that for incomplete observation in space dimension, the convergence rate is slowed down and the effectiveness of optimized initial data with incomplete observation in space strongly depends on the density of observations in space dimension.

In order to investigate the impact of background error covariance in the 4-D Var data assimilation, we added the background covariance term to the cost function.

There are two main reasons for using the background term. The first reason is that it is vital to have a background term in data sparse areas or data void areas. The background state propagates information from observations at earlier times into the data holes. The background term is also necessary for information spreading and balance constraints. By considering an assimilation with a single observation, it can be shown that the background error covariance matrix controls the way in which information is spread from that observation to provide statistically consistent increments at neighboring grid points. The background error covariance matrix can also be used to ensure that observations of one model variable provide dynamically consistent increments in other model variables, by enforcing balance constraints such as geostrophic balance.

There are 3 basic ways to specify the background error covariances: In data dense areas, observational or Hollingsworth-Lonnberg method (Hollingsworth and Lonnberg(1986)). The method assumes separability and homogeneity to calculate correlation functions. A histogram of empirical data correlations is found, and curves are fitted using the method of least squares. A variety of correlation functions were examined in Julian and Thiebaut(1975). These climatological covariance matrices do not take into account any synoptic dependence.

The NMC method (Parrish and Derber(1992)) uses the differences between forecasts and analyses verifying at the same time, to specify the background error covariance matrices.

To obtain more flow-dependent background error covariances, Kalman filter techniques can be

used. The Kalman filter data assimilation method propagates the background error covariances explicitly in time (Ghil and Malanotte-Rizzoli (1991)). However, in 4D Var, the error covariances are propagated implicitly. This means that the 4D Var background error covariance matrices are also flow dependent as shown in studies by Thepaut et al. (1996). By performing single observation studies it is possible to examine the structure functions (correlations) at the end of a 4D Var assimilation window (Thepaut et al. (1993)).

The problem is in the specification of the background error covariance matrix at the beginning of the assimilation time window. In theory, it is possible to fully transfer information from the previous assimilation time window to the next via the background term (Li and Navon (2001)). This allows the specification of the background error covariance matrix at the beginning of the assimilation window. By combining the Kalman Filter with 4D Var it would be possible to specify the \mathbf{B} , however, this is not feasible computationally.

The structure of this paper is as follows. In section 2, we present the basic description of parallel version of FSU GSM. In section 3, we briefly introduce the adjoint of FSU Global Spectral Model. In section 4 we describe the experiments related to incomplete observations in the spatial and temporal domains, their impact on ensuing forecast skill and the role played by the background error covariance term. The significance of the numerical results is then further discussed. Finally in 5 the summary and conclusions are presented.

2 Description of parallel version of FSU GSM

2.1 Model description

The FSUGSM is a global hydrostatic primitive equation model. The prognostic variables are vorticity, divergence, virtual temperature, moisture and log surface pressure. The model uses the spectral technique in the horizontal direction, and second order finite difference in the vertical.

The wave number truncation used here is T126 for real time forecasts. Higher resolutions may be used for research purposes. A σ coordinate is used in the vertical. The model physics include long and shortwave radiation, boundary layer processes, large scale precipitation, shallow and deep cumulus convection.

2.2 Governing equations

The original governing equations of FSU GSM are as follows.

the vorticity equation:

$$\frac{\partial \zeta}{\partial t} = -\nabla \cdot (\zeta + f)\mathbf{V} - \mathbf{k} \cdot \nabla \times (RT\nabla q + \dot{\sigma} \frac{\partial \mathbf{V}}{\partial \sigma} - \mathbf{F}) \quad (1)$$

the divergence equation:

$$\frac{\partial D}{\partial t} = \mathbf{k} \cdot \nabla \times (\zeta + f)\mathbf{V} - \nabla \cdot (RT\nabla q + \dot{\sigma} \frac{\partial \mathbf{V}}{\partial \sigma} - \mathbf{F}) - \nabla^2(\phi + \frac{\mathbf{V} \cdot \mathbf{V}}{2}) \quad (2)$$

the thermodynamic equation:

$$\frac{\partial T_v}{\partial t} = -\nabla \cdot \mathbf{V}T_v + T_v D + \dot{\sigma} \gamma - \frac{RT_v}{C_p} (D + \frac{\partial \dot{\sigma}}{\partial \sigma}) + H_T \quad (3)$$

the continuity equation:

$$\frac{\partial \ln p_s}{\partial t} = -D - \frac{\partial \dot{\sigma}}{\partial \sigma} - \mathbf{V} \cdot \nabla \ln p_s \quad (4)$$

the hydrostatic equation:

$$\sigma \frac{\partial \phi}{\partial \sigma} = -RT_v \quad (5)$$

the moisture equation:

$$\frac{\partial S}{\partial t} = -\nabla \cdot \mathbf{V}S + SD - \dot{\sigma} \frac{\partial S}{\partial \sigma} + H_T - H_M - \left[\frac{RT}{C_p} - \frac{RT_d^2}{\epsilon L(T_d)} \right] \left[D + \frac{\partial \dot{\sigma}}{\partial \sigma} - \frac{\dot{\sigma}}{\sigma} \right] \quad (6)$$

The spectral technique in the horizontal direction was used. For details we refer to [4].

2.3 Vertical discretization

Suppose there are N vertical levels in the FSU GSM, they are defined as $\sigma = \sigma_n = (2n - 1)/2N$, $1 \leq n \leq N$. $\sigma_{N+1} = 1$. The variables U, V, ζ, D, ϕ and P are carried on these levels. The vorticity equation and divergence equation are applied at these levels. q, ϕ_s and W_s are carried on σ_{N+1} level.

The variables $\tilde{T}, \tilde{S}, \tilde{\sigma}$ and \tilde{W} are defined on levels intermediate to the levels of the geopotential, which are defined as $\tilde{\sigma}_n = \sqrt{\sigma_n \sigma_{n+1}}$. The tilde notation is used to denote the variables which are carried in the layers ($\sigma = \tilde{\sigma}_n$).

It is convenient to define $d_n = \ln(\sigma_{n+1}/\sigma_n)$ and vertical increments $\tilde{\delta}_n = \tilde{\sigma}_{n+1} - \tilde{\sigma}_n$, $1 \leq n \leq N - 2$. For the top and bottom increments $\tilde{\delta}_0 = \tilde{\sigma}_1$ and $\tilde{\delta}_{N-1} = 1 - \tilde{\sigma}_{N-1}$.

In the program, we use the spectrum coefficients $\{\xi, D, P, q, S\}_l^m$ of the grid variables as the basic variables.

2.4 Parallelization of the FSU GSM

In this section we begin by first discussing first the organization of the FSU global spectral model. Initially, the spectral prognostic variables and their derivatives are transformed to the Gaussian grid, one latitude at a time. One routine computes the nonlinear dynamical tendencies for each grid point, while another routine computes the physical tendencies. Each of these routines are designed to operate on a single latitude at a time. The main physics routine calls a number of other routines which may act either on a single vertical column, or on a latitude band. When the tendencies have been computed, they are spectrally analyzed, again one latitude band at a time. These tendencies are accumulated from one band to the next. The spectral and grid point calculations are thus done in one large loop across the latitudes. Once the spectral tendencies

have been summed up, they are then used by the semi-implicit algorithm to obtain the prognostic spectral variables at the next time step. The process is then repeated until the forecast is complete.

Since the Legendre transform is rather slow, it can easily be done one latitude at a time. Given the band structure of the model, and that the calculations for each band are independent of each other, it is straightforward to parallelize the model by dividing the latitude bands across processors. Once the calculations are carried out for all processors, a reduction is done to obtain the total spectral global tendencies. The semi-implicit scheme is fast, so there is no need to implement it in parallel.

After the main latitude loop, a reduction is done by the master task, and the semi-implicit scheme is solved on a single processor. To achieve optimal load balance, any latitude band can be assigned to any processor. The main latitude loop is consecutively partitioned across logical latitudes, each of which may be mapped to any physical latitude by a mapping function.

Parallelization on distributed memory architecture is done by running a copy of the model on each processor and using Message Passing Interface (MPI) for communication. Each processor only operates on its portion of the latitude bands, which is determined at the beginning of the program execution. When all processors have completed their latitude calculations, a reduction is done using the MPI allreduce routine. Thus all processors have a copy of the total spectral tendencies, and each solves the semi-implicit scheme to obtain the global spectral variables at the next time step. While this computation is redundant, it is fast and minimizes communication.

3 Variational data assimilation problem

The cost function for 4-D variational data assimilation will assume the form :

$$J(\mathbf{X}_0) = \frac{1}{2} \sum_{r=0}^R (\mathbf{H}\mathbf{X}(t_r) - \mathbf{X}^{obs}(t_r))^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{X}(t_r) - \mathbf{X}^{obs}(t_r)) \quad (7)$$

where \mathbf{X}_0 is the control variable, a vector of dimension N representing the initial state of the model; $\mathbf{X}(t_r)$ is a vector of dimension N containing all the model variables; $\mathbf{X}^{obs}(t_r)$ represents observational data used for the assimilation purposes; \mathbf{H} is a transformation matrix that maps the model variables to the observations; \mathbf{R} is an $N \times N$ diagonal matrix which represents the covariance matrix of the observation errors. The values of the elements are usually determined by the dimensional scaling of various variables, relative importance and quality of the data set and other considerations.

3.1 Background error term

One of the main issues in data assimilation is the specification of the background error covariance matrix. In recent years, a number of research efforts was dedicated to the study of what is known as the background error denoted by J_b term. That is, the term given

$$J_b(\mathbf{X}) = (\mathbf{X} - \mathbf{X}_b)^T \mathbf{B}^{-1} (\mathbf{X} - \mathbf{X}_b).$$

Zou *et al.*[3] shows that for linear case, it can be proved that the minimization process with incomplete observation can still guarantee uniqueness in some cases, while for non-linear case, there is no theorem to guarantee the uniqueness, and experiments show that the minimization fails for the incomplete observation in space dimension.

In this case, the cost function J assumes the form

$$J(\mathbf{X}_0) = \underbrace{\frac{1}{2}(\mathbf{X}_0 - \mathbf{X}_b)^T \mathbf{B}^{-1} (\mathbf{X}_0 - \mathbf{X}_b)}_{J_b} + \underbrace{\frac{1}{2} \sum_{r=0}^R (\mathbf{H}\mathbf{X}(t_r) - \mathbf{X}^{obs}(t_r))^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{X}(t_r) - \mathbf{X}^{obs}(t_r))}_{J_o} \quad (8)$$

The first term on the RHS of the equation is the background term, $\mathbf{X}_0 - \mathbf{X}_b$ represents the departures of the model variables at the start of the analysis from the background field \mathbf{X}_b . \mathbf{B} is an approximation to the covariance matrix of background error.

With the background term, we can prove the uniqueness of the solution of minimizations process with incomplete observation for linear case(see Appendix).

Zou *et al.*(1992)[3] proved that for some particular case, one can derive sufficient conditions for the uniqueness of the solution. However in the general case, there is no general rule to provide for sufficient conditions.

With background term, no matter which form to give the projection operator H assumes, we still can guarantee the uniqueness of the solution.

For the nonlinear case, one can not derive sufficient conditions guaranteeing uniqueness of the solution.

3.2 Construction of background covariance matrix

It is not possible to completely specify the background error covariance matrix, \mathbf{B} , in full and various approximations are necessary due to its high dimensionality. Some of these assumptions are:

Separability Assume that $[\mathbf{B}]$ can be simplified into horizontal and vertical parts. $[\mathbf{B}] = [\mathbf{B}^{Horizontal}\mathbf{B}^{Vertical}]$.

Homogeneity The structure is spatially homogeneous if the error depends only on the relative displacement rather than on the absolute locations. In this case the diagonal elements of $[\mathbf{B}]$ are all equal.

Stationarity This is the temporal analogue of spatial homogeneity. $[\mathbf{B}]$ is only a function of the time difference and not the absolute times.

Isotropy For a single observation, the correlation with other grid points is independent of the direction, only the distance between them.

The problem of using the background matrix \mathbf{B} directly is that the inverse of \mathbf{B} cannot be calculated due to its huge dimensions. Instead we define a simple form of \mathbf{B}^{-1} . Based on the above assumption, following Bennett(2002), define:

$$\mathbf{B}^{-1} \approx w_0 I + w_2 \nabla^2 \nabla^2 \quad (9)$$

where w_0 and w_2 are chosen so that the power spectrum is similar to that of the Gaussian correlation function(Daley,1991[1]:

$$\mathbf{B}_{ij} = e^{-\frac{1}{2} \left(\frac{x_i - x_j}{l} \right)^2} \quad (10)$$

where $x_i - x_j$ is the distance between grid points and l is the correlation length scale.

Figure 1 shows distributions of a column of the covariance matrix \mathbf{B} and a relevant part of the inverse matrix of \mathbf{B}^{-1} along latitude -45° , the relative point is at longitude 90°). It is seen that their distribution is very similar.

