A LOCAL EMSEMBLE TRANSFORM KALMAN FILTER DATA ASSIMILATION SYSTEM FOR THE FSU GLOBAL ATMOSPHERIC MODEL

Rosangela Cintra[1,2], Steve Cocke[2]

[1,]Brazilian National Space Research, INPE, S. J. Campos, Sao Paulo, Brazil, rosangela.cintra@lac.inpe.br
[2]Center of Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida, scocke@fsu.edu

Abstract. *Projections of future climate or weather are produced using complex atmospheric general circulation models (AGCMs). Due to the inherent uncertainty of our knowledge of the weather/climate system it is inevitable that there exists model errors. Data assimilation is the process by which measurements and model predictions are combined to obtain an accurate representation of the state of the modeled system. Data assimilation is recognized as essential in weather prediction and climate analysis. All data assimilation systems require reasonable estimates of the initial condition (analysis) to run AGCMs considering the errors of the model, the observations and the analysis. In this work, a data assimilation system, the local ensemble transform Kalman filter (LETKF) was implemented. By "local" we mean that the analysis can be carried out independently at each grid point with the use of only local observations. Uncertainty is represented not by a covariance matrix, but by an ensemble of estimates in state space. The ensemble is evolved in time through the full model, which eliminates any need for a linear hypothesis as to the temporal evolution. The LETKF assimilation scheme was tested with Florida State University Global Spectral Model (FSUGSM). The model is a multilevel (27 vertical levels) spectral primitive equation model with a vertical σ-coordinate. All variables are expanded horizontally in a truncated series of spherical harmonic functions (at resolution T63) and a transform technique is applied to calculate the physical processes in real space. The LETKF data assimilation uses the synthetic conventional observations and satellite data (surface pressure, absolute temperature, zonal component wind, meridional component wind and humidity). The observations are localized at every other grid point of the model. The ensemble forecast size is 20 members, which run parallel (one single model member per computer node simultaneously) and the assimilation scheme is parallelized via MPI. The numerical experiment has a one-month assimilation cycle, for the period 01/01/2001 to 31/01/2001 at (00, 06, 12 and 18 GMT) for each day. An important source of information for the evaluation of the quality of any data assimilation is the observation-minus-forecast (OMF) and the observation-minus-analysis (OMA) statistics. The histogram of OMF and OMA for a range of spatial and temporal scales is calculated, and the results are consistent. The results showing the analysis from the assimilation of the observations will be presented.*

Keywords. *Data assimilation, Kalman filter, numerical weather prediction, global atmospheric model.*

1 INTRODUCTION

Predictions from computer models of the atmosphere made by integrating the Navier-Stokes equations for a three dimensional multi-constituent multi-phase rotation fluid, and coupled to representations of the ocean and land surface, are continually put to the test through the daily weather forecast (Bengtsson, 1999). However, the predictability is determined by the projection of uncertainties in both initial conditions and model formulation onto flow-dependent instabilities of the chaotic weather and climate attractor (Palmer, 1999). Uncertainty is a characteristic of the atmosphere, coupled with inevitable inadequacies in observations and computer models and increases errors in weather forecasts, seasonal climate and hydrological predictions.

Model forecasts have limits to the predictability of the behavior of the atmosphere. It is because the atmosphere is an inherently chaotic fluid. While chaos in the atmosphere is a reality, there have been significant efforts in making numerical weather prediction more accurate:

- Improving the model initializations with more frequent and accurate observations. Especially the amount of remotely sensed data, such as satellite and radar, has dramatically increased.
- Allowing higher resolution simulations because the available computers have rapidly increased in size and speed.
- Improving the physical parameterizations in the numerical models because of the increased knowledge of atmospheric processes by meteorological community.

The accuracy of weather forecasts is influenced by the ability to represent computationally the full equations of motion that governs the atmosphere, in addition to error in initial conditions. Chaotic dynamics are sensitive to the error in the initial state (Lorenz, 1960). The model is a numerical weather prediction (NWP) model and the initial condition is represented by an objective analysis of an atmospheric state. Data assimilation occurs when the observations and the model are combined, performing a smooth melding of observation data and forecast model, which carries out a set of procedures to determine the best initial condition. Some techniques are used to determine initial conditions for weather forecasts given a set of atmospheric and oceanic observations whose density is heterogeneous in both space and time. The analysis is useful in itself as a description of the physical system, but it can also be used as an initial state for the further time evolution of the system (Holm, 2008).

Ensemble forecasting is an approach to NWP where the model forecasts are conducted using slightly different initial conditions. Ensembles are used to capture forecasting uncertainties. Ensemble forecasts provide a range of equally probable forecast solutions; the uncertainties are conveniently described by probability distributions. The ensemble forecasts are often referred to as having *sensitive* dependence on the initial conditions. Ensemble prediction systems provide the means to estimate the flow-dependent growth of uncertainty during a forecast.

The Kalman filter (KF) (Kalman, 1960) is one approach to estimate an appropriate analysis for atmospheric models. The Bayesian scheme is approached using ensembles of integrations of comprehensive weather prediction models, with explicit perturbations to both initial conditions and model formulation; the resulting ensemble of forecasts can be interpreted as a probabilistic prediction. The ensemble Kalman filter (EnKF) (Evensen, 1994) and the particle filter (Doucet et al., 2000) use a probability density function associated with the initial condition, characterizing the Bayesian approaches (Daley, 1991). The algorithms are constantly updated to improve the computational performance of NWP systems. One example is the version of the EnKF that represents the uncertainty, not by an error covariance matrix, but by an ensemble of point estimates in state space, which are meant to sample the probability for the state of the system. The local ensemble Kalman filter (LEKF) (Ott et al., 2004) proposes the EnKF scheme restricted to small areas (local). Another version of the EnKF scheme uses the serial ensemble square root filter (EnSRF; Whitaker and Hamill, 2002). The local ensemble transform Kalman filter (LETKF; Hunt et al., 2007) is a kind of EnSRF, but the algorithm is designed to be particularly efficient on parallel computing architectures by taking an advantage of independent local analyses of the EnKF. A number of studies have shown promise of the LETKF with wide applications including global and regional atmosphere and ocean models (e.g.: Szunyogh et al., 2005, 2008; Miyoshi and Aranami, 2006; Miyoshi and Yamane, 2007; Miyoshi et al., 2010; Cintra, 2010) and even a Martian atmosphere model (Hoffman et al., 2010).

This paper presents the first results of the LETKF implemented with the general circulation model (AGCM) of Florida State University (FSU). This data assimilation scheme is part of the analysis research at the Center for Atmospheric-Ocean Prediction Studies of FSU, USA, and the research with data assimilation using artificial neural networks at Laboratory of Computational and Applied Mathematics in INPE, Brazil. This experiment was conducted with synthetic observations, simulating measurements from surface and upper-air, and provides the basis for future research on the FSUGSM data assimilation system. Section 2 describes the LETKF and FSUGSM system related software. The experiment setup is described in section 3. Finally, the discussion and summary are provided in section 4.

2 METHODOLOGY

2.1 Brief Description on Local Ensemble Transform Kalman Filter (LETKF)

Data Assimilation (DA) is the process of finding the model representation of the atmosphere, which is consistent with the observations. According Talagrand (2008), the purpose of assimilation is to reconstruct as accurately as possible the atmospheric or oceanic flow, using all available appropriate information. The latter essentially consists of:

- The observation proper, which vary in nature, resolution and accuracy, and are distributed more or less regularly in space and time.
- The physical laws governing the evolution of the flow, available in practice in the form of a discretized, and necessarily approximate, numerical model.

A summarized history of the main data assimilation algorithms used in meteorology and oceanography, roughly classified according to their complexity (and cost) of implementation, and their applicability to real-time problems, is found in the literature, see Daley (1991) and Kalnay (2003). The computational complexity involved in DA, for example, has been presented in Lyster et al. (2004). An important problem in atmospheric data assimilation lies in the large number of degrees of freedom of NWP models. According Talagrand (2008), very large numerical dimensions are required: 10^7-10^9 parameters to be estimated, 2.10^7 observations per 24-hour period. The large number of degrees of freedom of covariance matrices involved can prohibit the implementation of the best assimilation method known where there is a need for the forecast to be ready in a short amount of time. Many strategies can be adopted to fit the intensive computation with the operational period: the use of advanced computing, reduction of problem dimension to obtain a computer code feasible to run in real time; and the use of parallel computing with thousands of processors. The data assimilation algorithms are constantly updated and improved in performance.

