5.4 IMPROVED ANALYSES AND FORECASTS WITH AIRS RETRIEVALS USING THE LOCAL ENSEMBLE TRANSFORM KALMAN FILTER

Hong Li[1*], Junjie Liu[1], Elana Fertig[1]
Eugenia. Kalnay[1], Jose. A. Aravéquia[2], Istvan Szunyogh[1], Eric J. Kostelich[3], and Ricardo Todling[4]

[1]University of Maryland, College Park, MD, USA
[2] Institute of Space Research, Cachoeira Paulista, SP, Brazil
[3]Arizona State University, Tempe, AZ, USA
[4]NASA-GSFC Global Modeling and Assimilation Office, Greenbelt, MD, USA

1. INTRODUCTION

The Local Ensemble Transform Kalman Filter (LETKF) (Hunt et al. 2006) is an efficient data assimilation scheme of the square root ensemble Kalman filter family. It has been implemented to assimilate simulated observations in the NCEP GFS model (Szunyogh et al. 2005), and in the NASA fvGCM model (Liu et al. 2006). The results from LETKF are much better than those from 3DVAR in a perfect model scenario. With real data, LETKF has been shown to be superior to the operational NCEP SSI (operational 3DVAR) by Szunyogh et al 2006 when applied on the NCEP GFS model at T62L28 resolution, and verified against the NCEP T254/L64 analysis using all available operational observations. Unlike other square-root schemes that solve the Kalman filter equations in observation space (Anderson 2001, Bishop et al. 2001, Whitaker et al. 2004), LETKF solves the equations locally in model space. In this way, LETKF can utilize parallel computation and is more efficient when assimilating satellite observations, the number of which can be much larger than the number of degrees of freedom in the model.

The Atmospheric Infrared Sounder (AIRS) was lunched on EOS Aqua in 2002. Some positive impacts on global analysis and forecast have been found in 3DVAR (Marshall et al. 2006, Chahine et al. 2006). Since the LETKF analysis was shown to be better than 3DVAR analysis when using all operational observations except radiances, we now assess the impact of adding AIRS retrievals.

In this study we use the same system as Szunyogh et al (2006) assimilating real non-radiance observations on the NCEP GFS, and add AIRS temperature retrievals provided by Chris Barnet. The 4-dimension version of LETKF (Hunt et al. 2004), assimilates the retrievals at their right time. The analyses and forecasts are compared with those from Szunyogh et al (2006), which do not assimilate AIRS data.

* *Corresponding author address:* Hong Li 4339 CSS, College Park, MD 20742; e-mail: Lhhelen@atmos.umd.edu

2. 3D- and 4D-LETKF

LETKF is an ensemble square-root filter in which the observations are assimilated to update only the ensemble mean (shown in equation (1)) while the ensemble perturbations are updated by transforming the forecast perturbations through a transform matrix (equation (2)) introduced by Bishop et al (2001). The details of the 3D-LETKF scheme can be found in Hunt et al (2006).

$$\bar{x}^a = \bar{x}^b + X^b \tilde{P}^a (HX^b)^T R^{-1}[y^o - h(\bar{x}^b)] \quad (1)$$

$$X^a = X^b[(K-1)\tilde{P}^a]^{1/2} \quad (2)$$

Here K is the total ensemble number, X^a, X^b are the analysis and forecast ensemble perturbations, respectively. X^b is updated every analysis time step, therefore the forecast error covariance $P^f = \frac{1}{K-1} X^b X^{b^T}$ is flow-dependent, more accurate than the constant forecast error covariance used in 3DVAR. $\tilde{P}^{a^{1/2}}$ is the square-root of the matrix $(K-1)\tilde{P}^a$ where $\tilde{P}^a$, the analysis error covariance in ensemble space, is given by

$$\tilde{P}^a = \left[(K-1)I + (HX^b)^T R^{-1} (HX^b)\right]^{-1} \quad (3)$$

which has dimension K by K, much smaller than both the dimension of the model and the number of observations. Thus, the LETKF performs the matrix inverse in the space spanned by the forecast ensemble members, which greatly reduces the computational cost.