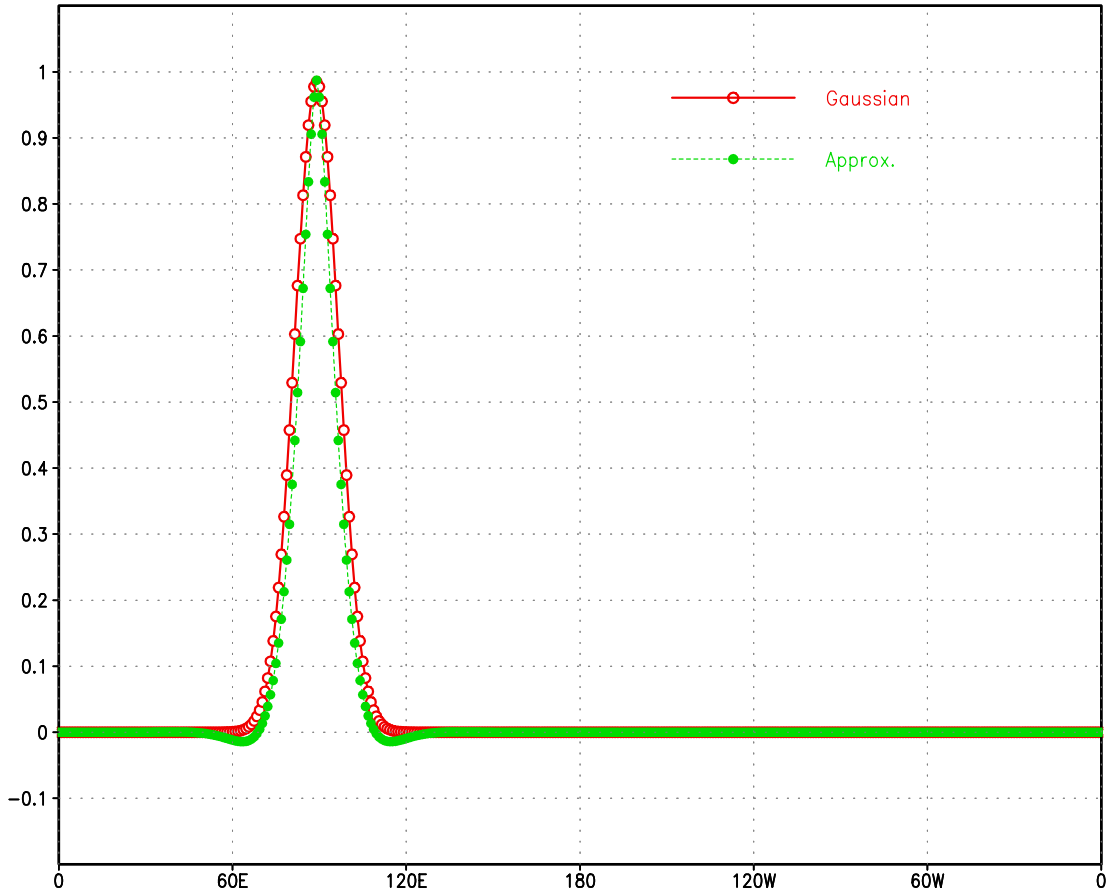


Figure 1: Distribution of a column of \mathbf{B} and a relevant part of the inverse matrix of \mathbf{B}^{-1}

4 Numerical experiments with incomplete observations

4.1 Description of the experiments

A twin experiment was carried out using the dynamic core of FSU global spectral model and its adjoint model with complete observations which served as a control run.

The data assimilation window was set to 6 hours from 06UTC Oct. 2, 2002 to 12UTC Oct. 2, 2002.

The observation data were generated by integrating the forward model 6 hours from the initialized ECMWF analysis at 06UTC Oct. 2, 2002, and were available at every timestep and each

Gaussian grid point.

The initial guess of the initial condition was taken from the initialized ECMWF analysis at 00UTC Oct. 2, 2002, which is 6 hours prior to the initial time level.

A suitable choice of the background state \mathbf{X}_b is important, since it will strongly affect the retrieved initial data. In our experiments, the data which were generated by integrating the forward model 6 hours at the initial time 00UTC Oct. 2, 2002, when is 6 hour prior the initialtime of data assimilation window.

The limited-memory quasi-Newton method of Liu and Nocedal(1989)(L-BFGS) [14] was used to carry out the unconstrained minimization process to obtain ti he optimal initial data.

Then, we reduced the number of observations in the space and the time dimension respectively and carried out the minimization process related to 4-D Var data assimilation.

We first reduced the number of observations in the space dimension, namely observation were available only every 2, 4 or 8 grid points respectively.

In order to investigate the impact of incomplete observations over data void areas, *i.e.* observation data was subtracted over such areas, we carried out an experiment on incomplete observations over the ocean where data were missing, specially, the observations over all grid points located over oceans of South hemisphere were subtracted.

In another experiment we reduced the observation in time dimension to be available only every 2, 4 or 8 timesteps. The case of observation being available only the initial time and at the end of the window of data assimilation were also investigated.

The incomplete observation experiments were divided to two parts, one with the cost function including the background term, the other without the background term, in order to investigate and assess the impact of background term on the incomplete observation of 4D Var data assimilation.

4.2 Incomplete observations experiments without background term

Figure 2 shows the decrease of the cost function for the minimization with both complete and incomplete observations in the space dimension versus number of minimization iterations. we observe that the decrease in cost functional in the cases of incomplete observations in space dimension is slowed down and decrease by roughly 3 orders of magnitude and could not be decreases any further, while in the case of complete observations, the decrease in the cost functional was by about 4.5 orders of magnitude and could decrease further with additional minimization iterations. The figure also shows that the density of incomplete had only a minor impact in the minimization process, *i.e.* there were only slight differences as the numbers of grid points where observations were available was varied from every 2 to every 8 grid points.

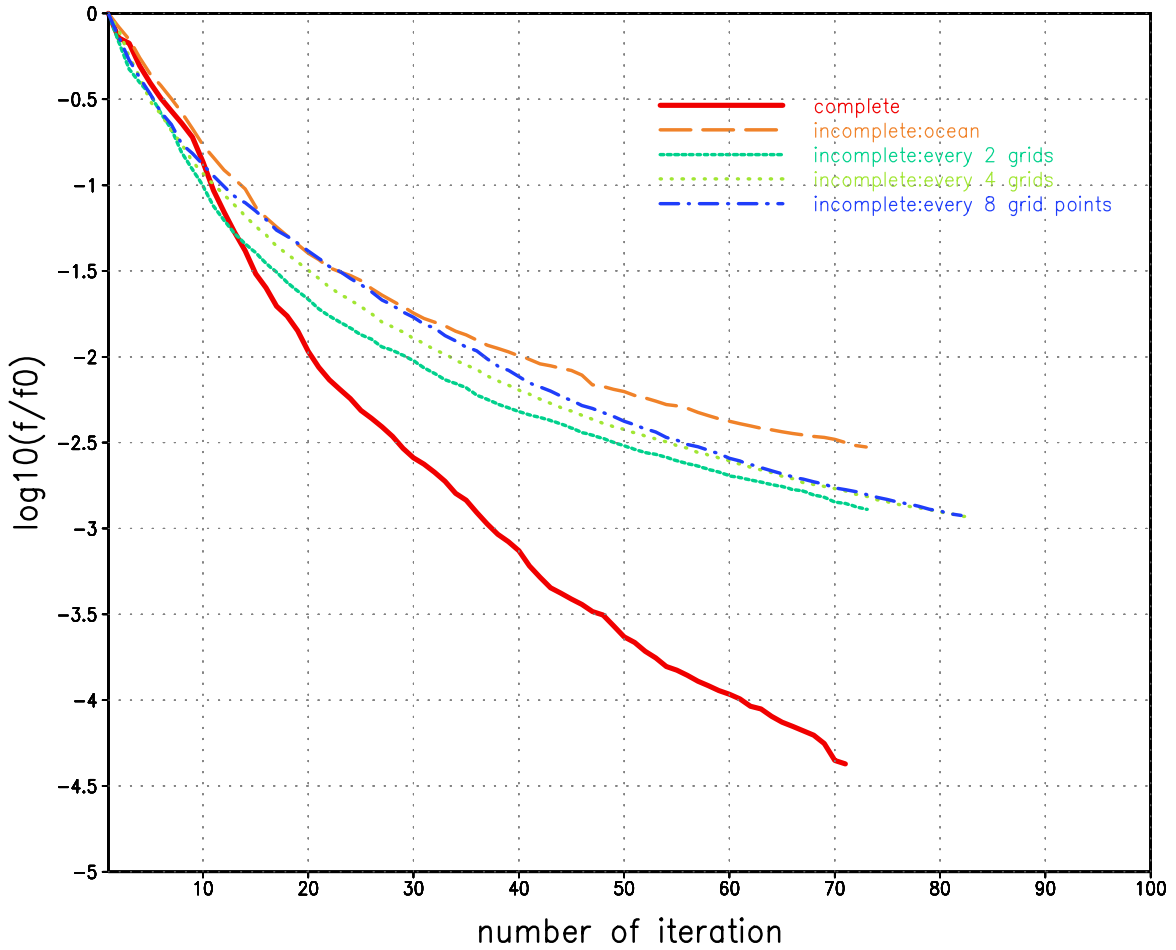


Figure 2: Normalized cost function for the incomplete observation in space versus number of iterations

Figure 3 shows the decrease of the normalized gradient of cost functional versus number of minimization iterations. It also shows that the decrease rate of the cost functional for the case of incomplete observations in space dimension displays a slowdown and cannot decrease beyond a certain threshold, while that for the case of complete observations continues to decrease as the number of minimization iterations increase.

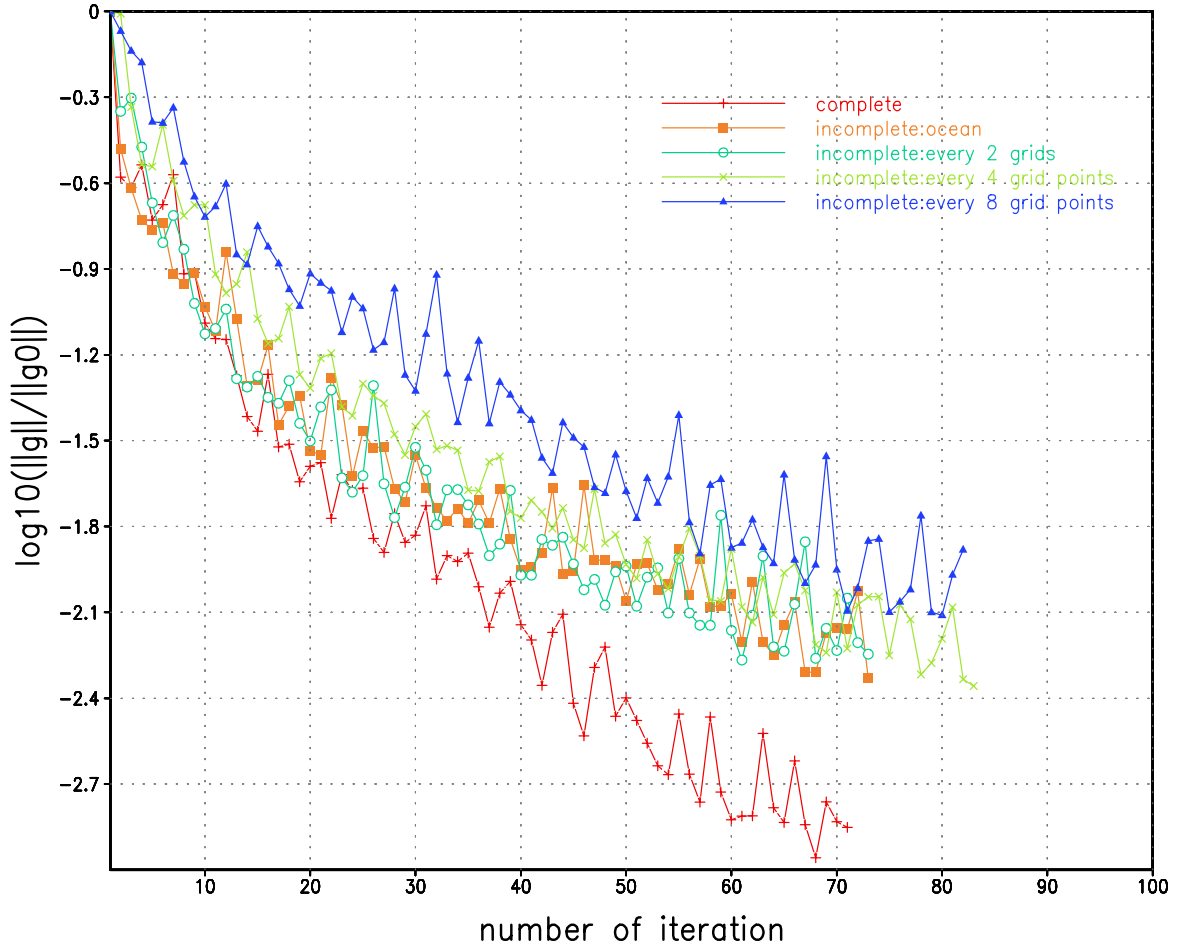


Figure 3: L_2 -norm of normalized gradient of the cost functional for the incomplete observation in space dimension vs. number of iteration

For comparison, we also plotted the height field at 500 hPa and the difference between height fields at 500 hPa corresponding to the model generated observation and the runs by perturbed initial data after 12 hours of model integration.

Figure 4 shows the height field of model generated observations after 12 hours of integration from initial time, Figure 5 shows the height field after 12 hour of integration from the retrieved initial data for the case of complete observation.