One common approach at present is the EnKF where the uncertainty is represented, not by an error covariance matrix, but by an ensemble of point estimates in state space, which are meant to sample the conditional probability distribution for the state of the system. The ensemble is evolved in time through the full model, which eliminates any need for a linear hypothesis as to the temporal evolution. The ensemble forecasts are used to evaluate the probability distribution. Based on ensemble forecasting, the probabilistic state space formulation and the requirement for updating information when observations are encountered, the Bayesian approach is used to get the "errors of the day" of the predictions. The Bayesian approach is a set of efficient and flexible Monte Carlo methods for solving the optimal filtering problem.

The Kalman Filter, a sequential assimilation scheme, is the best linear unbiased estimate of analysis, where the equations are obtained from an analytical solution from setting the gradient of the cost function to zero, considering recursive least square and the assumption of the Gaussian probability density functions (pdf).

A brief description for Kalman filter algorithm is expressed below:

- Analysis step: update the analysis covariance matrix

$$x_n^a = x_n^f + P_n^f H_n^T [H_n P_n^f H_n^T + R]^{-1} \left(x_{n+1}^{obs} - H_{n+1} x_{n+1}^f \right)$$

$$P_n^a = P_n^f - P_n^f H_n^T [H_n P_n^f H_n^T + R]^{-1} H_n P_n^f .$$

- Forecast step:

$$x_{n+1}^f = M_n \left(x_n^a \right)$$

$$P_{n+1}^f = M_n P_n^a M_n^T + W_n^b .$$

The analysis x_n^a and the associated covariance matrix P_n^a are updated at analysis step. The matrices M and H represent the dynamical system and observation operator, respectively. The covariance matrix R identifies the observation error. The covariance matrix P_{n+1}^f is associated with the forecast model x_{n+1}^f updated at forecast step and W_n^b is the modeling error.

In the EnKF approach, an ensemble of estimates in state space represents the covariance matrix P_{n+1}^f . The nonlinear evolution problem for the error covariance is calculated by sampling from the pdf and propagating the samples, the model states, forward in time with the fully nonlinear model equations. At any time the samples can be used to calculate an approximate mean and error covariance. The best implementation is in localization, in which applying a cut-off radius of influence for each observation eliminates spurious correlations. This is the Local Ensemble Kalman Filter (LEnKF) algorithm that captures the space of forecast uncertainties, formulated by ensemble-based Kalman filter scheme.

The LETKF algorithm is a LEnKF-based scheme, in which the analysis ensemble members are constructed by a linear combination of the forecast ensemble members (Myioshi, 2005). The ensemble transform matrix, composed of the weights of the linear combination, is computed for each local subset of the state vector independently and thus allows parallel computation. The local subset depends on the error covariance localization (Myioshi and Yamane, 2007). In the experiment with LETKF-FSUGSM,a local subset of the state vector contains all variables at the region centered at given grid point.

Each member of the ensemble gets its forecast:

$$\left\{ x_{n-1}^f \right\}^{(i)} : i = 1,2,3,\ldots,k \, ,$$

where k is the total members at time t_n , to estimate the state vector $\overline{x^f}$ of the reference model. The ensemble is used to calculate the forecast by the average:

$$\overline{x^f} \equiv k^{-1} \sum_{i=1}^{k} \{ x^f \}^{(i)}$$

and the model error matrix is given by:

$$P^f = (k-1)^{-1}\sum_{i=1}^{k}(\{x^f\}^{(i)} - \overline{x^f})(\{x^f\}^{(i)} - \overline{x^f})^T$$

The LETKF determines an analysis $\left\{x_{n-1}^a\right\}^{(i)} : i = 1,2,3,...,k$ for each member of the ensemble and an appropriate sample mean state estimate: $\overline{x^a} \equiv k^{-1}\sum_{i=1}^{k}\{x^a\}^{(i)}$, and covariance:

$$P^a = (k-1)^{-1}\sum_{i=1}^{k}(\{x^a\}^{(i)} - \overline{x^a})(\{x^a\}^{(i)} - \overline{x^a})^T$$

Data assimilation problems are often limited by the high dimensionality of states created by spatial discretization over large high-resolution grids and the extensive spatial structure of observations. The LETKF is suitable for such problems, promising computational efficiency and accuracy in using the localization method.

The code of the LETKF in this experiment is based on the system initially developed by Miyoshi (2005) and has been continuously improved. The current version is a Message Passing Interface (MPI)-parallelized Fortran90 code and includes spatial covariance localization with physical distance (Miyoshi et al. 2007), four-dimensional EnKF (4D-EnKF) for appropriate treatment of asynchronous observations (Hunt et al. 2004) and temporal covariance localization. The LETKF code has been applied to and assessed with the Lorenz 40-dimensional model (Lorenz, 1996), a low-dimensional AGCM known as the SPEEDY model (Molteni, 2003; Miyoshi, 2005; Cintra, 2010), realistic atmospheric models such as the AGCM for the Earth Simulator (Miyoshi and Yamane, 2007) and the Japan Meteorological Agency operational global and mesoscale models (Miyoshi and Aranami, 2006; Miyoshi and Sato, 2007; Miyoshi et al.; 2010), the Geophysical Fluid Dynamics Laboratory (GFDL) Mars AGCM (Wilson and Hamilton, 1996, Greybush, 2011) and Center for Weather Forecast and Climate Studies (CPTEC) AGCM (Medeiros et al. 2010). All applications showed successful data assimilation using the LETKF code. The core part of the LETKF code is shared and improvements from applications can benefit other applications directly. This research with FSUGSM is an important application in the use LETKF system; this model has some different characteristics, which may lead improvements of the LETKF system after this evaluation.

2.2 FLORIDA STATE UNIVERSITY GLOBAL SPECTRAL MODEL (FSUGSM)

The FSUGSM is a general circulation model, a global spectral model based on primitive equations. The vertical coordinates are defined on sigma (σ) surfaces. The horizontal coordinates are latitude and longitude on a Gaussian grid in real space. The spectral model, used in this study, runs with T63 horizontal resolution (approximately 1.875°) and 27 unevenly spaced vertical levels. Details of this model can be found in Cocke and LaRow (2000).

The dynamical processes are the six primitive equations to forecast atmospheric motion: vorticity, divergence, thermodynamic, continuity, hydrostatic and moisture, which are expanded in their spectral form. The nonlinear terms are calculated on a Gaussian grid using a transform method. The vertical discretization of the FSUGSM uses a finite difference scheme and a semi-implicit leapfrog scheme is used for time integration. The full physical packages include orography, planetary boundary layer, dry adjustment, large-scale precipitation, moist-convection, horizontal diffusion, and radiation processes. The horizontal diffusion term is usually incorporated in a numerical weather prediction model to parameterize the effects of motions on the unresolved scales and to inhibit spectral blocking, that is, the growth of small scales in the dynamic model variables due to the accumulation of energy at high wavenumbers. The presence of any dissipation, physical or computational, can attenuate the amplitude of the short wavelengths very significantly, as cited by Zhuin and Navon (2000).

According Cocke and LaRow (2000), the global model has been developed to take advantage of scalable parallel architectures. The grid calculations are done use a domain decomposition approach. In this experiment, we use a small number of processors, and then the domain decomposition is simply a one-dimensional partitioning of latitude bands. Each latitude band may be arbitrarily assigned to any processor to achieve optimal load balance. The vertical calculations for any given domain are done on the same processor.

3 Experimental LETKF-FSUGSM

Data assimilation (DA) is the process where observations are incorporated into model representation of the atmospheric state, and are adjusted in real time as new data becomes available. The result of the DA process is a model state that is consistent with the observed data, which can be used as an initial condition for next model prediction period; this run is called the DA cycle. The LETKF-FSUGSM is tested with synthetic observations simulating surface and upper-air observations at the model grid point localization. The LETKF analyses the prognostic variables: zonal component wind (*u*), meridional component wind (*v*), temperature (*T*), specific humidity (*q*), and surface pressure (*ps*), for the period starting on 01/01/2001 until 31/01/2001. Observations are taken and analyzed every 6 hours (00, 06, 12, 18 UTC). The control or "nature run" model fields for this assimilation experiment are obtained from the integration of the model without analysis, i.e. the initial condition for the next run is the previous model forecast. The first analysis, to run the control model, is taken from the National Centers for Environmental Prediction (NCEP). The FSUGSM control run is started with NCEP analysis for 31/12/2000-18UTC and the 6 hour forecast (01/01/2001 00 UTC), it is the initial condition used to run the control model for the next run. The control model generates the 6 hour forecast which is the initial condition for the next model and repeated for entire period of experiment

3.1 Observations

The data assimilation experiments in this study are based on synthetic observation simulation experiments, where the control model fields are assumed to be known, and observations are simulated by adding Gaussian random noise to the control or "nature run" model; this noise is calibrated according to observational errors.