The mean analysis state generated by this 3D-LETKF is the linear combination of the background ensemble states which best fits the available observations at analysis time (Hunt et al., 2006). 4D-LETKF modifies LETKF by seeking the linear combination of the ensemble trajectories that best fits the observations made within the assimilation window between analysis times (Hunt et al., 2004). Specifically, 4D-LETKF solves the following analysis equations:

$$\bar{x}_n^a = \bar{x}_n^b + X_n^b \tilde{P}^a \left(\sum_{l=1}^{n} (H_l X_l^b)^T R_l^{-1} [y_l^o - h_l(\bar{x}_l^b)] \right) \quad (4)$$

$$X_n^a = X_n^b [(k-1)\tilde{P}^a]^{1/2} \quad (5)$$

$$\tilde{P}^a = \left[(k-1)I + \sum_{l=1}^{n} (H_l X_l^b)^T R_l^{-1} (H_l X_l^b) \right]^{-1} \quad (6)$$

where the subscript l refers to the corresponding model state at time $l \in \{1,2,\ldots,n\}$, and n is the current analysis time (Harlim and Hunt, 2006). In this way, 4D-LETKF can assimilate observations at the correct time.

3. OBSERVATIONS

We assimilate two observational data sets in this study. The first set contains all operationally available data except radiances, the same set used by Szunyogh et al (2006). The second observation set is like the first set but augmented by the 3x3 degree resolution AIRS temperature retrievals, NESDIS.

The AIRS retrieval algorithm is a v5 emulation system, based on the operational version 4 (Susskind et al 2003, Susskind et al 2006) but headed toward version 5 (Chris Barnet, personal communication). The quality control flag resembles v4 qual_temp_mid=0 flag, but is applied to the whole column. In order to assimilate the AIRS temperature retrievals, we have to estimate their error statistics, but it is very difficult to estimate the error correlations among retrievals at different latitude and longitude locations. To simplify the problem, we ignore the error correlations but increase the error standard deviations to compensate for the reduced magnitude of the observation error covariance matrix.

4. EXPERIMENTAL SETUP

Using 4D-LETKF, we assimilate the first observation data set in January 2004 with the NECP GFS model (control run). We verify the analysis and forecasts against the NCEP T254/L64 analysis that uses all available operational observations. Then, we assimilate the second observation data set including the AIRS temperature retrievals for the same time period (AIRS run). We assess the AIRS impact by comparing the results from the AIRS run with the control run.

5. RESULTS

5.1 Experiment with vertically constant AIRS error standard deviation

The v4 AIRS retrievals have produced 1km tropospheric layer mean temperatures with an RMS error of 1K, in general (Susskind et al. 2006). As mentioned in section 3, we ignore the retrievals correlation and therefore have to increase the error standard deviation. For the first test, we assume the error standard deviation of temperature retrievals used in the observation error covariance matrix is 2K at all levels, doubling the estimated error of about 1K.

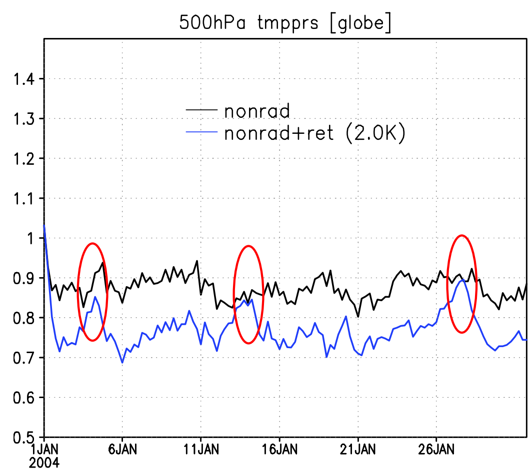

Fig.1. Time series of the global averaged analysis RMS error for 500hPa temperature field in January 2004, for the control run (black, assimilating all the NCEP operational non-radiance data) and the AIRS run (blue, adding AIRS temperature retrievals). The red ovals indicate days in which AIRS retrievals were missing.