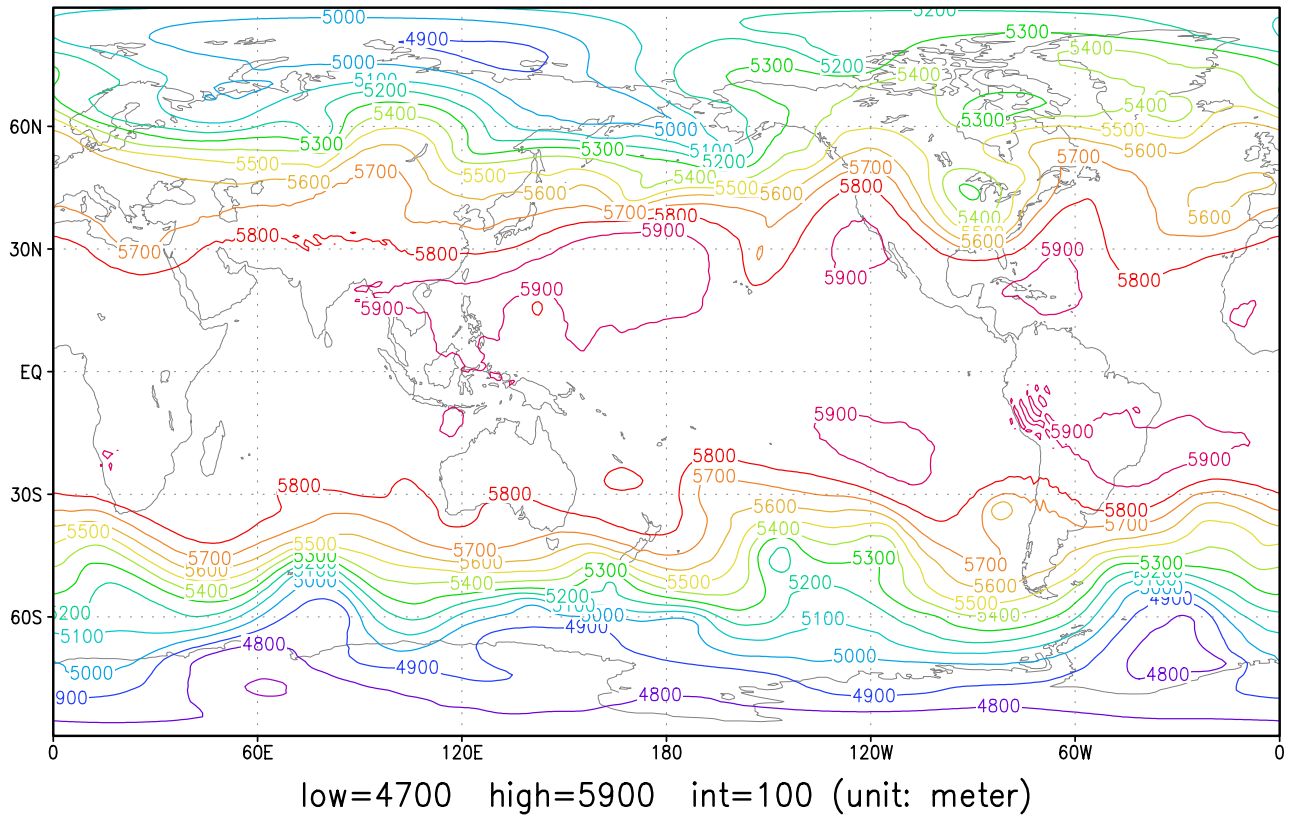


Figure 4: Height field of model generated observations at 500 hPa after 12 hours from initial time

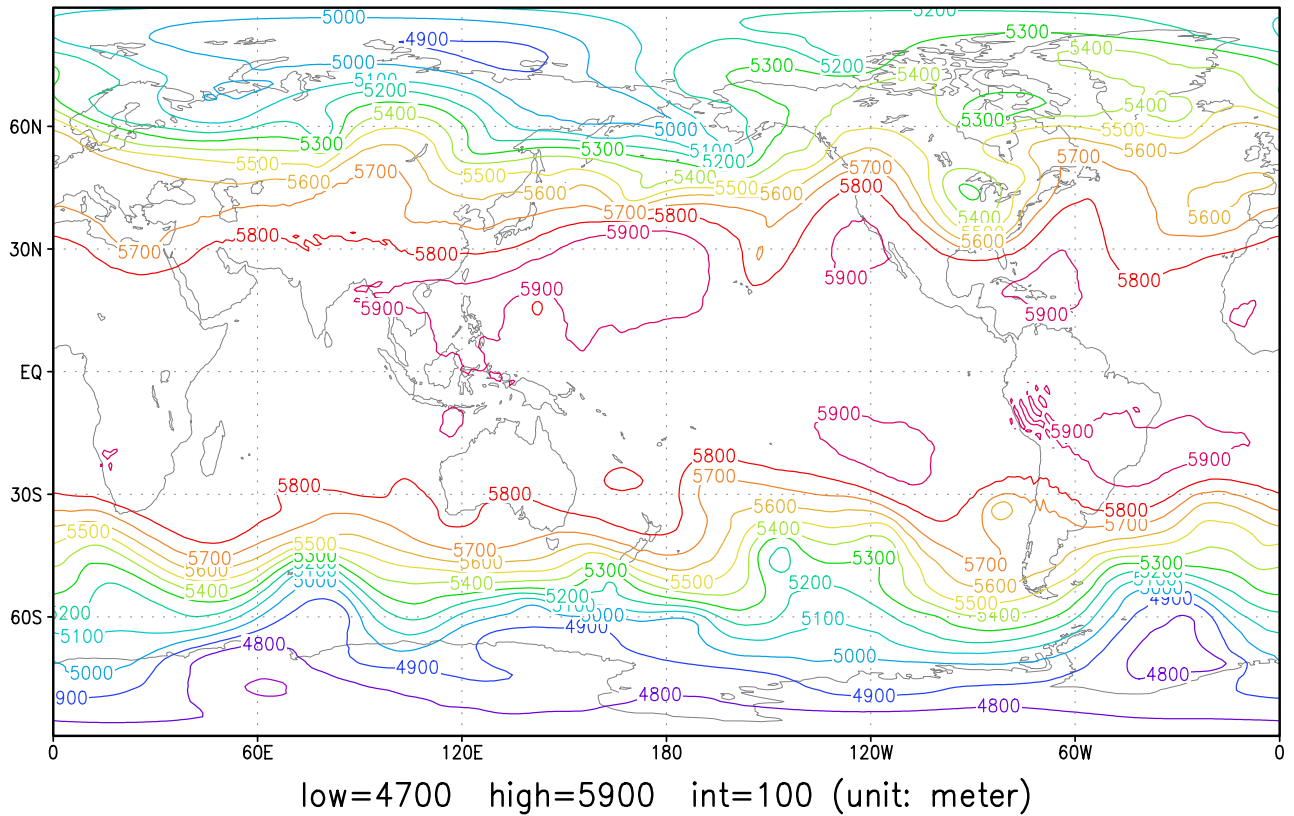


Figure 5: height field at 500hPa after 12 hour integration starting from retrieved initial data with complete observations

Figure 6 shows the difference between the height fields at 500 hPa corresponding to the runs by using the unperturbed and perturbed initial data after 12 hours integration. For the perturbed initial data, the forecast field follows a different trajectory.

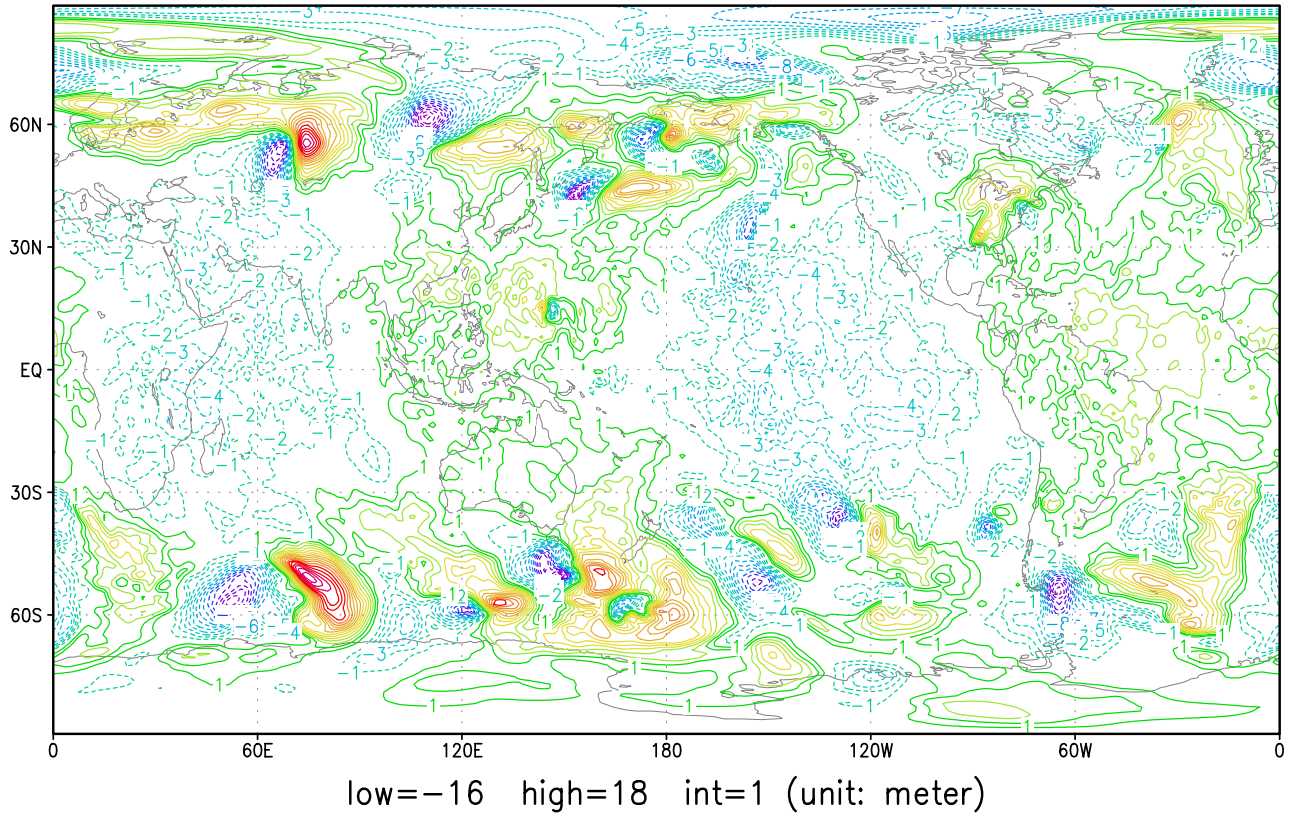


Figure 6: Difference between the height fields at 500 hPa corresponding to the runs by using unperturbed and perturbed initial data after 12 hours integration

When we employed the retrieved initial data with complete observations, the forecast fields follow to the same trajectory as observation, the difference between the height fields at 500 hPa after 12 hours integration was reduced from 18m to below 1m after 40 minimization iterations(see Figure 7).

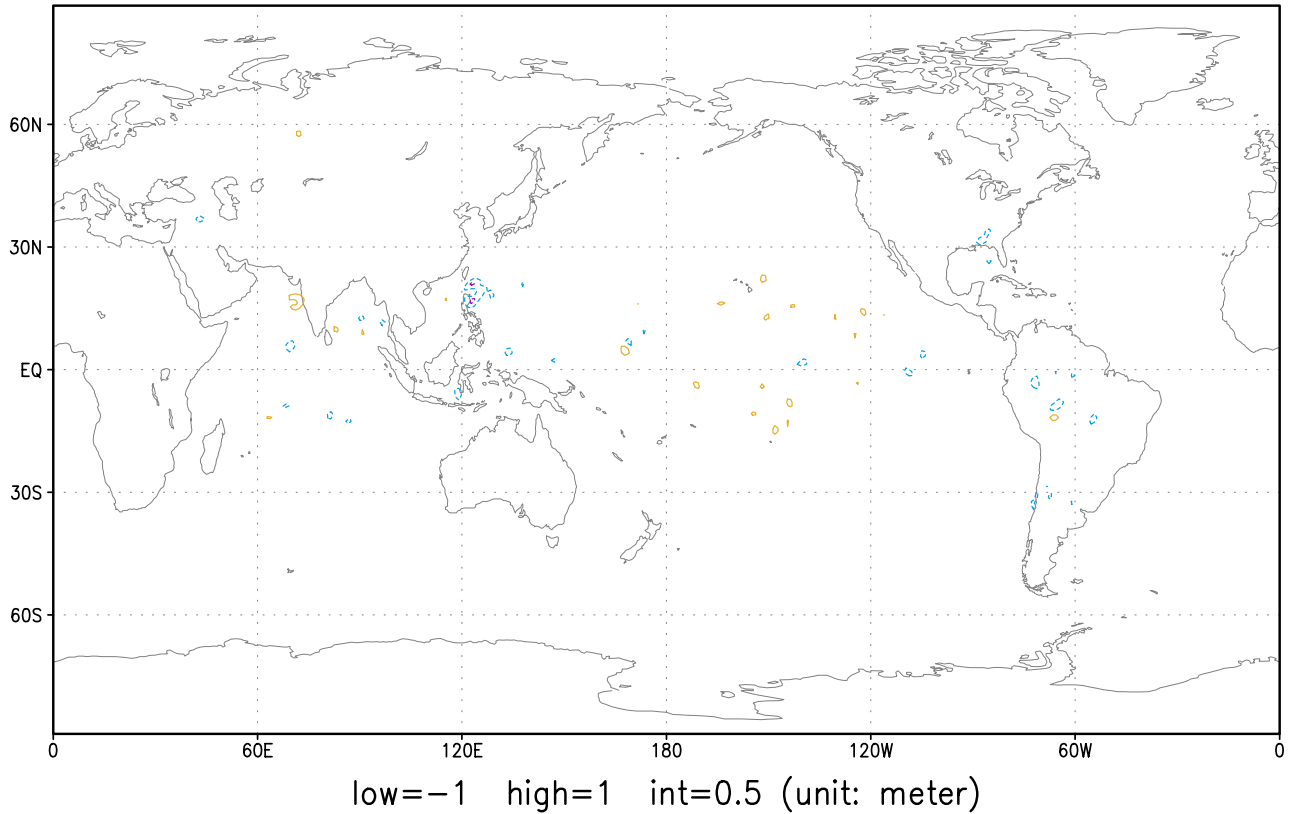


Figure 7: Difference between the height fields at 500 hPa corresponding to the runs by using unperturbed and retrieved initial data after 12 hours integration

The forecast fields follow a different trajectory when the initial data are obtained by the minimization with incomplete observations in space dimension. Figure 8 shows the difference in the height fields at 500 hPa between unperturbed initial conditions and initial data obtained by incomplete observations available at every 2 grid points. Although the errors decrease in the minimization process, the difference cannot be reduced beyond a given threshold level after 40 minimization iterations.