The observational grid is a regularly distributed dense network; it has (45 x 96 x 27) grid points for latitude, longitude, and vertical directions, respectively for four upper-air variables (*u*, *v*, *T* and *q*) and (45 x 96) for the surface variable (*ps*). This grid localization is every other latitude/longitude grid point of the FSUGSM native grid of (96 x 192, 27). We exclude the extreme north/south points of latitude, simulating no observations near the poles. Large errors exist in the polar region where few observations are available. Figure 1 shows the observation grid for one vertical level. Figure 1 shows an example of the temperature observation grid at one level that is about 4320 observations.

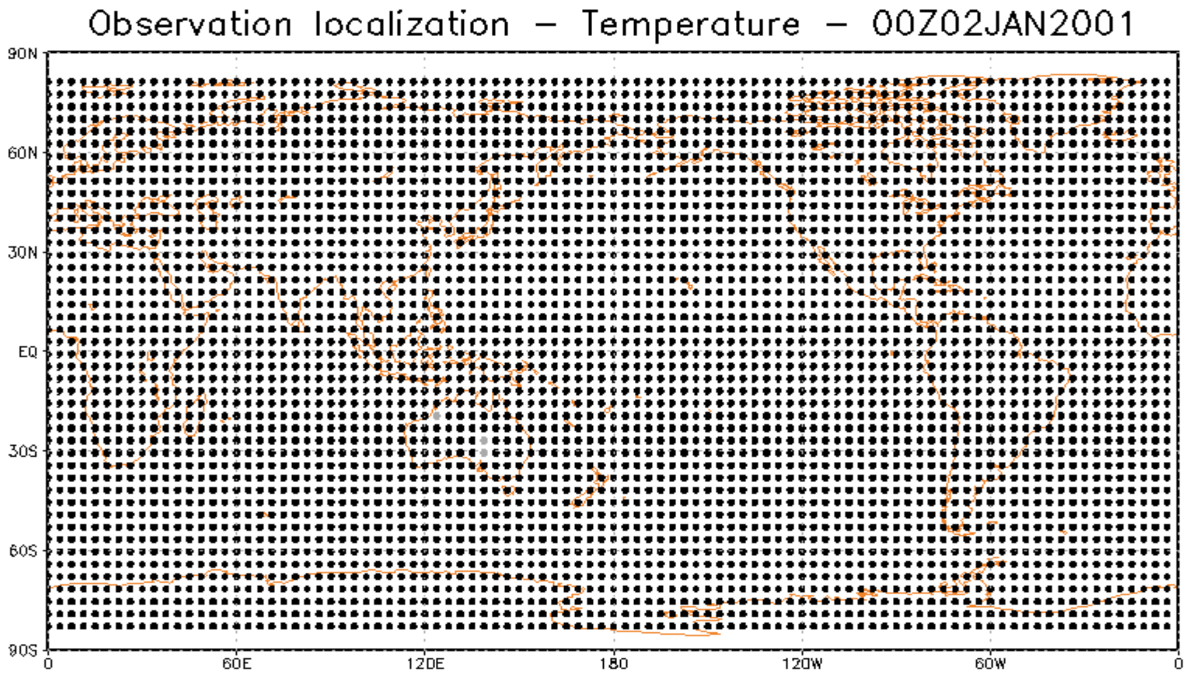

Figure 1 – The dot points shows the Temperature observation grid for 02/01/2001 at 00 UTC.

3.2 Routines

Firstly, we run the FSUGSM to generate the control or "nature run" fields and then we perform the observational routine to collect the synthetic observations based on the model fields. The next step is to perform the analysis-forecast cycle. We run 124 analysis-forecast cycles. The first forecast to initiate the analysis cycle is the control field model from 01/01/2001 at 00 UTC (the result of the model performed with NCEP analysis for 31/12/2000-18UTC).

The LETKF system is run with 20 members, the first 6-hour forecast is the same for the members. The LETKF program runs parallel with 4 processors. After LETKF is performed, we obtain the error covariance matrix, the analysis for each member, the first-guess mean of the members and the analysis mean of the members. Then, we run the FSUGSM model for each member; we run a forecast for each member simultaneously, submitting a job to one processor for each member at same time. Each job runs independently for the 6-hour forecast. These forecasts are the first-guess for the next assimilation cycle. The next assimilation cycle begins (for 6 hours) as soon as the first-guess members and observations are ready. These tasks continue until the end of the assimilation period 31/01/2001 at 18 UTC.

According Laroz et al. (2007), the observation minus forecast (OMF) increment gives a raw estimation of the agreement of the forecast information (i.e., the first guess) with the observation information prior to assimilation. Usually, a small OMF increment indicates a high quality forecast, and OMF increments are used as a primary measure of the quality of the assimilation. The observation minus analysis (OMA) increment represents the changes to the model forecast that are derived from the analysis algorithm. If the assimilation system weighs the observations heavily relative to the forecast, then the OMA increments will have significant differences relative to the OMF increments. If the model information is weighed more heavily than the observational information then there will be little change represented by the OMF increments.

The computer used to perform this experiment is the FSU High Performance Computing (HPC) Cluster that provides 403 computer nodes and 6,464 CPU cores. Jobs are managed by scheduling software (e.g. batch processing). All programs are developed using Fortran90 and bash scripts are developed to implement the operation of data assimilation cycle.

Although parallelization is employed, the focus of this LETKF implementation is the evaluation of the analysis for the FSUGSM, and not on the computational performance.

3.3 Results

In this section we show the results of the LETKF-FSUGSM analysis-forecast cycle experiment. We compare the behavior of the model with analysis by comparing with its first-guess, observations and control model fields.

Figure 2 and Fig. 3 show the model fields of the LETKF analysis, first-guess (6-hours forecast), control model and the differences between analysis and control model. Fig. 2 presents temperature in Celsius degree (C) on the second vertical level (bottom to top), generated from assimilation cycle 08/01/2001-00UTC. Fig. 3 presents surface pressure in hector-Pascal (hPa) generated from assimilation cycle 13/01/2001-18UTC.