Fig.1 shows the global averaged 500hPa temperature analysis RMS difference from the NCEP T254/L64 analysis for the control run (black) and the AIRS run (blue). There is a consistent reduction of errors when assimilating AIRS retrievals. It is remarkable that on some days, such as Jan 3rd, 13th and 27th, when the AIRS retrievals were missing, the AIRS run has almost the same RMS error as the control run. This further proves the significant positive impact of AIRS on the analysis. As expected, the AIRS data improved the analysis significantly more over the Southern Hemisphere. The improvement is about 30%. The results for the Northern Hemisphere show a smaller but still consistently positive impact in the analysis accuracy (Fig 2).

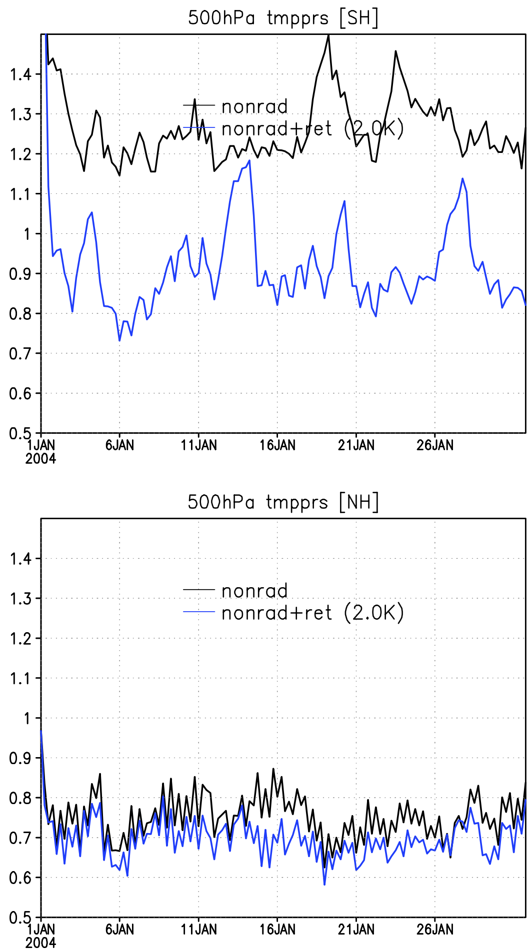

Fig. 2. As for Fig.1, except for the Southern Hemisphere (upper panel) and the Northern Hemisphere (bottom panel), separately.

Though only AIRS temperature retrievals were assimilated, the improvements are also seen in other variables. The temperature information benefits other variables through the evolving dynamical cross correlation between the observed variable and the others estimated by the LETKF. This is seen in Fig. 3 with an improved AIRS analysis for the 500hPa zonal wind field.

Fig.4 shows the AIRS impact at other levels. In general, the analysis of the AIRS run is more accurate than that of control run for most of the tropospheric levels. However, a negative impact is observed in high levels. Since the NCEP T254/L64 analysis has errors of its own, this may be partly due to lower accuracy at these levels. The degradation of positive impact in the near-surface levels suggests that the retrievals are less accurate near the surface.

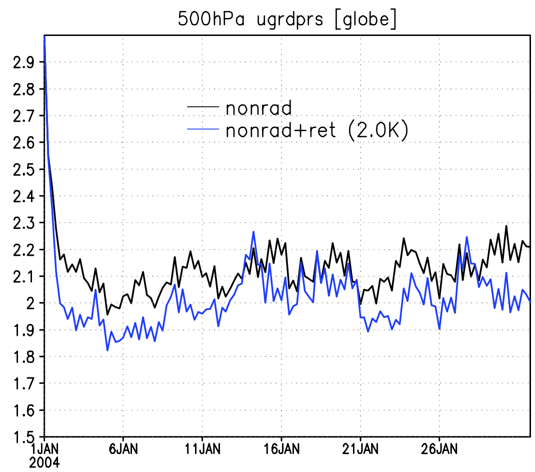

Fig.3. As for Fig.1, except for the zonal wind field.