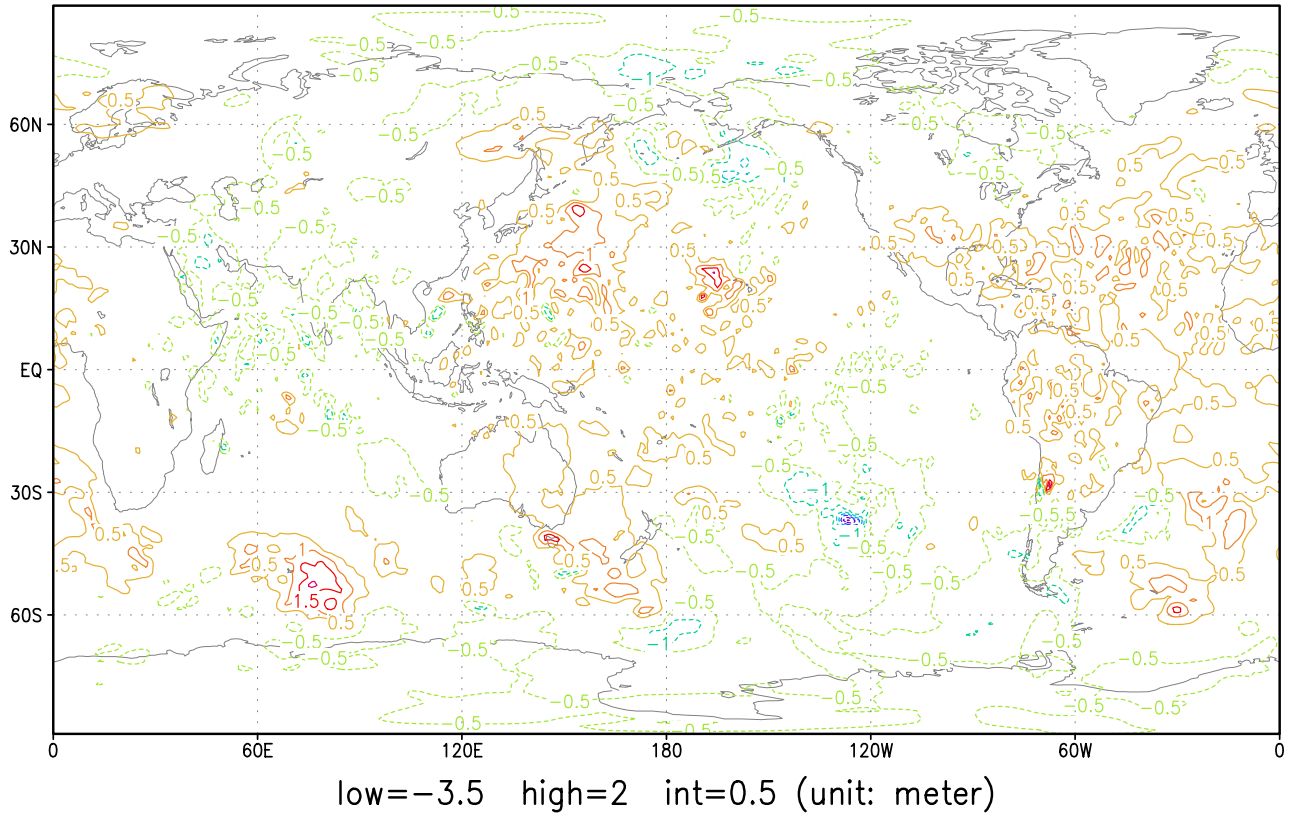


Figure 8: Difference between the height fields at 500 hPa corresponding the runs by using unperturbed initial conditions and the initial data obtained by incomplete observations available at every 2 grid points after 12 hours integration

For incomplete observations available at every 4 grid points, the difference cannot be reduced beyond same threshold even after 100 minimization iterations.(see Figure 9).

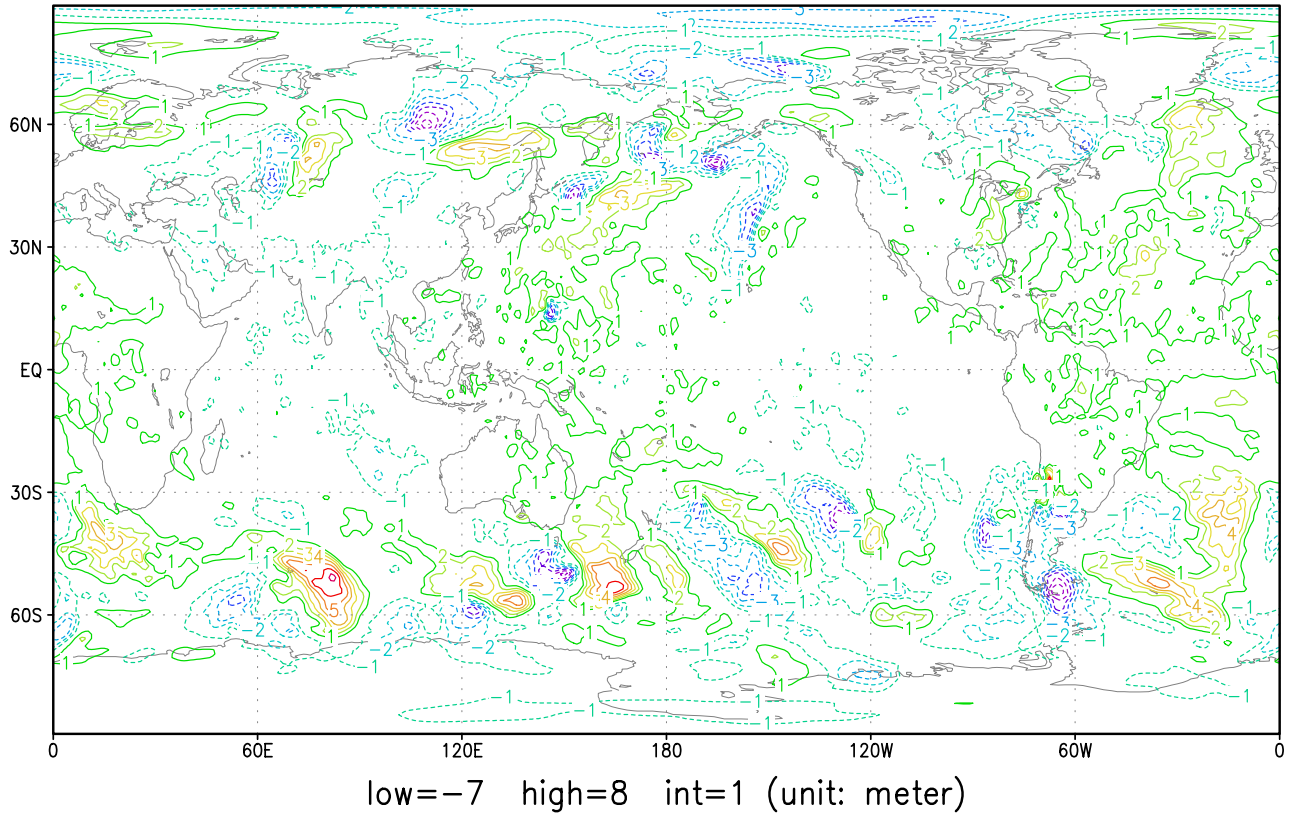


Figure 9: Difference between the height fields at 500 hPa corresponding to the runs by using unperturbed initial conditions and initial data obtained by incomplete observations available at every 4 grid points after 12 hours integration

The results are even worse for the case when the observations over ocean are missing. Figure 10 shows the difference between the height fields at 500 hPa corresponding the runs by using the unperturbed initial conditions and the initial data obtained by the minimization of incomplete observations over ocean after 12 hours integration. It is clearly seen that the lack of fit over ocean cannot beyond a given threshold.

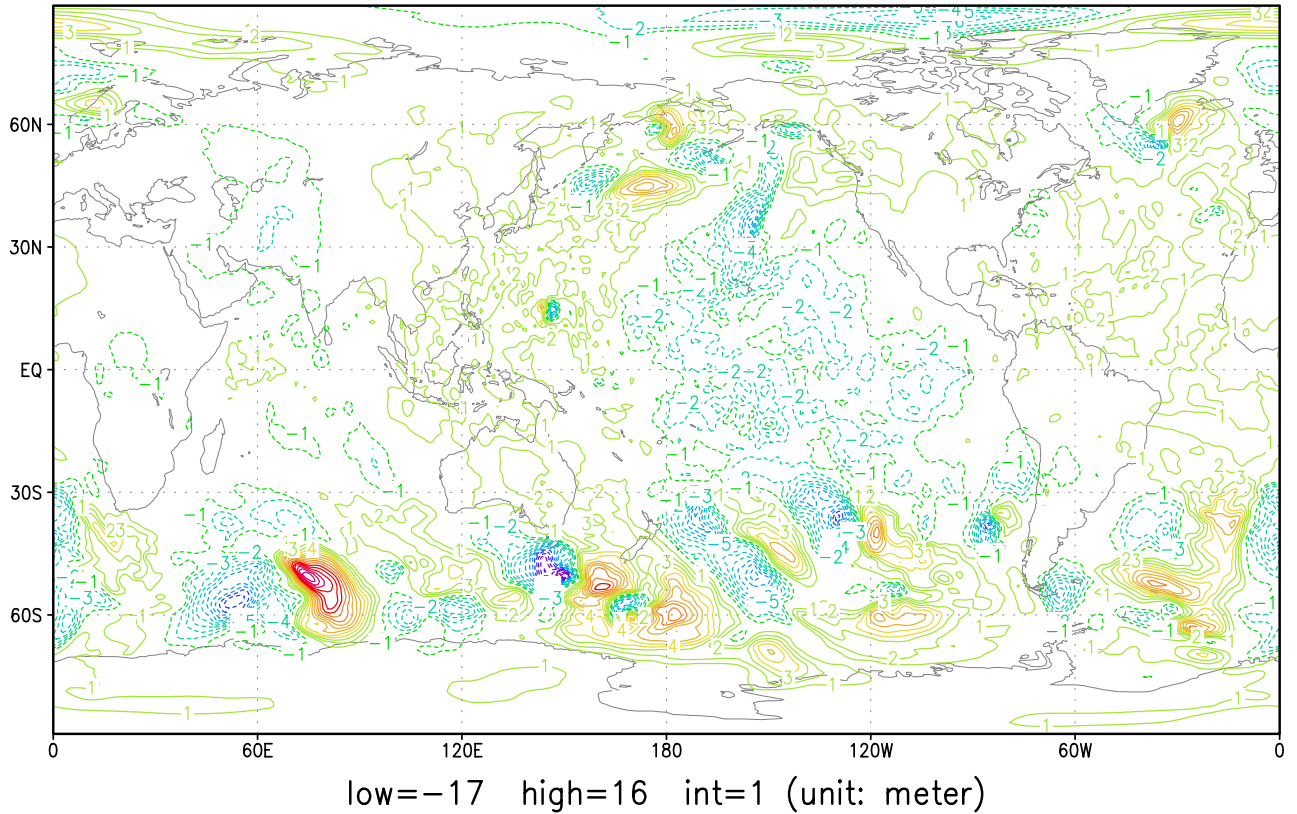


Figure 10: Difference between the height fields at 500 hPa corresponding to the runs by using unperturbed initial conditions and the initial data obtained by incomplete observations missing over the location of ocean after 12 hours integration

The variations of the normalized cost function and the L_2 norm of the normalized gradient of cost functional can also be seen in Figures 2 and 3 respectively.

In contrast to the incomplete observations in space dimension, incomplete observations in time dimension can still retrieve the initial data.

Figures 11 and 12 shows the decrease of the log of the normalized cost function and L_2 norm of log of gradient of cost functional with incomplete observations available only at every 2, 4, 8 time steps respectively.

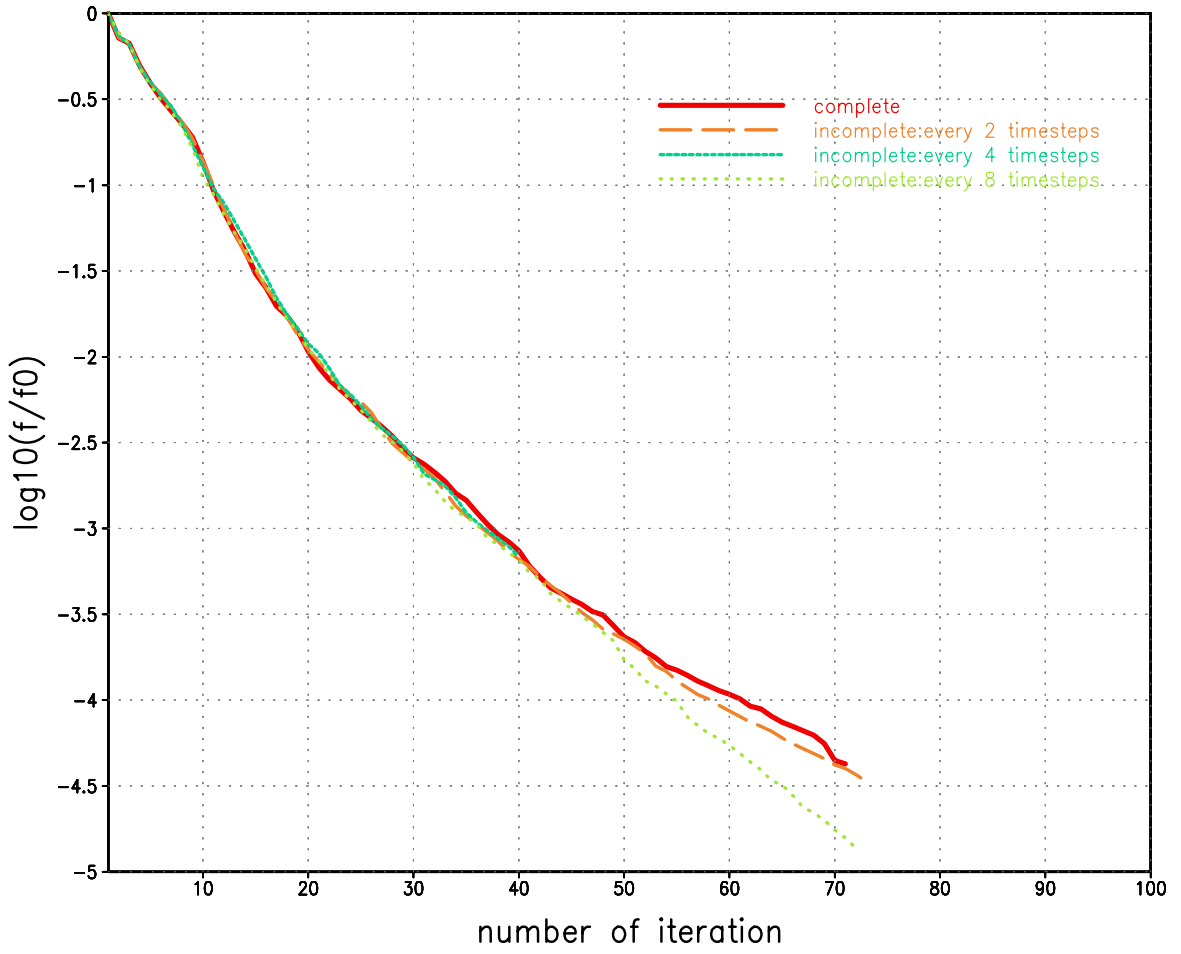


Figure 11: Log of the normalized cost function for the incomplete observation in time vs the number of minimization iterations