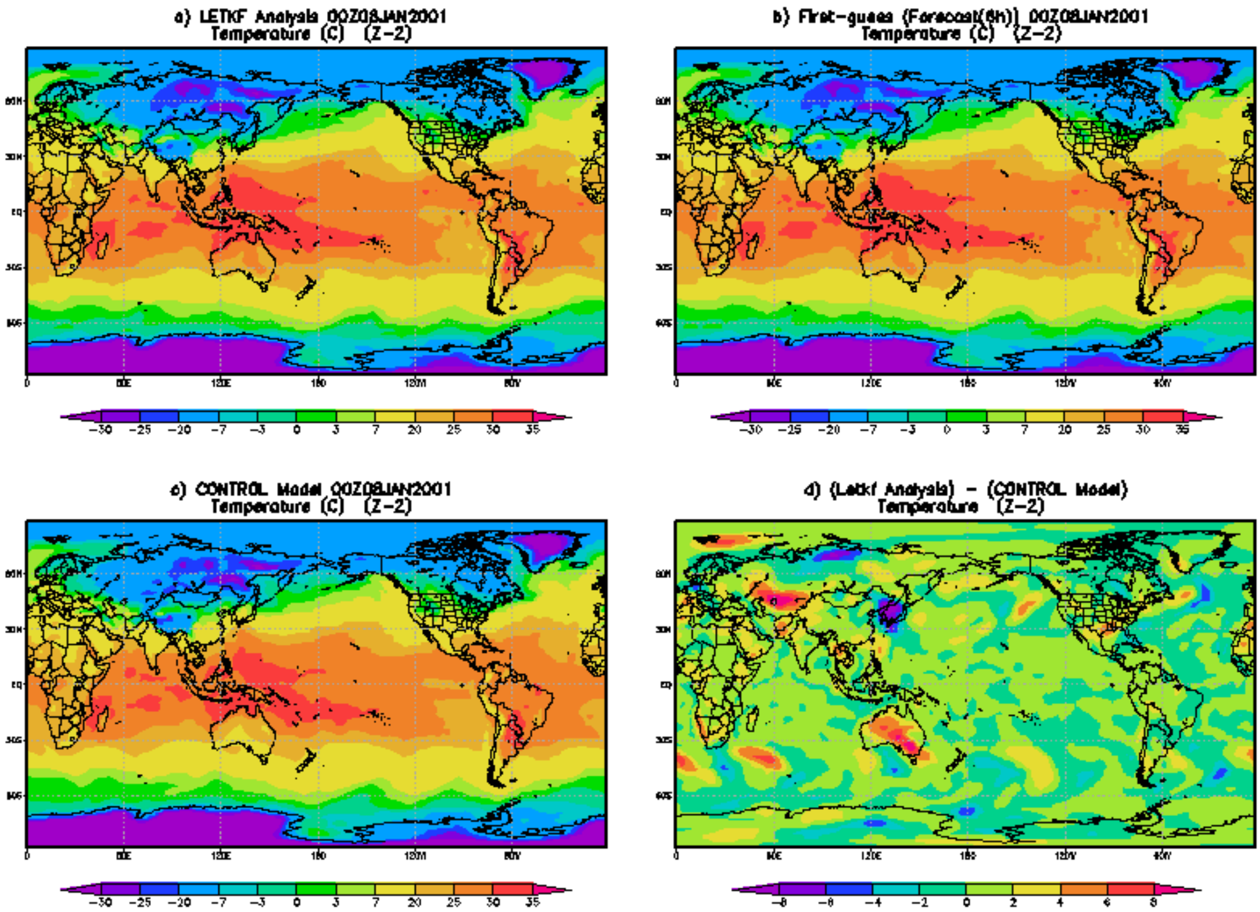

Figure 2 – Temperature from FSUGSM (^{o}C) at sigma level 0.976 on 08/Jan/2001-00UTC: a) Letkf analysis; b) first-guess; c) control model; d) differences between analysis and control model.

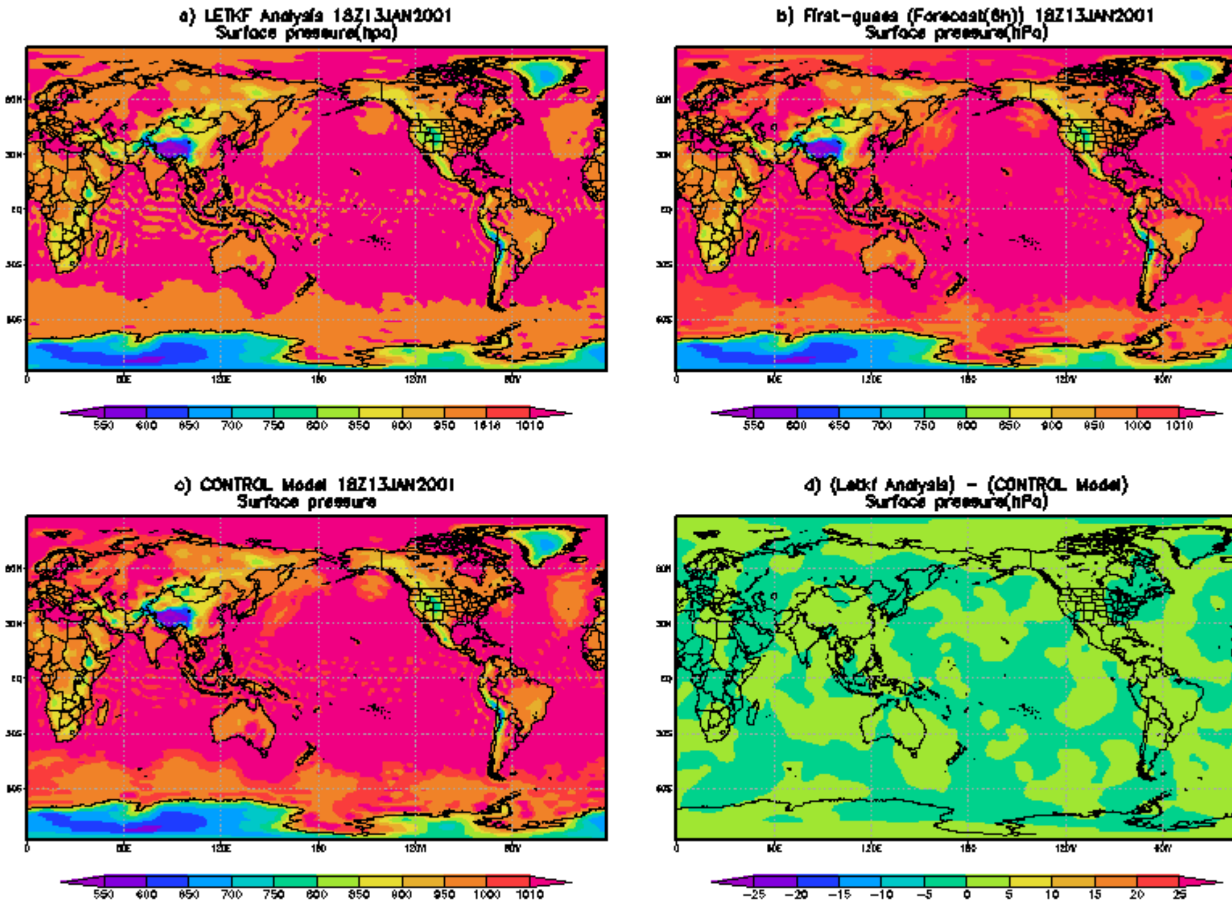

Figure 3 – Surface Pressure from the FSUGSM (hPa) on 13/Jan/2001-18UTC: a) Letkf analysis; b) first-guess; c) control model; d) differences between analysis and control model.

We note that analysis is similar to the forecast, and the differences from control model are only large at some points, the largest differences in surface pressure are around -5 hPa to 5 hPa, and in temperature are about -6°C to 4°C. These results show the effectiveness of data assimilation; the net result is a combination of forecast and observation based on the control model. The analysis thus retains much of the forecast bias.

Figure 4 shows pressure fields over United States of America and Fig. 5 shows temperature fields over Brazil. These fields are from LETKF analysis, control field (base of observations), and differences between those. Fig. 4 shows surface pressure generated for 15/01/2001 at the 12UTC assimilation cycle and Fig. 5 shows temperature at the 3rd level from bottom for 03/01/2001 at 06 UTC. These results present small differences between the analysis and control model.

Figure 6 and Fig. 7 present the zonal global means from three latitudes: 30° North (30N), Equator (EQ) and 30° South (30S) during the 124 assimilation cycles of January 2001. The means are from the control model, 6-hour forecast and the LETKF analysis. Fig. 6 shows the surface pressure zonal means: the mean ranges between 960 and 976 hPa at 30N; between 992 and 1008 hPa at Equator; and between 1007 and 1016 at 30S. Fig. 7 shows surface temperature zonal means: the mean ranging between 10 and 16 °C at 30N; between 27 and 29 °C at Equator, and between 22 and 24 °C at 30S. Considering that the summer season is occurring in the southern hemisphere and winter season is occurring in the northern hemisphere, the results are coherent and the forecast and analysis are stable.

Figure 8 presents the monthly average of specific humidity differences between control model and the LETKF analysis during January 2001. The result shows that the analysis is consistent with the control model; the major differences are between 0.05 and 0.2 Kg/Kg.

Since observations are constructed from the control model, the OMF and OMA are expected to match each other. Furthermore, good performance means that the root mean square error (RMSE) indicated by OMA is smaller than RMSE in OMF. The quality of the analysis is evaluated in terms of bias (observation minus analysis), and the root mean square (RMS) of the bias. These increments are shown in Fig. 9 (bias) and Fig. 10 (RMSE) , Fig. 9 presents the OMA of surface pressure and temperature, in red line, are smaller than OMF in blue marks, it confirms that LETKF-FSUGSM works well.