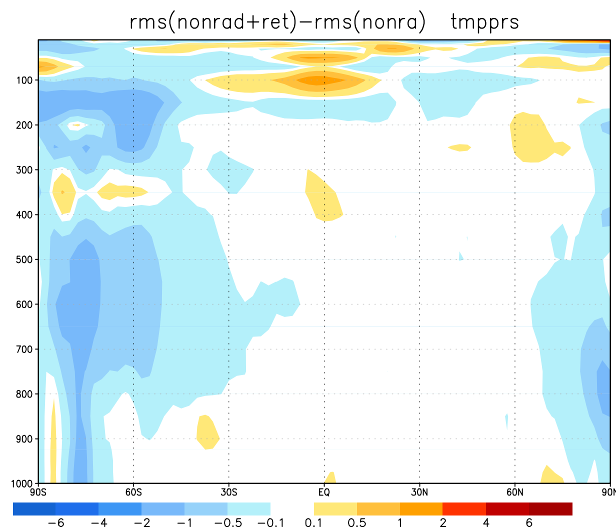

Fig.4. Time-longitude averaged vertical cross-section of the RMS error of the AIRS run minus the RMS error of the control run for temperature field, averaged over the last 10 days in January 2004.

We have shown a significant beneficial impact of AIRS retrievals on the analysis, at least compared with the higher resolution operational NCEP analysis. We now test their impact on forecast skill. Fig. 5 is the global averaged 500hPa temperature RMS error for 48-hour forecasts for the control run and the AIRS run. It demonstrates that 48-hour forecasts consistently show a positive impact from the assimilation of AIRS temperature retrievals. This improvement is also seen at other levels and for the other variables (not shown).

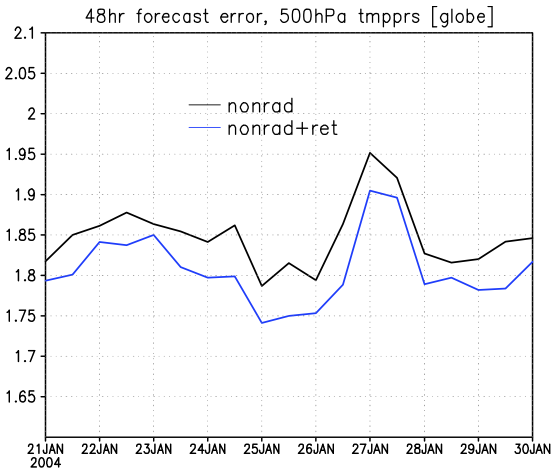

Fig 5. As for Fig.1, except for the 48-hours forecast and only showing the last 10 days.

5.2 Experiment with vertically dependent AIRS error standard deviation

In section 5.1, we showed promising analysis and forecast results in the upper and middle troposphere when assimilating AIRS retrievals, but the results were degraded near the surface. In section 5.1 we assumed a constant 2K retrieval error for the whole column. Though simple, this assumption is crude because in reality the AIRS temperature errors vary with height. It appears that 2K is a reasonable assumption for retrievals in the upper-to-middle troposphere (given that we neglected horizontal error correlations), but the RMS profile shown in Fig 5 from Susskind et al. (2003) suggests that the errors are larger in the low troposphere and high stratosphere.

We performed another preliminary experiment by using level-dependent AIRS temperature errors. We retain 2K errors in the upper and middle troposphere, but increase them to about 2.4K at low troposphere and high stratosphere, and 2.6K near the surface, values roughly proportional to the RMS profile in Susskind et al (2003).

Fig 6 shows 850hPa global averaged analysis error for the control run, the AIRS run with 2K error and the AIRS run with the level-dependent error. Although the result of the AIRS run with 2K error were better than the control run, the improvement was not as large as that at 500hPa (Fig 1). By increasing the errors to be 2.4 K~2.6 K at the low to surface levels, the result of assimilating AIRS temperature is further improved at 850hPa. This improvement is demonstrated further in the time-longitude averaged RMS error difference between these runs (Fig 7). We find that some negative AIRS impact observed in Fig 4 are now neutral or positive impacts. The AIRS impact is positive almost everywhere below 400hPa. The negative impact at the high levels still appears, but its magnitude is reduced, although, again this error could be arising from the errors in the verification analysis.