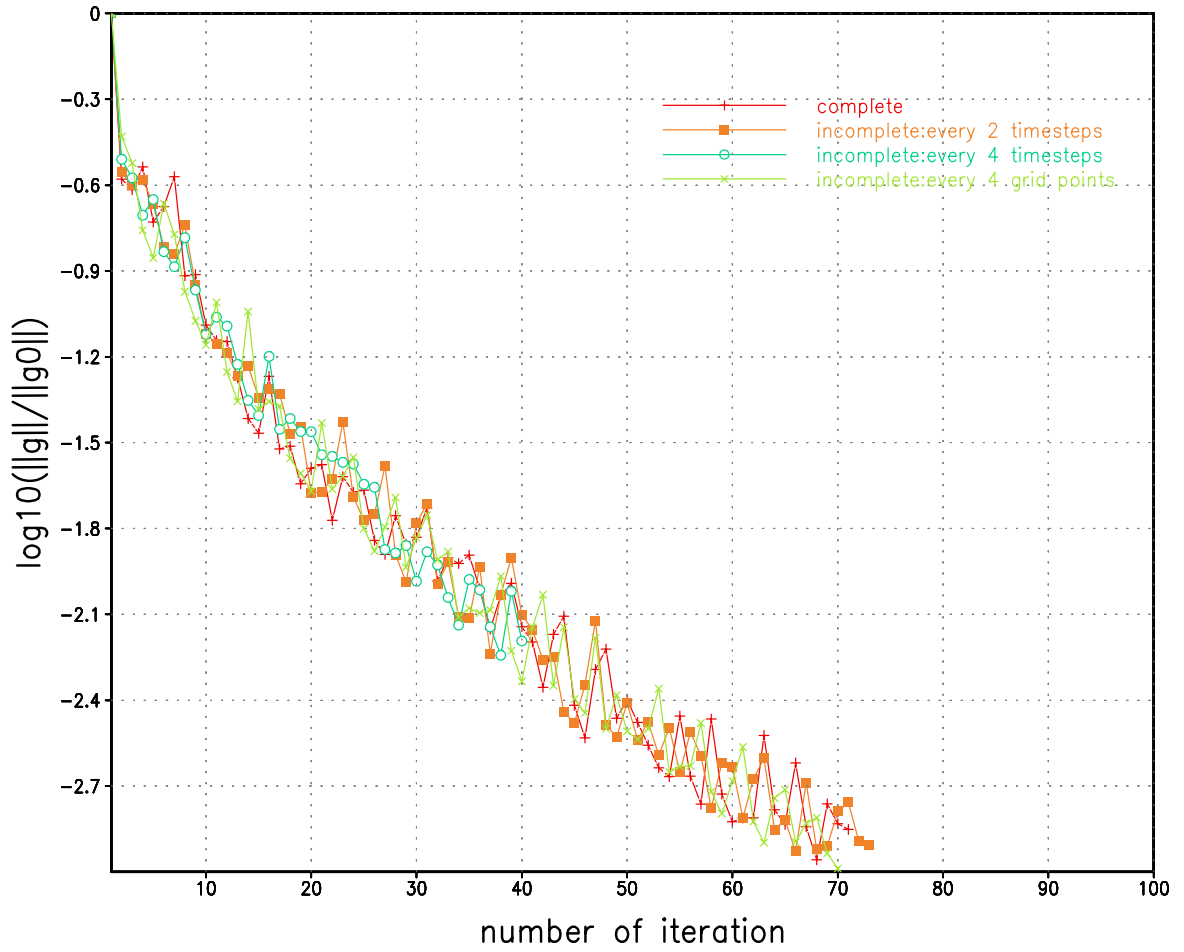


Figure 12: L_2 norm of log of Gradient of the cost functional for the incomplete observation in time vs the number of minimization iterations

It is seen that a good convergence is obtained and that the minimization of the cost functional allows a good retrieval of real initial data(see Figure 13).

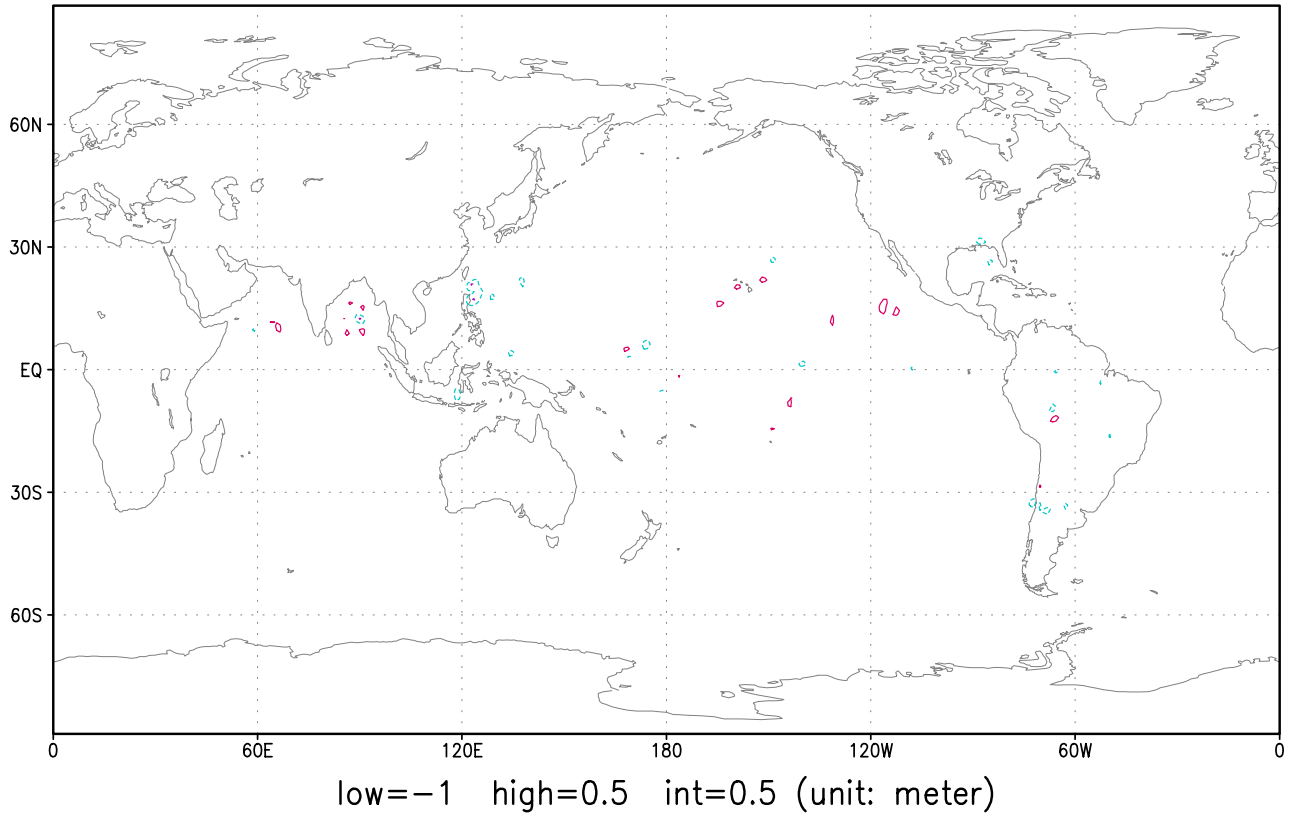


Figure 13: Difference between the height fields at 500 hPa corresponding to unperturbed initial conditions and the initial data obtained by incomplete observations available at every 2 time-step after 12 hours integration

Figure 14 provides the RMS errors of height field at 500 hPa, calculated between fields of model generated observations and the one obtained by integrating the optimized initial data with incomplete observation after 40 minimization iterations. It shows that for incomplete observations in space dimension, the error reduction obtained by the minimization process with incomplete observations depends on the density of the observations in space dimension. For observations available at every 2 grid points, though the decrease rate of the cost function is slowed down, it still can retrieve the initial data to certain degree, while for observations available only at every 4 or 8 grid points, the errors increase to a unacceptable degree. The sparser the density of

the observations, the larger the departure from the observation is obtained with minimization of the incomplete observation, *i.e.* observations available at every 2,4 or 8 grid points respectively. This is more evident in the case where void area is concerned, *e.g.*, over the Southern hemisphere ocean. In this case practically was no retrieved initial data. For incomplete observation in the time dimension, the retrieved initial data is almost identical to the one obtained by the minimization process with complete observation.

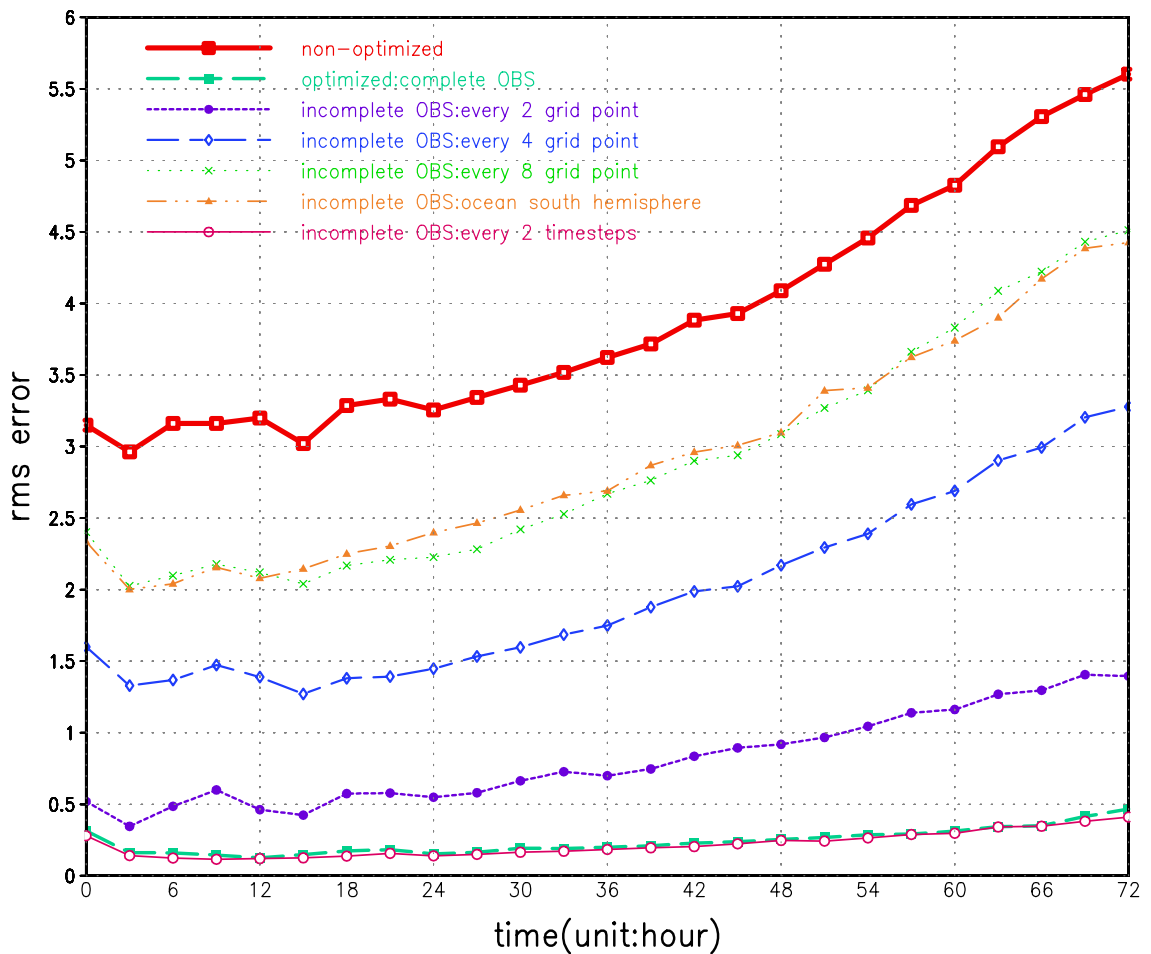


Figure 14: Time evolution of RMS of the height field at 500 hPa for the incomplete observation

There was no significant change when we increased the the number of minimization iteration steps. Figure 15 provides the same figure as figure 14 but the number of minimization iteration

step is 100. It shows only a slight amount of error reduction for the case of incomplete observation in space dimension. But the same conclusion remains. It also can be seen, for either the complete observations case or for the incomplete observation in time dimension, that the RMS errors retain a low value even after a 72 hours forecast integration (below 0.5 for iteration step 40, below 0.25 for iteration step 100). The values of RMS of height field at 500 hPa remains almost flat, while for the case of incomplete observations in space dimension, the RMS errors of height field at 500 hPa are always increasing after 15 hours of model integration.