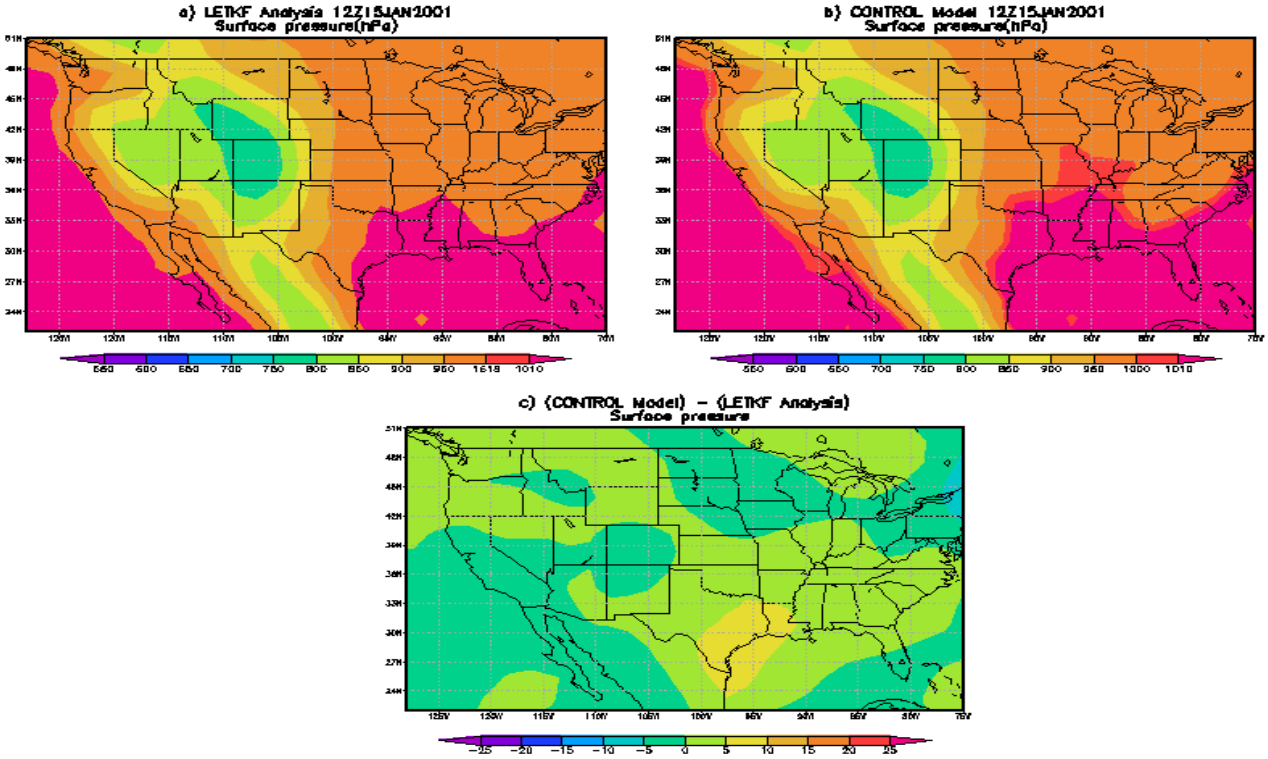

Figure 4 – Surface Pressure from FSUGSM (hPa) over USA on 15/Jan/2001-12UTC: a) Letkf analysis; b) control model; c) differences between analysis and control model.

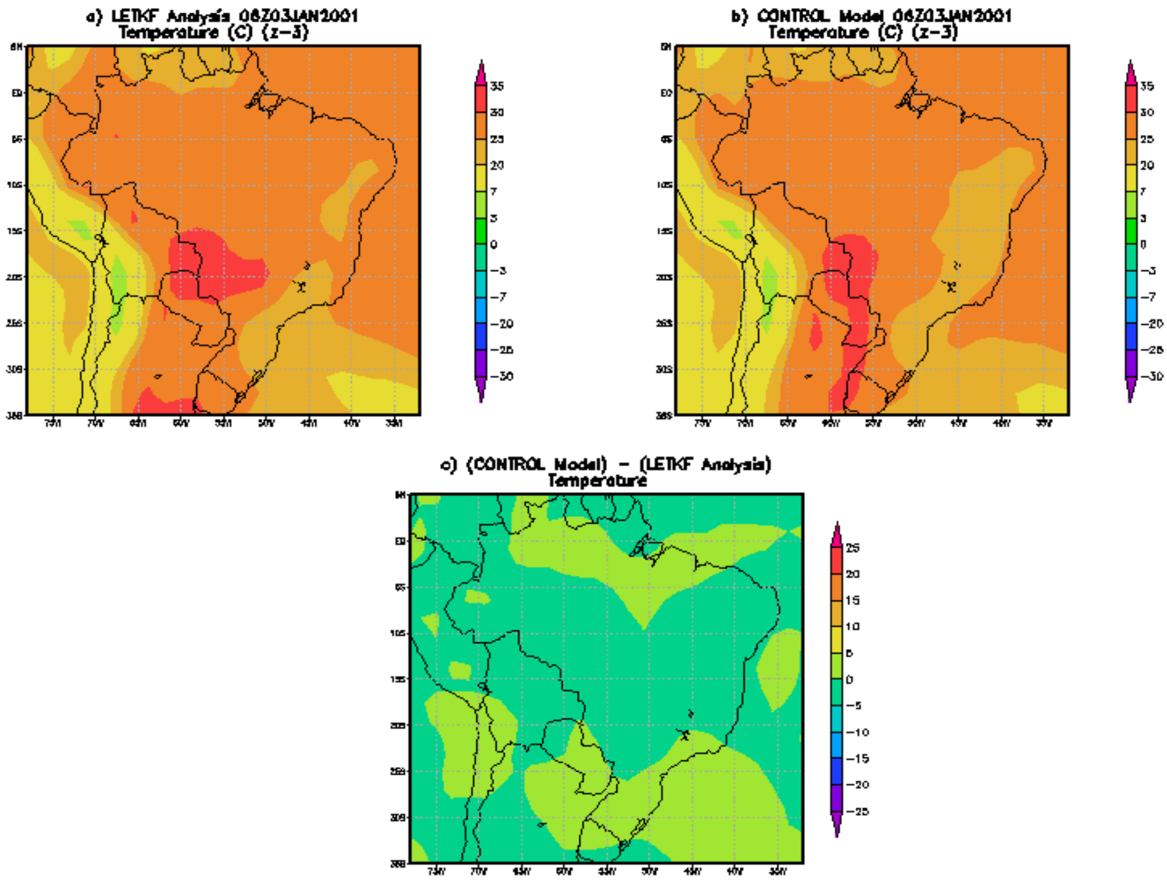

Figure 5 – Temperature at sigma level 0.953 from FSUGSM (^{o}C) over Brazil on 03/Jan/2001-06UTC: a) Letkf analysis; b) control model; c) differences between analysis and control model.