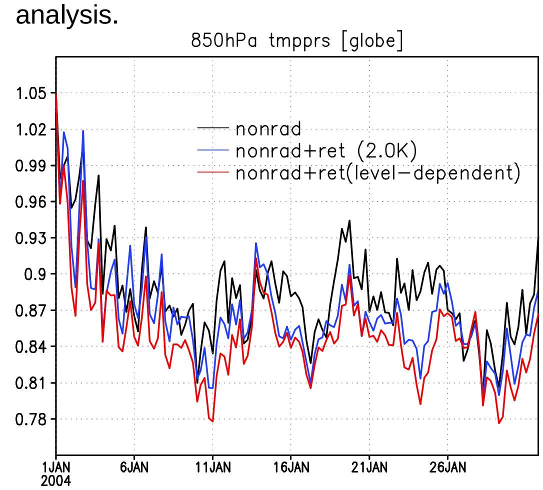

Fig.6. Time series of the global averaged analysis RMS error for 850hPa temperature field from the control run (black), the AIRS run assuming a 2K error at all levels (blue), and the AIRS run assuming a level-dependent error (red).

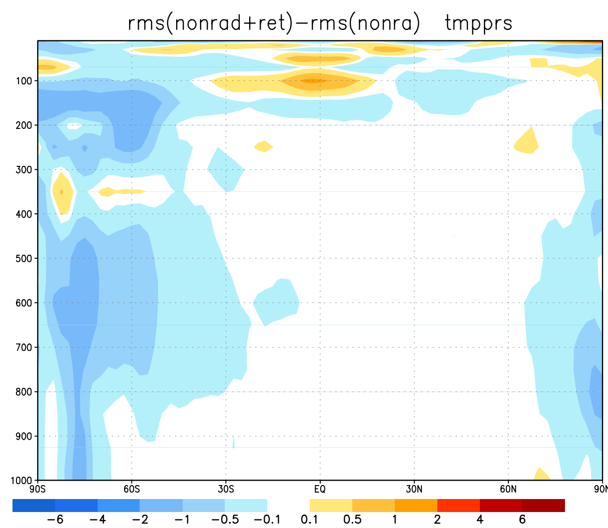

Fig.7. As for Fig.4, except for assuming a level dependent retrieval error in the AIRS run.

6. CONCLUSON, DISCUSSON AND FUTURE WORK

The AIRS temperature retrievals have consistent positive impact on both analysis and forecast, found not only in the temperature field but also in the other variables. This positive impact is biggest in Southern Hemisphere but still significant in Northern Hemisphere.

We have arbitrarily assumed either constant 2K errors or increased the errors near the surface and in the upper levels. The results indicate that there is a significant impact from the assumed error levels. We have also neglected the correlation of retrieval errors from different locations. We are exploring the possibility to estimate the observation error covariance of the retrievals based on the covariances of analysis and observational increments (Desroziers et al. 2005). This can be done online within the LETKF, and preliminary results with simpler models are encouraging (Kalnay et al. 2006).

We are planning to assimilate AIRS retrievals generated by other algorithms with the 4D-LETKF. The MIT stochastic/neural network retrievals trained with ECMWF analyses, are apparently able to cloud clear AIRS radiances and produce retrievals with accuracy that does not degrade with cloud cover. William Blackwell has offered to provide us with these retrievals for January-February 2004 to test their impact on analyses and forecasts, a more robust way to test whether these retrievals do indeed contain more information than the standard retrievals.

In addition to the AIRS temperature retrievals, we will assimilate the humidity to see if the AIRS humidity product further benefits the analysis and forecasting (Liu et al. 2006).