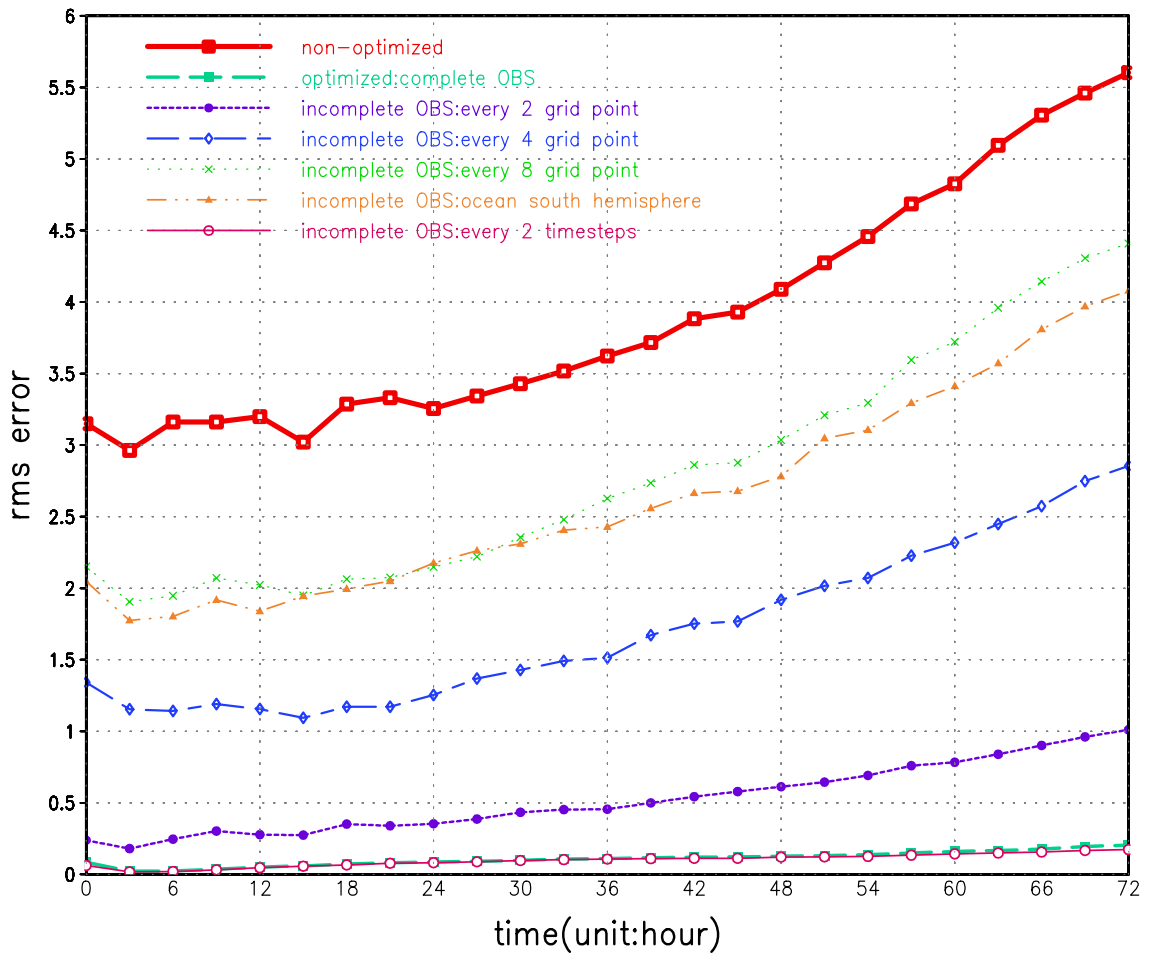


Figure 15: Time evolution of RMS of the height field at 500 hPa for the incomplete observations

4.3 Impact of background term on the experiments of incomplete observation

Since the background error term will provide some informations of the initial data, we carried out a number of experiments on the incomplete observations in space dimension with a background error term. The results show that the behavior of the minimization process with background covariance term are quite different from the one without background covariance term.

Figure 16 shows the variations of the total cost function (i.e. $J = J_b + J_o$) with complete observations vs number of iteration. In order to investigate whether the retrieved initial data were forced to the observations or the background field, we plotted both J_o and J_b . It clearly show that the J_o decreased faster than J_b , which may means, for complete observations, the retrieved initial data follows to the observations. The decrease of the total cost function J was much slower than the one without background error covariance term.

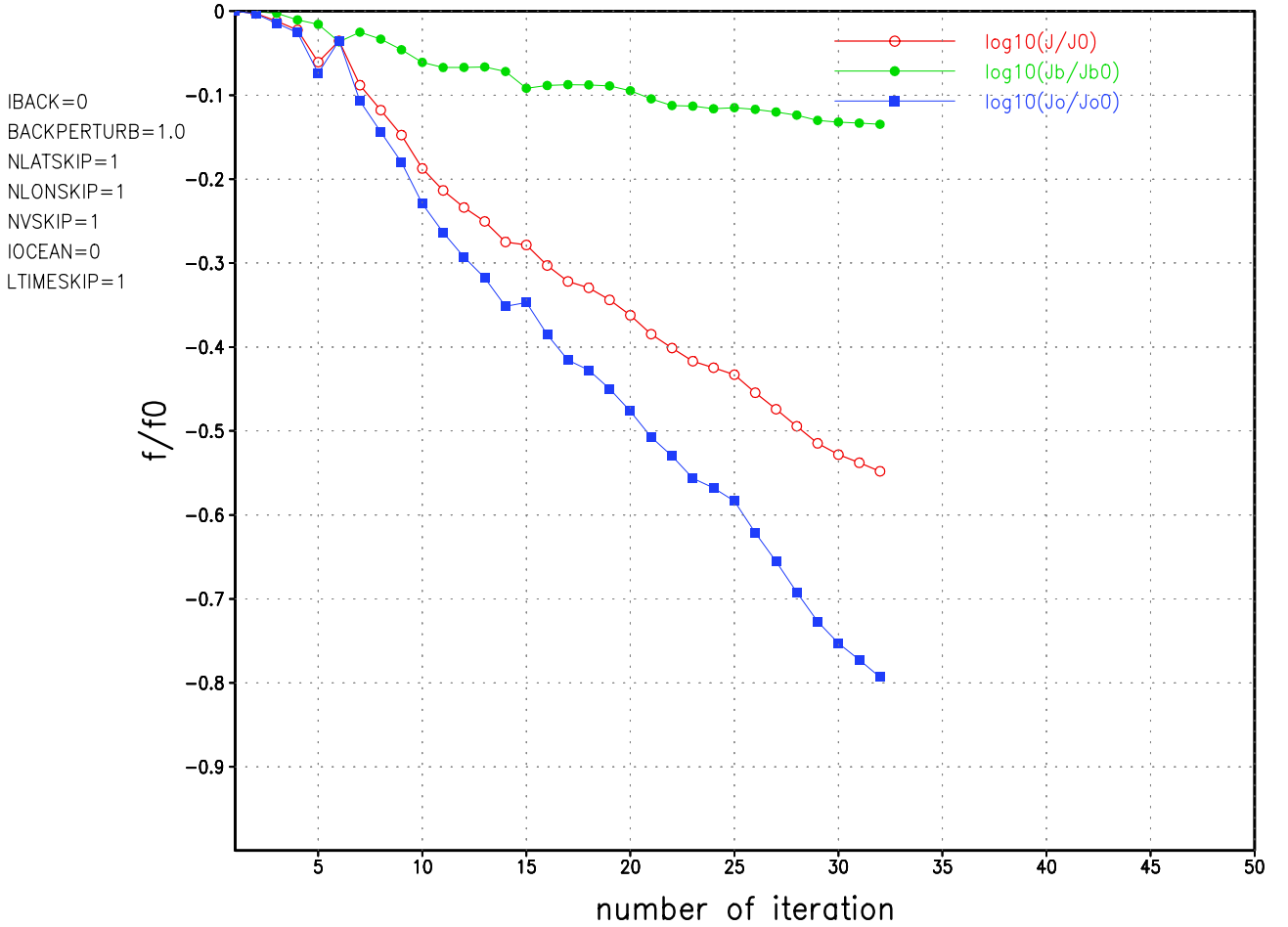


Figure 16: Variation of the total cost function J , observation cost function J_o and background cost function J_b vs number of minimization iterations

Figure 17 shows the variation of the cost function J , J_o and J_b with the observations available only every 2 grid points. The solution shows the decrease rate was not smooth as the one without background error covariance term. It is seen that the characteristic variations of J_o and J_b with the number of minimization iteration were totally reversed, *i.e.* when J_o exhibits a faster decrease, J_b exhibits slower decrease, vice versa, when the decrease of J_b exhibits a faster decrease, the decrease of J_o exhibits a slower rate. This means, for incomplete observation, during some stages of minimization process, the retrieved initial data follows the observation, while at other stages,

the retrieved initial data follows the background field. For the case of incomplete observations in space dimension where the observations are available only every 2 grid points, the retrieved initial data finally follows the background field.

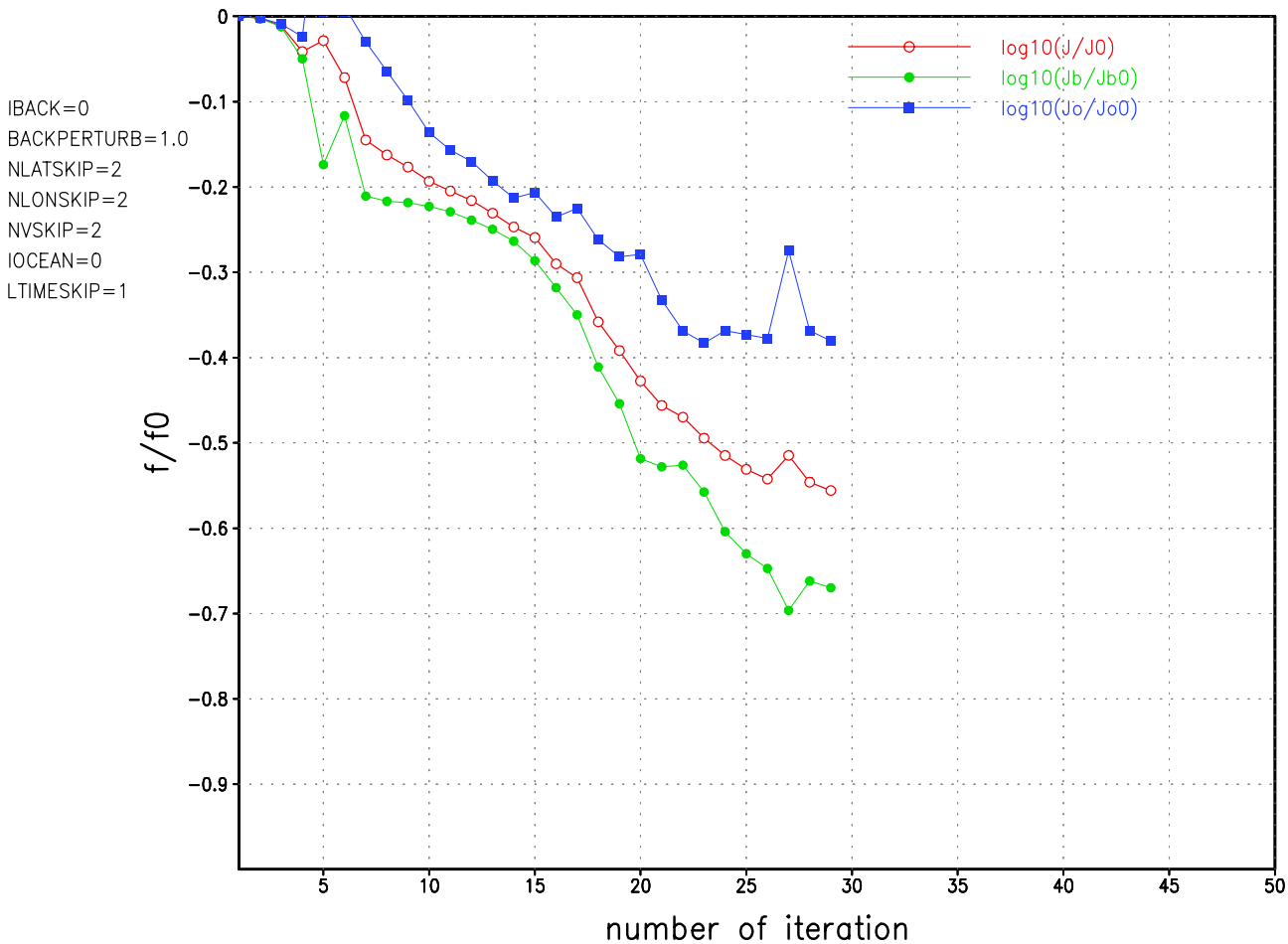


Figure 17: Variation of the log of total cost function J , observation cost function J_o and background cost function J_b with the number of minimization iterations

Figure 18 shows the variation of the cost function J , J_o and J_b with the observation available only at every 8 grid point. This time, the optimized initial data follows background field immediately, since the case of incomplete observations being available only every 8 grid point causes the term J_o the cost functional to have a lesser impact compared to J_b than for the case of incomplete

data being available every 2 grid points.