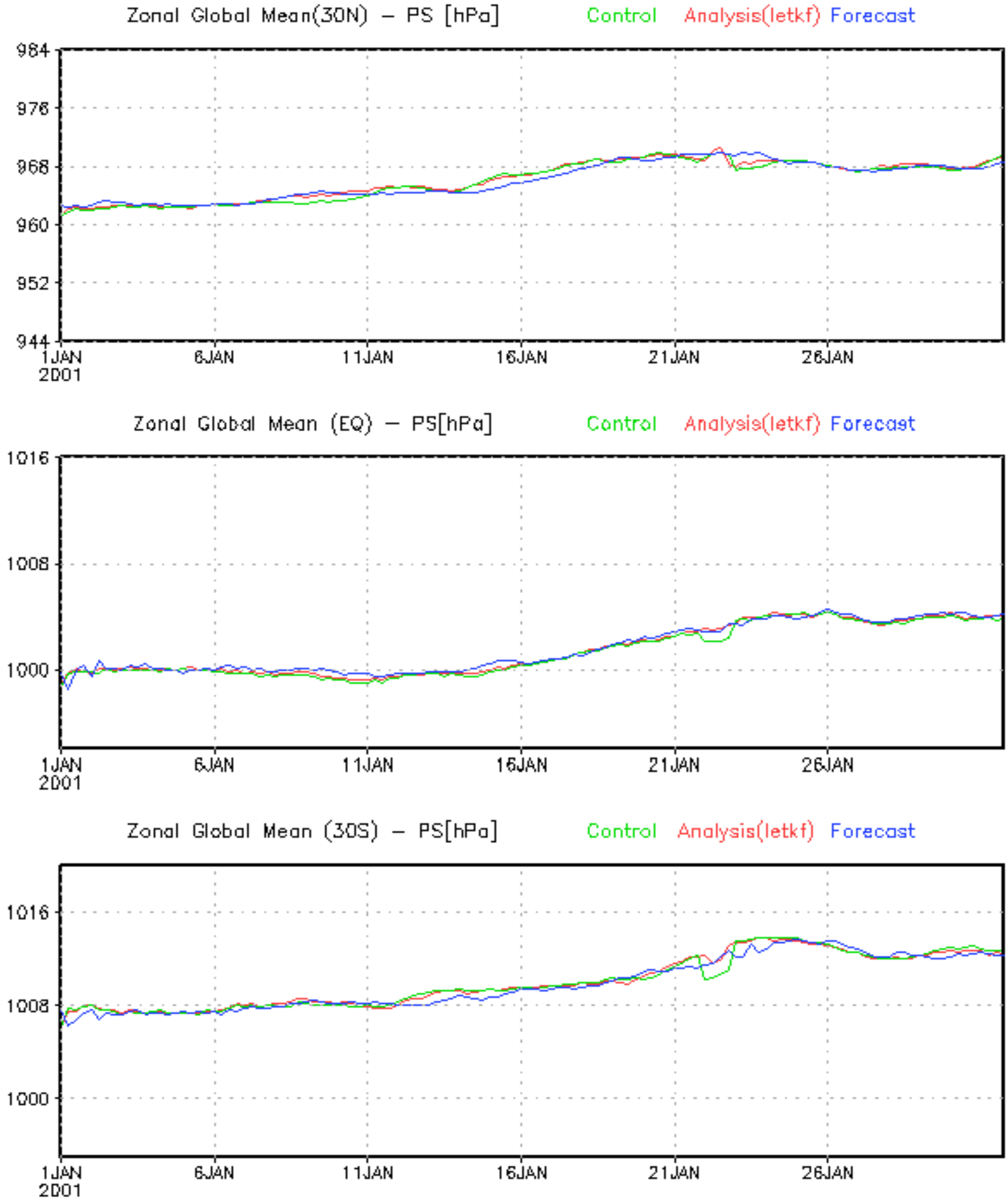

Figure 6 – Zonal mean of surface pressure from FSUGSM (hPa) during January, 2001 at latitudes : 30° North (30N); Equator (EQ); 30° South (30S). The blue lines are 6-hours forecast means, the red lines are LETKF analysis and the green line is from the control model.

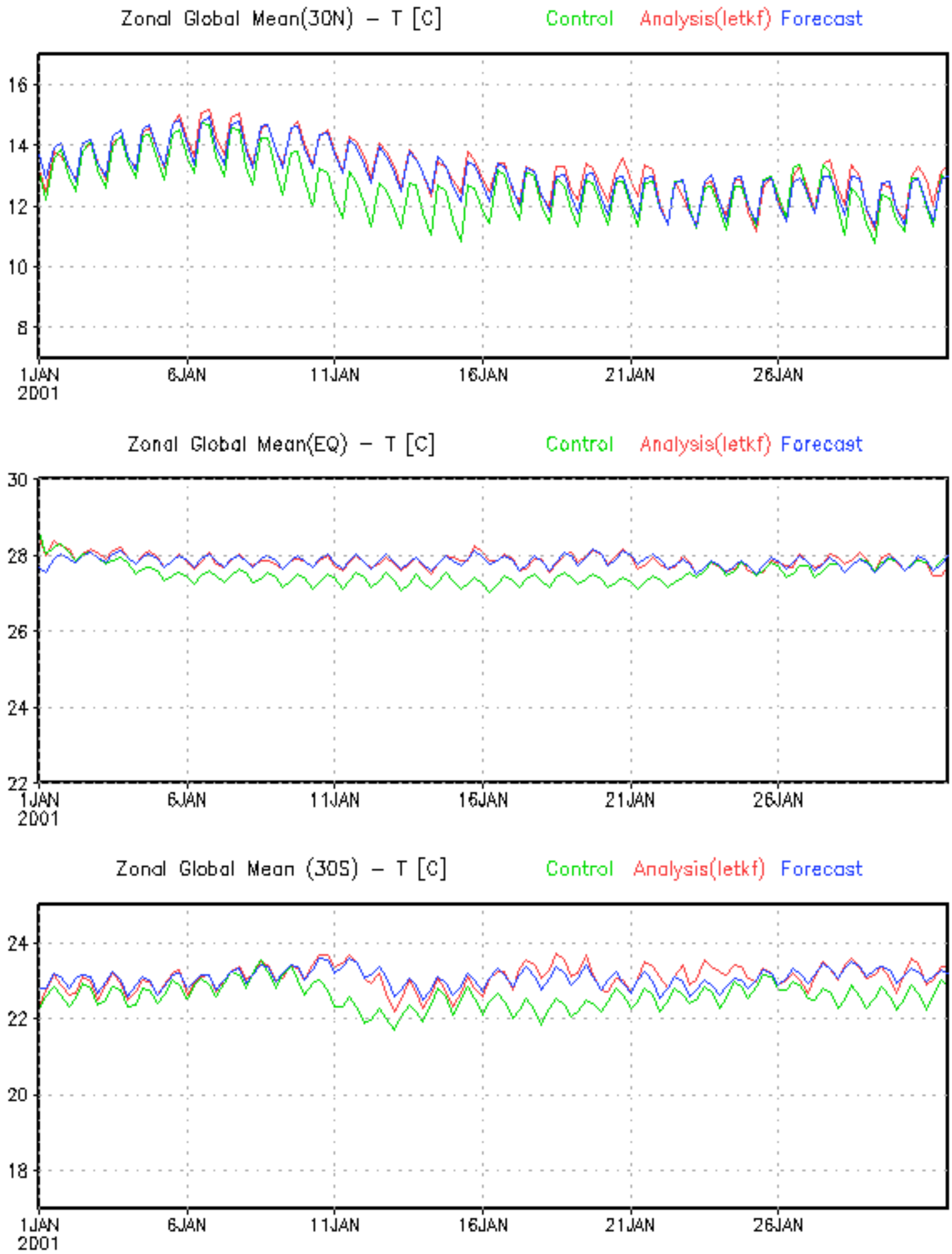

Figure 7 – Zonal mean of surface temperature from FSUGSM (^{o}C) during January 2001 at latitudes: 30^{o} North (30N); Equator (EQ); 30^{o} South (30S). The blue lines are 6-hours forecast means, the red lines are LETKF analysis and the green line is the control model.