Finally, we plan to perform assimilation of radiances and compare with the retrievals and with the assimilation of cloud-cleared radiances obtained in the process of computing the retrievals. Clear radiances depend only on instrument errors, which are less correlated than the retrievals. The LETKF has an advantage for assimilating radiance observations because it does not require either the adjoint or the linear tangent models of the radiance transform model: it only needs differences between nonlinear integrations. Cloud-cleared radiances may have less correlated errors than the retrievals, but are much more abundant than the clear radiances currently used in operations.

REFERENCES

Anderson, J. L. 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903.

Bishop, C. H., B.J. Etherton, and S.J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, **129**, 420-436.

Chahine, M.T, and Coauthors 2006: Improving weather forecasting and providing new data on greenhouse gases. *Bulletin of the American Meteorological Society*, **87**, 911-926.

Desroziers G., L. Berre, B.Chapnik, and P.Poli, 2005: Diagnosis of observation,background and analysis error statistics in observation space. *Quart. J. Roy. Meteor. Soc,* **131**, 3385-3396.

Harlim, J. and B.R. Hunt, 2006: Four-dimensional local ensemble transform Kalman filter: Variational formulation and numerical experiments with a global circulation model. Submitted to *Tellus A.*

Houtekamer, P.L, H.L. Mtchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with the ensemble Kalman filter: Results with real observations. *Monthly Weather Review*, **133**, 604-620.

Hunt, B. R., E. Kalnay, E. Kostelich, E. Ott, D.J. Patil, T. Sauer, I. Szunyogh, J. A. Yorke, A.V. Zimin, 2004: Four-Dimensional Ensemble Kalman Filter. *Tellus*, **56A**, 273-277.

Hunt, B. R., E. Kostelich, and I. Szunyogh, 2006: Efficient Data Assimilation for Spatiotemporal Chaos: a Local Ensemble Transform Kalman Filter. In preparation.

Kalnay, E., H. Li and T. Miyoshi, 2006: Adaptive estimation of background and observation errors within Local Ensemble Transform Kalman Filter. Submitted to the 87th AMS annual meeting.

Liu, J. and E. Kalnay, 2005: Assimilating specific humidity observations with Local Ensemble Transform Kalman Filter. Submitted to the 87th AMS annual meeting.

Liu, J., E. Fertig, H. Li, I. Szunyogh, B. R. Hunt, E. Kalnay, E. Kostelich, and R. Todling, 2006: Comparison between Local Ensemble Transform Kalman Filter and PSAS in NASA finite volume GCM. In preparation.

Marshall, J.Le, J. Jung, J. Derber, M. Chahine, R.Treadon, S. J. Lord, M. Goldberg. W.Wolf, H. C. Liu, J. Joiner, J. Woollen, R. Todling, P.Van Delst and Y. Tahara, 2006: Improving global analysis and forecasting with AIRS.

Bulletin of the American Meteorological Society, **87**, 891-894.

Susskind, J., C.D. Barnet and J.M. Blaisdell 2003. Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. *IEEE Trans. Geosci. Remote Sens.* **v.41**, 390-409.

Susskind, J., C.D. Barnet, J.M. Blaisdell, L. Iredell, F. Keita, L. Kouvaris, G. Molnar and M.T. Chahine, 2006. Accuracy of geophysical parameters derived from Atmospheric Infrared Sounder/ Advanced Microwave Sounding Unit as a function of fractional cloud cover. *J. Geophys. Res*. **v.111** doi:10.1029/2005JD006272, 19 pgs.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, 2005: Assessing a Local Ensemble Kalman Filter: Perfect Model Experiments with the NCEP Global Model. *Tellus*, **57A**, 528-545.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, and J. A. Yorke, 2006: Assessing a Local Ensemble Kalman filter: Assimilating real observations with the NCEP global model. In preparation.

Whitaker, J. S., G. P. Compo, X. Wei, and T.H. Hammil, 2004: Reanalysis without Radiosondes using Ensemble Data Assimilation. *Monthly Weather Review*, **132**, 1190-1200.