3.4 Assimilating specific humidity observations with Local Ensemble Transform Kalman Filter

Junjie Liu[1*], E. Kalnay[1]
[1]University of Maryland, College Park, MD

1. Introduction

The atmospheric humidity is directly related with cloud coverage and precipitation processes, and with latent heat release. Therefore, accuracy of the initial humidity field is a necessary condition to improve the forecast of the other meteorological fields, especially in the cloudy and precipitating regions. Since water vapor is an active absorption gases in the infrared frequency, the amount of water vapor in the atmosphere determines the magnitude of radiances measured by some bands of satellites. The assimilation of radiances, related to both temperature and humidity through the radiative transfer model, requires accurate humidity information. Because of the important role the humidity plays in the hydrological cycle and the atmospheric circulation, the humidity data assimilation is an important topic.

In spite of the important role the humidity field plays in NWP and climate studies, the current humidity observation assimilation methods are behind of the other mass field, such as temperature, and wind assimilation in several aspects. First, in the current variational assimilation system, the background constraint is uni-variate in specific humidity analysis (e.g. ECMWF). The specific humidity is not fully coupled with the other dynamical variables. Second, the background error-variances are not cycled but are given through an empirical relationship with T and q of the background state (Rabier et al. 1998).

The difficulty in humidity data assimilation is related with the special characteristics of humidity field. The humidity has large variability in both space and time. The large variability in small spatial scales makes the interpolation of the humidity information difficult, which requires subtle treatment. The temporal variation requires the time changing error statistics. However, the current operational data assimilation schemes, which are variational data assimilation schemes (e.g. NCEP, ECMWF), assume constant background error covariance. Furthermore, the humidity field is a variable that has the least Gaussian error distribution among the other dynamical variables, while Gaussian error distribution is an assumption of most data assimilation schemes. Further complicating the problem is the poor quality of the humidity observations and the large model error related with the parameterization process.

In contrast with variational methods, ensemble Kalman filter estimates and uses the time-changing background error statistics, which are estimated from ensembles (Anderson, 2001, Whitaker and Hamill, 2002, Bishop et al., 2001, Tippett et al., 2003). Furthermore, ensemble Kalman filter automatically estimates the covariance statistics between different variables. The multivariate analysis in which the mismatch between background and observed humidity implies an accompanying adjustment of the other dynamical fields, especially the vertical velocity, creates the consistent initial condition between humidity field and the other dynamical fields. The forecast initialized with such initial condition should have less spin-up, which seriously affects the short-term forecasts. These characteristics are inherent to ensemble Kalman filter, while variational methods have to do extra work to reduce spin-up. Because of these characteristics, multivariate humidity data assimilation should be tested with ensemble Kalman filter. Ensemble Kalman filter has been proven to be a more accurate data assimilation scheme than 3D-Var (Whitaker et al., 2006, Liu et al., 2006), and comparable with 4D-Var with the assimilation of temperature, winds, and pressure fields (Houtekamer, et al., 2006, pers. communication). However, there are no specific studies about humidity data assimilation with ensemble Kalman filter. In this study, we will use Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2006), which is very efficient and accurate.

This paper is organized as follows. Section 2 will give a brief review of the model we use. Section 3 briefly describes the 3D-Var developed

[1*] *Corresponding author address, 2113 CSS, College Park, MD. Email: jjliu@atmos.umd.edu*

by Miyoshi (2005) on the SPEEDY model, and the local ensemble transform Kalman filter. Section 4 includes preliminary results, and results and discussions will be in section 5.

2. SPEEDY model

The SPEEDY model (Molteni 2003) is a recently developed atmospheric general circulation model (AGCM) with simplified physical parameterization schemes that are computationally efficient, but that maintain the basic characteristics of a state-of-the-art AGCM with complex physics. It has triangular truncation T30 with 7 sigma levels. The dynamical variables include zonal and meridional wind components, temperature, specific humidity, and surface pressure.

3. Implementation of 3D-VAR and LETKF on SPEEDY model

3.1 Description of 3D-Var Miyoshi (2005) in SPEEDY model

Miyoshi (2005) followed the 3D-Var formulation of Barker et al., (2004). He considers the error standard deviation, spatial error correlation for each variable, and inter-variable correlation based on the geostrophic balance. These statistics were computed using the NMC method (Parrish and Derber, 1992). The humidity analysis is univariate.