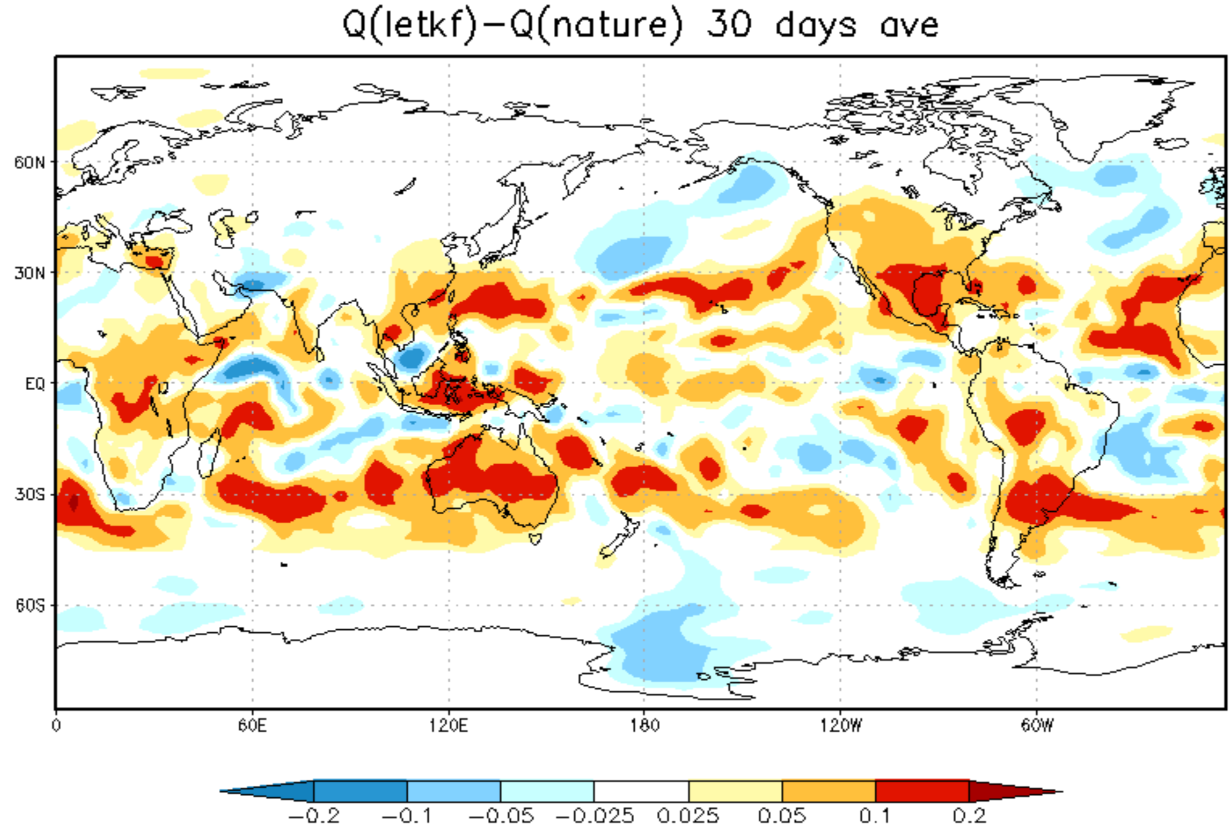

Figure 8 – Difference in global mean of humidity (Kg/Kg) between FSUGSM control model and LETKF-FSUGSM analysis during January 2001.

4 SUMMARY AND DISCUSSION

The results using LETKF data assimilation with the FSUGSM model are summarized as follows:

- The LETKF data assimilation cycle experiment was performed using a dense synthetic observational network, in which the synthetic observations are provided every 6 hours on a subset of the native model grid. The preliminary analysis shown here indicates that the performance of the assimilation is reasonable.
- The ensemble size is chosen to be 20 members.
- The FSUGSM is a computationally efficient numerical weather prediction model and was used as the forecast model at T63L27 resolution in the assimilation system.
- The LETKF data assimilation method performs an analysis at each grid point simultaneously using the state variables and all observations in the region centered at a given grid point.
- The ensemble is used to calculate the average of the forecasted members, and then used to get the model error covariance matrix to run the LETKF method.

The LETKF-FSUGSM assimilation system was stable and was able to generate analyses to initialize the FSU model and the results shows it is suitable for weather prediction.

The investigation is ongoing, and we plan to get more results performing the analysis for a longer period to evaluate errors and forecasts. Furthermore, we are implementing another strategy that uses NCEP reanalysis to generate the "observations" rather than deriving them from the FSUGSM.

5 ACKNOWLEDGEMENTS

The authors thank Dr. Takemasa Miyoshi for providing computer routines of the LETKF code and Dr. Haroldo de Campos Velho for fruitful discussions. This paper is a contribution of the Brazilian National Institute of Science and Technology (INCT) for Climate Change funded by CNPq Grant Number 573797/2008-0 e FAPESP Grant Number 2008/57719-9.

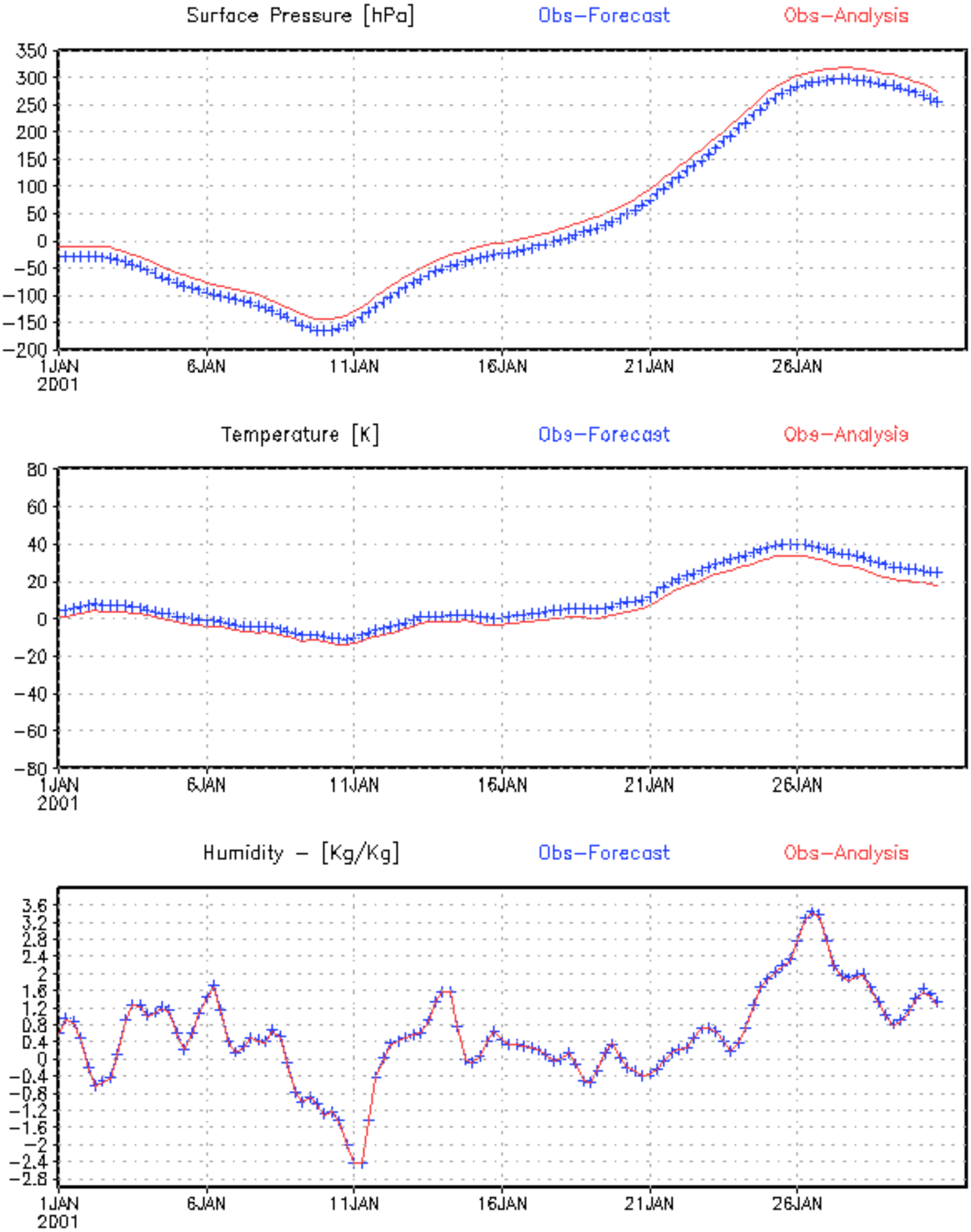

Figure 9 – Observation-minus-Forecast (OMF) increment in blue marks compared with Observation-minus-Analysis (OMA) in red line, to January 2001.