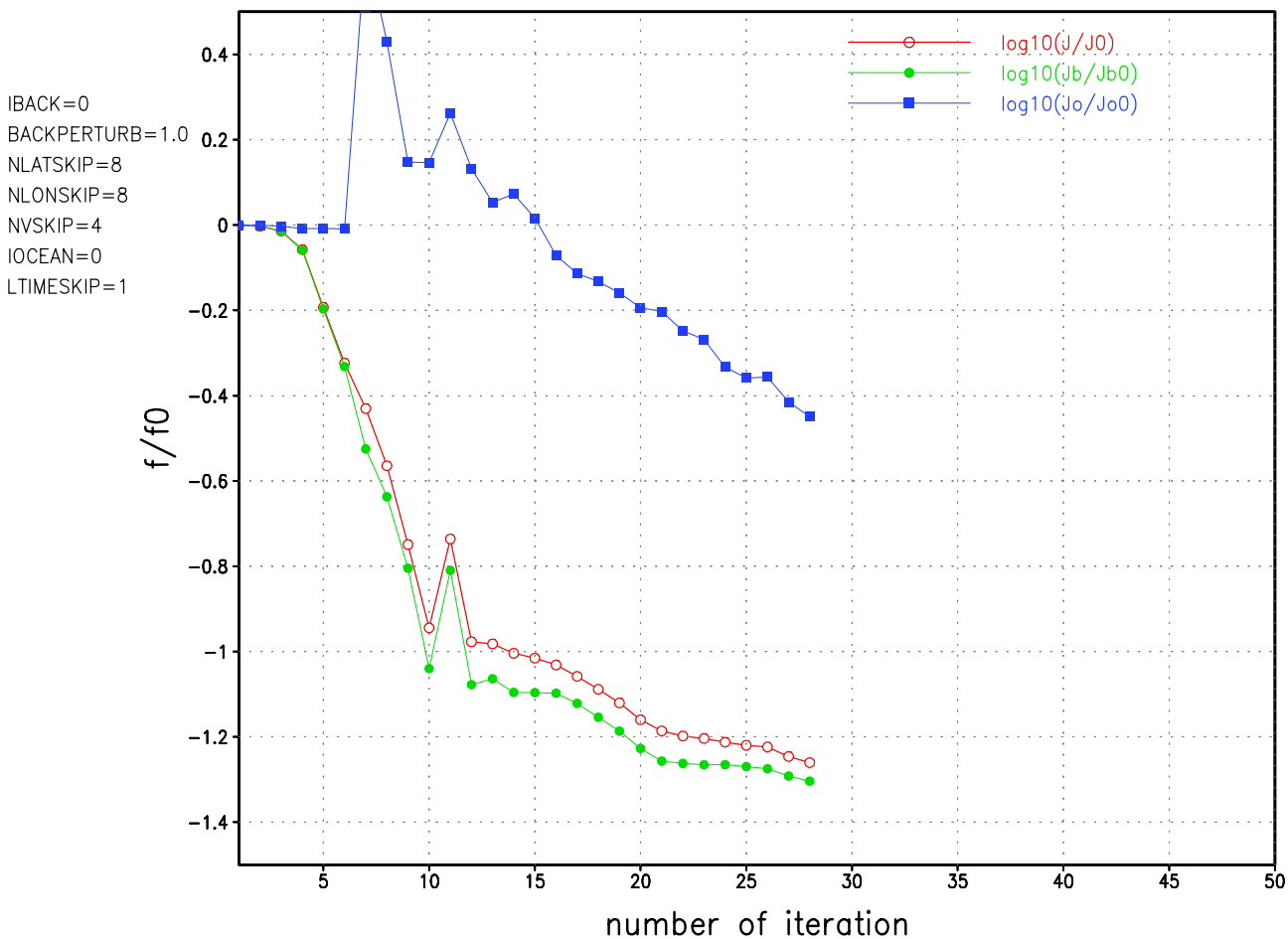


Figure 18: Variation of the total cost function J , observation cost function J_o and background cost function J_b with the number of minimization iteration

Figure 19 shows the variation of the cost function with incomplete observations where data were missing over the ocean region in the southern hemisphere. In contrast to the aforementioned cases of incomplete observations where observations were available only every 2 or 8 grid points, here the optimized initial data follows the observations for ocean data void area.

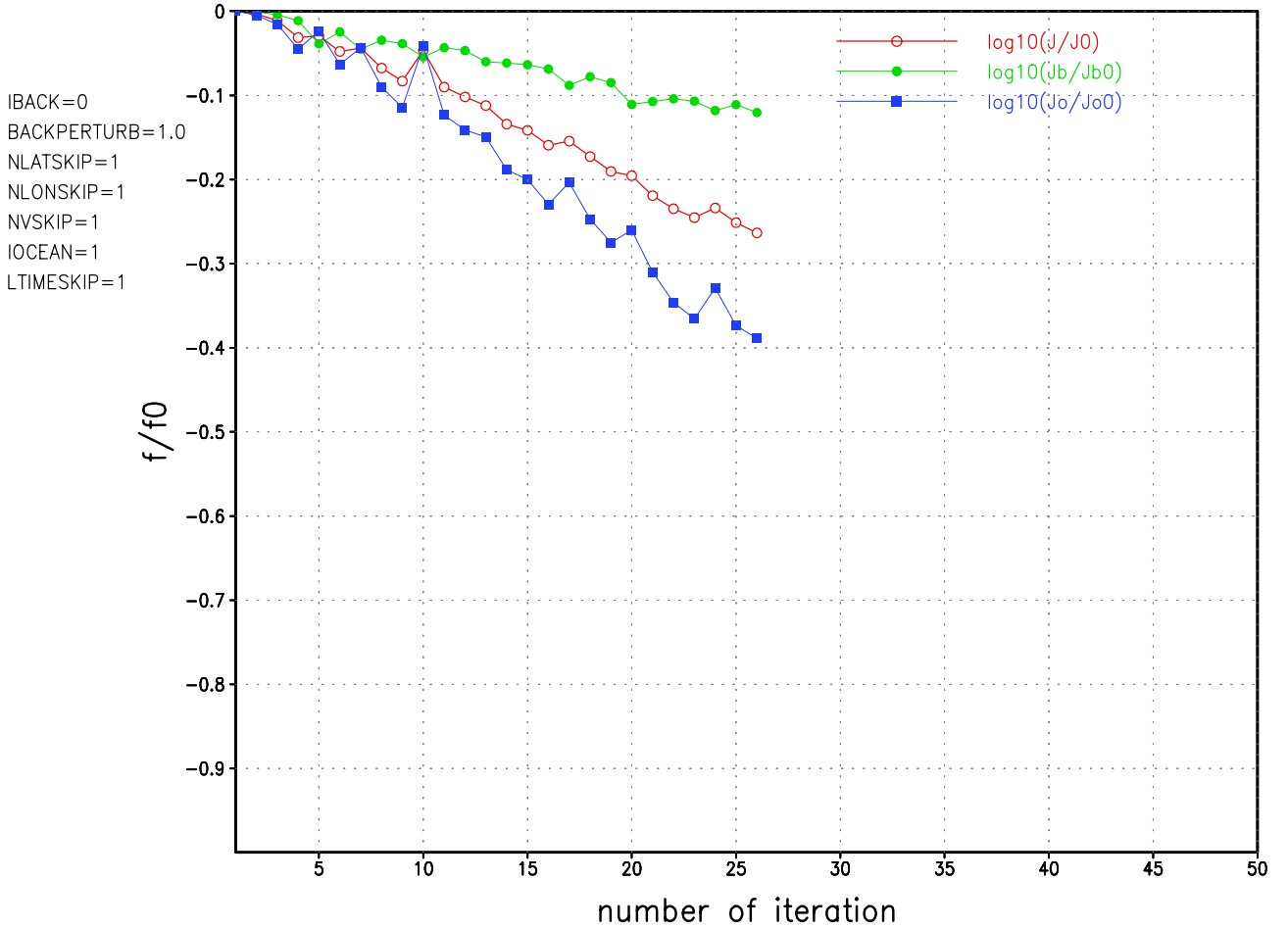


Figure 19: Variation of the total cost function J , observation cost function J_o and background cost function J_b for the case of ocean data void area with number of minimization iteration

The results are qualitatively different for the case when the observations are incomplete in time dimension.

Figure 20 shows the variation of the cost function with the observations being available only every 2 timesteps. Since the density of the observation in time is high, the retrieved initial data follows the observation. However for the case when observations are available every 4 timesteps, the retrieved initial data appears to swing from the background field to observation field. It first follows the background field then after a number of iterations it follows a trajectory that

is in between the background field and observation field, and finally ends up by following the observation field(see Figure 21).

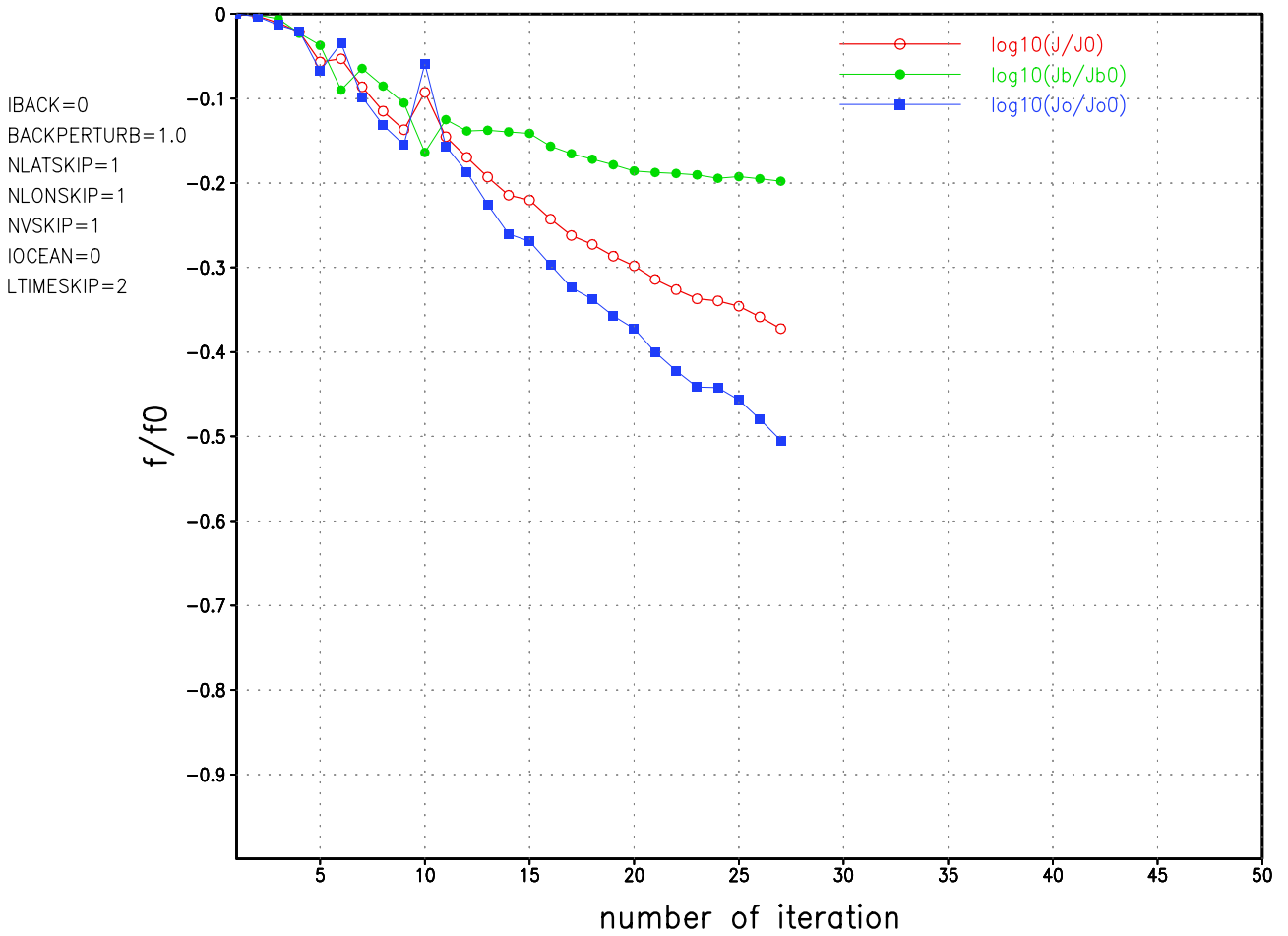


Figure 20: Variation of the total cost function J , observation cost function J_o and background cost function J_b

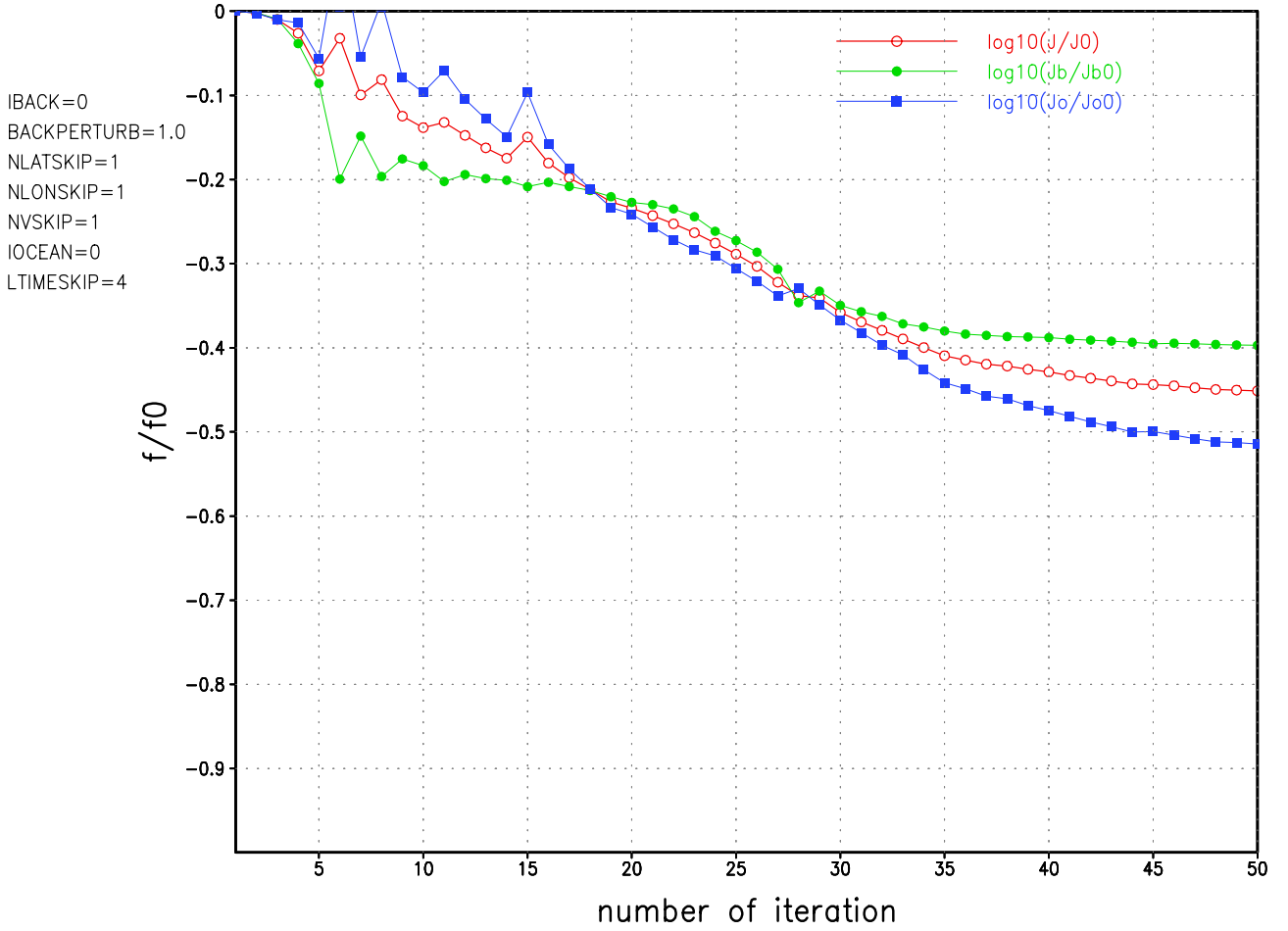


Figure 21: Variation of the total cost function J , observation cost function J_o and background cost function J_b

These results show that with the inclusion of the background error covariance term in the cost functional, the retrieved initial data strongly depend on the size of term J_b compared that of J_o . For the case of complete observations, the impact of J_o is dominant, so the retrieved initial data follows observation field. For the case of incomplete observations, the impact of J_o is reduced, when the size of J_o becomes comparable with the size of J_b , the retrieved initial data will swing from observation field to background field or *vice versa*. When the amount of available observations continues to be reduce, the impact of the J_b term becomes dominant, so the retrieved initial data

will follow the background field.