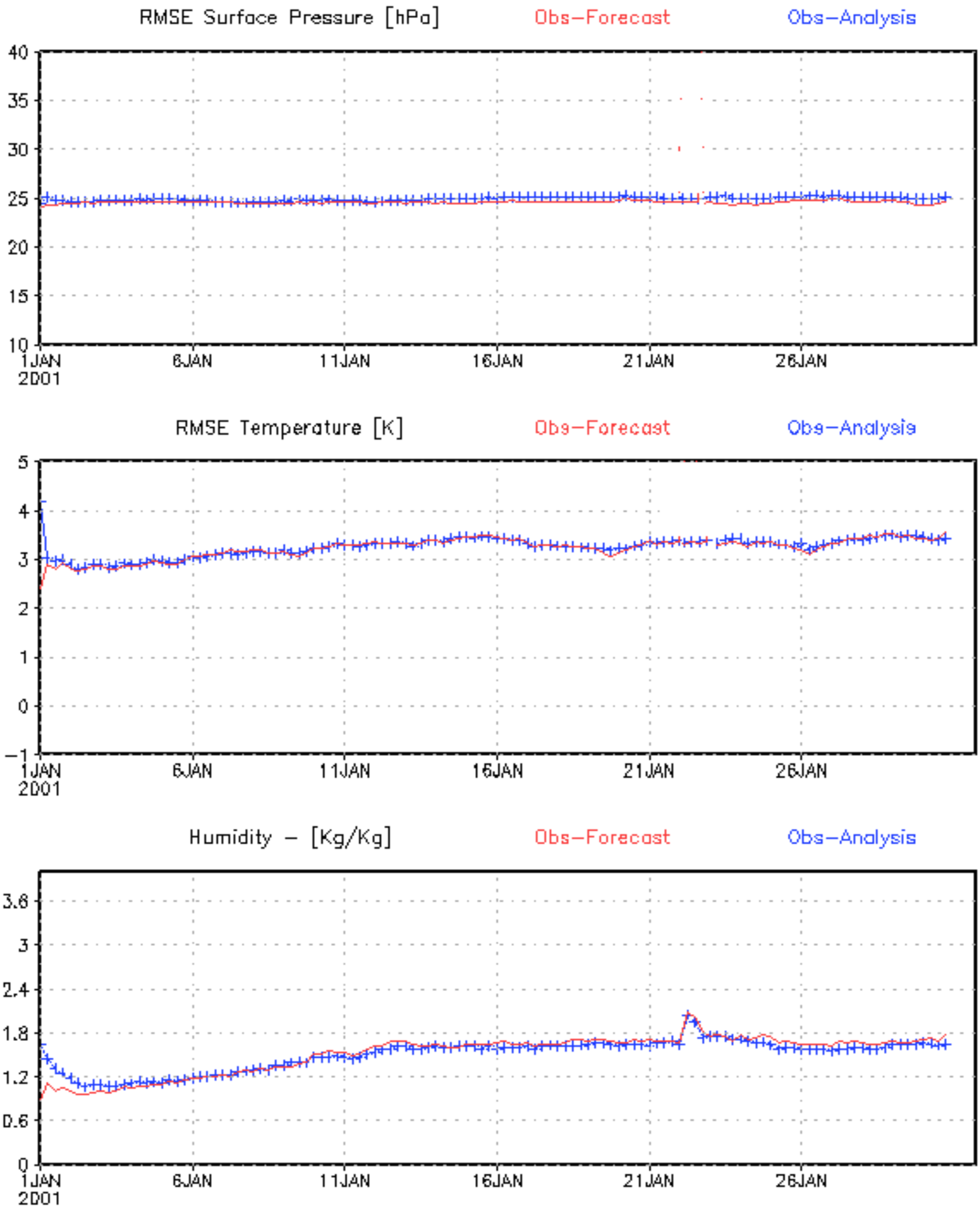

Figure 10 – Root Mean Square errors of OMF in blue marks and OMA in red line for January 2001.

6 REFERENCES

Anderson, J. L. and S. L. Anderson , 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon. Wea. Rev., 127, 2741–2758.

Bengtsson, L., 1999: From short range barotropic modelling to extended range global weather prediction - a 40 year

perspective. *Tellus* 51 A-B, 1, 13-32.
Bishop, C. H., B. J. ETHERTON, and S. J. MAJUMDAR, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. Mon. Wea. Rev., 129, 420–436.
Cintra, R. C., 2010: Assimilacao de dados com redes neurais artificiais em modelo de circulacao geral da atmosfera. Thesis (D.Sc. on Applied Computing), National Institute for Space Research, 200 p., Sao Jose dos Campos. 2010. Available on-line: <http://urlib.net/8JMKD3MGP7W/389UKPS>.
Cocke, S., and T. E. LaRow, 2000: Seasonal Predictions Using a Regional Spectral Model Embedded within a Coupled Ocean-Atmosphere Model. Mon. Wea. Rev. 128, 689-708.
Daley, R., 1991: Atmospheric Data analysis. Cambridge University Press, New York, USA, 471 pp.
Doucet, A., S. Godsill, and C. Andrieu, 2000: On sequential monte carlo sampling methods for Bayesian filtering. Statistics and Computing, 10(3), 197{208}.
Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi{geostrophic model using monte carlo methods to forecast error statistics. Journal of Geophysics Research, 99, 10 143{10 162.
Greybush, S., 2011: Mars Weather and Predictability: Modeling and Ensemble Data Assimilation of Spacecraft Observations. Ph.D. Dissertation, University of Maryland, College Park.
Hoffman, M. J., S. J. Greybush, R. J. Wilson, G. Gyarmati, R. N. Hoffman, E. Kalnay, K. Ide, E. J. Kostelich, T.Miyoshi, and I. Szunyogh, 2010: An Ensemble Kalman Filter Data Assimilation System for the Martian Atmosphere: Implementation and Simulation Experiments. Icarus, 209, 470–481.
Hunt, B. R., E. J. Kostelichi and I. Szunyogh, 2007: Efficient Data Assimilation for Spatiotemporal Chaos: A Local Ensemble Transform Kalman Filter. Physica D, 230, 112–126.
Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82 (1): pp.35–45.
Kalnay, E., 2003: Atmospheric Modeling, Data Assimilation and Predictability. 2d ed., Cambridge University Press, New York.
Krishnamurti, T. N., M. Kanamitsu, B. Ceselski, M. B. Mathur, 1973: Florida State University's Tropical Prediction Model. *Tellus* **25**:6, 523-535
Lyster, J. Guo, T. Clune, and J.W. Larson The Computational Complexity and Parallel Scalability of Atmospheric Data Assimilation Algorithms, Journal of Atmospheric and Oceanic Technology, 21(11), 1689-1700, 2004.
Medeiros, M. S., Herdies, D. L., Aravequia, J. A., Souza, S. S., 2010, "The Impact of Assimilation with the Inclusion of Amsu-A Radiances in 4D-LETKF/AGCM System", American Geophysical Union, Fall Meeting 2010, abstract #A13B-0197.
Miyoshi, T., 2005: Ensemble Kalman filter experiments with a primitive-equation global model. Ph.D. dissertation, University of Maryland, College Park, 197 pp.
Miyoshi, T., 2011: The Gaussian Approach to Adaptive Covariance Inflation and its Implementation with the Local Ensemble Transform Kalman Filter. Mon. Wea. Rev., 139, 1519–1535. doi: 10.1175/2010MWR3570.1.
Miyoshi,T. and K. Aranami 2006: Applying a Four-dimensional Local Ensemble Transform Kalman Filter (4D-LETKF) to the JMA Nonhydrostatic Model (NHM). SOLA, 2, 128–131.
Miyoshi, T. and Y. Sato, 2007: Assimilating Satellite Radiances with a Local Ensemble Transform Kalman Filter (LETKF) Applied to the JMA Global Model (GSM). SOLA, 3, 37–40.
Miyoshi, T. and S. Yamane, 2007: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. Mon. Wea. Rev., 135, 3841–3861.
Miyoshi,T., S. Yamane, and T. ENOMOTO, 2007: Localizing the Error Covariance by Physical Distances within a Local Ensemble Transform Kalman Filter (LETKF). SOLA, 3, 89–92. doi:10.2151/sola.2007-023.
Miyoshi,T., Y. SATO, and T. Kadowaki, 2010: Ensemble Kalman filter and 4D-Var inter-comparison with the Japanese operational global analysis and prediction system. Mon. Wea. Rev., 138, 2846–2866.
Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. Climate Dynamics, 20, 175–191.
Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data assimilation. Tellus, 56A, 415–428.
Palmer, T. N., 1999: A Nonlinear Dynamical Perspective on Climate Prediction. *J. Climate*, **12**, 575–591.
Penny, S., 2011: Data Assimilation of the Global Ocean Using the 4D Local Ensemble Transform Kalman Filter (4D-LETKF) and theModular Ocean Model (MOM2). Ph.D. Dissertation, University of Maryland, College Park.
Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B.R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, 2005: Assesing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model. Tellus, 57A, 528–545.
Talagrand, O, 2008: Data Assimilation in Meteorology and Oceanography. Academic Press.
Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. Mon. Wea. Rev., 130, 1913–1924.
Wilson, R. J. and K. P. Hamilton, 1996: Comprehensive model simulation of thermal tides in the martian atmosphere. J. Atmos. Sci., 53, 1290–1326.

Zhijin Li , I. Michael Navon and Yanqiu Zhu, 2000: Performance of $D-Var with different strategies for the use of adjoint physics with FSU global spectral model, *Monthly Weather Review* , 128, No.3, 668-688.