It should be pointed out, that although the retrieved initial data appears to follow either the observation or the background fields, this does not mean that the retrieved initial data will be close to either field, due to the existence of the other term (either J_o or J_b), the retrieved initial data will follow a trajectory between the trajectories determined by either the observations or the background term.

5 Summary and conclusions

Using the new MPI version of FSU Global Spectral Model, we developed tangent linear and adjoint models of the dynamical core of FSU model in a framework of MPI parallel processing computation. The correctness of tangent linear and adjoint of the model were fully tested.

The adjoint model was applied to carry out the experiments on the impact of the background error term on incomplete observations in the framework of 4-D Var data assimilation.

Results show that for the incomplete observations in the space dimension, in absence of the background term, the minimization process fails to retrieve the initial data, while for incomplete observations in time dimension or in variables, the minimization process can still retrieve the initial data.

We also carried out the experiments on the impact of background term,

Acknowledgements

The authors thank T.N. Krishnamurti who gave us useful help, thanks also go to Steven Cocke who provided the new MPI version of FSU GSM code.

Appendix

A The uniqueness of solution

In order to ensure the uniqueness of the solution of minimization of (8), we need to show that the cost functional J is strictly convex, *i.e.*, if $\forall U, V \in R^n$

$$J(\lambda U + (1 - \lambda)V) < \lambda J(U) + (1 - \lambda)J(V), \quad \lambda \in [0, 1] \quad (11)$$

If J is not strictly convex, *i.e.* the equality may exist for some U, V and λ

$$J(\lambda U + (1 - \lambda)V) = \lambda J(U) + (1 - \lambda)J(V), \quad \lambda \in [0, 1] \quad (12)$$

then the minimum is not necessarily unique and the function J may have no minima. J may also have stationary points which are not minima, but rather maxima or saddle points.

When J is second order differentiable, the convexity of cost function J is equivalent to that the Hessian matrix \mathcal{H} being positive semi-definite: $\forall X \in R^n, X^T \mathcal{H} X \geq 0$ and J is strictly convex if \mathcal{H} is positive definite, *i.e.* $\forall X \in R^n, X^T \mathcal{H} X > 0$.

Suppose we have a nonlinear forecasting model:

$$\mathbf{X}_{k+1} = \mathbf{M}_k \mathbf{X}_k \quad (13)$$

where \mathbf{M}_k is an nonlinear operator.

The tangent linear model of (13) is given by

$$\mathbf{X}'_{k+1} = \mathcal{M}_k \mathbf{X}'_k \quad (14)$$

where \mathcal{M}_k is a linear operator.

Then the solution of (13) can be written as

$$\mathbf{X}_{k+1} = \mathbf{M}_k \mathbf{X}_k = \mathbf{M}_k \mathbf{M}_{k-1} \mathbf{X}_{k-1} = \cdots = \mathbf{M}_k \mathbf{M}_{k-1} \cdots \mathbf{M}_0 \mathbf{X}_{t_0} = \bar{\mathbf{M}}_t \mathbf{X}_0 \quad (15)$$

where $\bar{\mathbf{M}}_t = \mathbf{M}_k \mathbf{M}_{k-1} \cdots \mathbf{M}_0$ is a nonlinear operator.

Then the solution of (14) can be written as

$$\mathbf{X}'_{k+1} = \mathcal{M}_k \mathbf{X}_k = \mathcal{M}_k \mathcal{M}_{k-1} \mathbf{X}'_{k-1} = \cdots = \mathcal{M}_k \mathcal{M}_{k-1} \cdots \mathcal{M}_0 \mathbf{X}'_{t_0} = \bar{\mathcal{M}}_t \mathbf{X}'_0 \quad (16)$$

where $\bar{\mathcal{M}}_t = \mathcal{M}_k \mathcal{M}_{k-1} \cdots \mathcal{M}_0$ is defined as TLM operator.

Thus the cost function J of (8) can be written as

$$\begin{aligned} J(\mathbf{X}_0) = & \frac{1}{2} (\mathbf{X}_0 - \mathbf{X}_b)^T \mathbf{B}^{-1} (\mathbf{X}_0 - \mathbf{X}_b) \\ & + \frac{1}{2} \sum_{r=0}^R (\mathbf{H} \bar{\mathbf{M}}_{t_r} \mathbf{X}_0 - \mathbf{X}^{obs}(t_r))^T \mathbf{R}^{-1} (\mathbf{H} \bar{\mathbf{M}}_{t_r} \mathbf{X}_0 - \mathbf{X}^{obs}(t_r)) \end{aligned} \quad (17)$$

In order to prove that J is strictly convex, we use the increment $\delta \mathbf{x} = (\mathbf{X}_0 - \mathbf{X}_b)$ as variable.

Since \mathbf{B} is symmetric and positive definite, it can be constructed by using its square root:

$$\mathbf{B} = \sqrt{\mathbf{B}} \sqrt{\mathbf{B}}$$

Then the inverse of \mathbf{B} can be constructed as

$$\mathbf{B}^{-1} = \sqrt{\mathbf{B}^{-1}} \sqrt{\mathbf{B}^{-1}}$$

Now introducing the following variable transformation:

$$\mathbf{Z} = \sqrt{\mathbf{B}^{-1}} \delta \mathbf{x}$$

then

$$\mathbf{X}_0 = \sqrt{\mathbf{B}} \mathbf{Z} + \mathbf{X}_b$$

and

$$\mathbf{X}_{t_r} = \bar{\mathbf{M}}_t \mathbf{X}_0 = \bar{\mathbf{M}}_t (\sqrt{\mathbf{B}} \mathbf{Z} + \mathbf{X}_b)$$

Adopting the method of Courtier *et al.*(1994)[9] and Courtier(1997)[12], the cost function can be written as

$$J(\mathbf{Z}) = \frac{1}{2} \mathbf{Z}^T \mathbf{Z} + \frac{1}{2} \sum_{r=0}^R (\mathbf{H} \bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}} \mathbf{Z} - \mathbf{d}_{t_r})^T \mathbf{R}^{-1} (\mathbf{H} \bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}} \mathbf{Z} - \mathbf{d}_{t_r}) \quad (18)$$

where for linear case

$$d_{t_r} = \mathbf{X}^{obs}(t_r) - \mathbf{H} \bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}} \mathbf{X}_b$$

Thus the cost functional J becomes a function of \mathbf{Z} .

We will use (18) to demonstrate the strictly convexity of J .

Following the idea of [2], it is easy to obtain the gradient of the cost function J

$$\nabla J(\mathbf{Z}) = \mathbf{Z} + \sum_{r=0}^R \sqrt{\mathbf{B}}^T \bar{\mathcal{M}}_t^T H^T \mathbf{R}^{-1} (\mathbf{H} \bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}} \mathbf{Z} - \mathbf{d}_{t_r}) \quad (19)$$

where

$$\bar{\mathcal{M}}_t^T = \bar{\mathcal{M}}_0^T \mathcal{M}_1^T \cdots \bar{\mathcal{M}}_{t_r}^T$$

is the adjoint of $\bar{\mathcal{M}}_t$.

Thus the Hessian of cost functional J can be derived as

$$\mathcal{H} = \nabla^2 J(\mathbf{Z}) = \mathbf{I} + \sum_{r=0}^R \sqrt{\mathbf{B}}^T \bar{\mathcal{M}}_{t_r}^T H^T \mathbf{R}^{-1} \mathbf{H} \bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}} \quad (20)$$

The solution is unique if the Hessian of the cost function is positive definite. *i.e.*

$$\mathbf{X}^T \mathcal{H} \mathbf{X} > 0, \quad \forall \mathbf{X} \in R^n$$

Now for $\mathbf{X} \in R^n$,

$$\begin{aligned} \mathbf{X}^T \mathcal{H} \mathbf{X} &= \mathbf{X}^T \mathbf{X} + \sum_{r=0}^R \mathbf{X}^T \sqrt{\mathbf{B}}^T \bar{\mathcal{M}}_{t_r}^T H^T \mathbf{R}^{-1} \mathbf{H} \bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}} \mathbf{X} \\ &= \mathbf{X}^T \mathbf{X} + \sum_{r=0}^R (\mathbf{H} \bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}} \mathbf{X})^T \mathbf{R}^{-1} (\mathbf{H} \bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}} \mathbf{X}) \end{aligned} \quad (21)$$

Since \mathbf{R} is the observation error covariance matrix, by the definition, \mathbf{R} symmetric and positive definite. So as its inverse \mathbf{R}^{-1} .

Defining a R-norm:

$$\|\cdot\|_R = (R^{-1}\cdot, \cdot) = (\cdot)^T R^{-1}(\cdot)$$

The above equation (21) can be written as:

$$\mathbf{X}^T \mathcal{H} \mathbf{X} = \|\mathbf{X}\|_2 + \sum_{r=0}^R \|(H\bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}}) \mathbf{X}\|_R \quad (22)$$

The second term of (22) is a summation of series R-norm, and \mathbf{P} is positive definite, it should be non-negative, and

$$\|(H\bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}}) \mathbf{X}\|_R = 0 \quad \text{if and only if} \quad (H\bar{\mathcal{M}}_{t_r} \sqrt{\mathbf{B}}) \mathbf{X} = 0$$

The first term of (22) is a L_2 -norm, it is always positive as long as $\mathbf{X} \neq 0$.

Consequently,

$$\mathbf{X}^T \mathcal{H} \mathbf{X} = 0 \quad \text{if and only if} \quad \mathbf{X} = 0$$

This demonstrate the Hessian of the cost function with background term is positive definite, and the uniqueness of the solution of minimization is guaranteed.

It should be pointed out that without background term, Hessian matrix has the form of

$$\mathcal{H} = \nabla^2 J(\mathbf{Z}) = \sum_{r=0}^R \bar{\mathcal{M}}_{t_r}^T H^T \mathbf{R}^{-1} \mathbf{H} \bar{\mathcal{M}}_{t_r} \quad (23)$$

and

$$\mathbf{X}^T \mathcal{H} \mathbf{X} = \sum_{r=0}^R \|(H\bar{\mathcal{M}}_{t_r}) \mathbf{X}\|_p \quad (24)$$

Since H is the projection (related to the incomplete observation), one cannot guarantee that

$$\sum_{r=0}^R \|(H\bar{\mathcal{M}}_{t_r}) \mathbf{X}\|_p \neq 0 \quad \text{when} \quad \mathbf{X} \neq 0$$

References

- [1] Daley, R. 1991, Atmospheric Data Analysis. Cambridge University Press.
- [2] Navon, I.M., X. Zou, J. Derber and J. Sela, 1992, Variational data assimilation with an adiabatic version of the NMC spectral model, *Monthly Weather Review*, **120**, 1433–1446.
- [3] 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. 253pp.
- [4] Krishnamurti T.N., H.S. Bedi and V.M Hardiker, 1998, An introduction to global spectral modeling, Oxford University Press
- [5] Cohn, S.E., An introduction to estimation theory. *J. Meteorol. Soc. Japan*. **75**(1B)(1997)257–288.
- [6] Dee, D.P. and Da Silva, A.M., Data assimilation in presence of forecast bias. *Quart. J. Roy. Meteorol. Soc.* **124**(1998)269-295
- [7] Griffith, A.K. and Nichols, N.K., Accounting for model error in data assimilation using adjoint methods.In: *Computational Differentiation: Techniques, Application and Tools*. **SIAM**, Philadelphia, PA(1996), pp.195–204
- [8] Zou, X., Navon, I.M and Le Dimet, F.X., An optional nudging data assimilation scheme using parameter estimation. *Quart. J. Roy. Meteorol. Soc.* **118**(1992)1163-1186
- [9] Courtier, P., Thépaut, J.-N. and Hollingsworth, A., 1994, A strategy for operational implementation of 4D-Var, using an incremental approach, Quarterly Journal of the Royal Meteorological Society, **120**,1367-1388

- [10] Courtier, P. and Coauthors, The ECMWF implementation of three dimensional variational assimilation(3D-Var), Part I: Formulation, *Quarterly Journal of the Royal Meteorological Society*,**124**(1998),1783–1808
- [11] Courtier, P. , Variational methods, *Journal of Meteorological Society Japan*,**75**(1997),211–218
- [12] Courtier, P., 1997, Dual formulation of 4d-Var assimilation. Quarterly Journal of the Royal Meteorological Society, **123**, 2449-2461.
- [13] Weiyu Yang and I.M. Navon, Documentation of the Tangent Linear Model and its Adjoint of the Adiabatic Version of the NASA GEOS-1 C-Grid GCM–Version 5.2, *NASA Technical Memorandum* 104606, **Vol.8**(1996)
- [Liu *et al*, 2002] Zhuo Liu and I.M. Navon, Documentation of the Tangent Linear and Adjoint Models of New MPI Version of the FSU Global Spectral Model.
- [Wang,1993] Wang, Zhi, 1993, Variational data assimilation with 2-D shallow water equations and 3-D FSU global spectral models, Ph.D Dissertation
- [14] Liu, D.C., and J. Nocedal, 1989, On the limited memory BFGS method for large scale optimization, *Math. Programm.*, **45**,503–528
- [15] J. Nocedal, 1980, Updating quasi-Newton matrices with limited storage, *Mathematics of computation*, Vol.35, pp. 773–782, 1980.
- [16] Zhu K., I.M. Navon and X. Zou, 1994, Variational Data Assimilation with a Variable Resolution Finite-Element Shallow-Water Equations Model, *Monthly Weather Review*, Vol. 122, No.5, 947–965