3.2 Brief description of LETKF scheme

LETKF (Hunt, 2005, Hunt et al, 2006) is an ensemble square-root filter in which the observations are assimilated simultaneously to update only the ensemble mean, as shown by (3.2.1), while the ensemble perturbations are updated by transforming the forecast perturbations through a transform matrix term $[(k-1)\tilde{P}^a]^{1/2}$ in (3.2.2), introduced by Bishop et al (2001). The basic formulas used in the LETKF are given by

$$\overline{\mathbf{x}}^{\mathbf{a}} = \overline{\mathbf{x}}^{\mathbf{b}} + \mathbf{X}^{\mathbf{b}}\tilde{\mathbf{P}}^{\mathbf{a}}(\mathbf{H}\mathbf{X}^{\mathbf{b}})^{\mathbf{T}}\mathbf{R}^{-1}[\mathbf{y}^{\mathbf{o}} - \mathbf{h}(\overline{\mathbf{x}}^{b})] \qquad (3.2.1)$$

$$\mathbf{X}^{\mathbf{a}} = \mathbf{X}^{\mathbf{b}}[(\mathbf{k}-1)\tilde{\mathbf{P}}^{a}]^{1/2} \qquad (3.2.2)$$

Here $\overline{\mathbf{x}}^a$ ($\overline{\mathbf{x}}^b$) is the analysis (forecast) ensemble mean, and $\mathbf{X}^a$ ($\mathbf{X}^b$) are the analysis (forecast) ensemble perturbation matrices, whose columns are the difference of ensemble member and ensemble mean. The transform matrix $[(k-1)\tilde{\mathbf{P}}^a]^{1/2}$ is the square-root of the matrix $(k-1)\tilde{\mathbf{P}}^a$ where $\tilde{\mathbf{P}}^a$, the analysis error covariance in ensemble space is given by

$$\tilde{\mathbf{P}}^a = \left[(k-1)\mathbf{I} + (\mathbf{H}\mathbf{X}^b)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{X}^b)\right]^{-1} \qquad (3.2.3)$$

The dimension of $\tilde{\mathbf{P}}^a$ is k*k, where k is the ensemble size, a number much smaller than the dimension of the model state. In this study, we use 20 ensemble members. Thus, the LETKF performs the analysis in the space spanned by the forecast ensembles, which greatly reduces the computational cost. Furthermore, since the analysis is computed independently at each grid point, the LETKF computations can be performed in parallel.

4. Experimental design

In these perfect model simulations, the observations are obtained from the “truth”, which is a long time integration of the model, by adding Gaussian random errors. The standard deviation of the random error is 1 m/s for both zonal wind and meridional wind. It is 1hPa for surface pressure, 1K for temperature. The specific humidity observation has Gaussian observation errors with 0.1g/Kg standard deviation. The observations are uniformly distributed, with the observation available every 2 by 2 grid points. Overall, the observation coverage is 25% of the grid points. The initial condition is the truth at the 06Z01Jan, 2003 plus random errors with triple observation error standard deviation.

In order to test the effectiveness of the specific humidity observation assimilation with LETKF, we designed four types of experiments. We name the first experiment as **noq** experiment in which the specific humidity is not assimilated at all: the final “analysis” of the specific humidity is copied from background. This is the approach used so far in most EnKF experiments (e.g., Szunyogh et al, 2005, Whitaker et al, 2006). The second one is named as **uniq** experiment, in which the specific humidity is updated by itself, and not affected by the other variables (u, v, T and ps), which is the way it is done operationally in most centers. The third one is the **passiveq** experiment in which the specific humidity is updated by the other variables (winds, temperature, and surface pressure), but the other variables are not updated by the specific humidity, which is a conservative multivariate approach allowing for the possibility that the humidity errors are correlated with, for example, wind errors due to advection, but the use of these correlations may be not robust enough to improve the wind analysis. The fourth one is the experiment with fully coupled variables (**full-var**), in which all the dynamical variables are included in one vector. Table 1 summarizes the characteristics of these four experiments. Since the inter-variable correlation between specific humidity and other variables is not considered in the background error covariance used in current 3D-Var we cannot carry out the **full-var** and **passiveq** experiments with 3D-Var. For 3D-Var we only carried out the **noq** and **uniq** experiments. In the next section, we will discuss the interactive impact between the humidity and the other variables during analysis by comparing the RMS errors from each experiment. We will also discuss the different characteristics of LETKF and 3D-Var by comparing the impact of specific humidity observations assimilation on the other dynamical variables.

Experiment Name	Observed variable	Dynamical variables	Updated variables
noq	u, v, T, Ps	u, v T, Ps	u, v, T, Ps
uniq	u, v, T, Ps	u, v, T, Ps	u, v, T, Ps
	q	q	q
passiveq	u, v, T, Ps	u, v, T, Ps	u, v, T, Ps
	u, v, T, q, Ps	u, v, T, q, Ps	q
Full-var	u, v, T, q, Ps	u, v, T, q, Ps	u, v, T, q, Ps

Table 1 Four types of experiments carried out by LETKF. Dynamical variables represent the variables included in the background vector. The updated variables are the final analysis variables.

5. Results

5.1 The accuracy of humidity analysis

Without the assimilation of specific humidity (**noq** experiment), the accuracy of specific humidity is comparable in both 3D-Var and LETKF, since in both schemes the specific humidity is copied from background. With the assimilation of specific humidity observations (**uniq** experiment), the specific humidity analysis is improved in both 3D-Var and LETKF (Fig. 1). The improvement is much more significant in LETKF than that of 3D-Var due to the higher accuracy of LETKF assimilation. With the same observations, LETKF can extract more information than 3D-Var (Liu et al, 2006).

The accuracy of specific humidity analysis is further slightly improved in the **full-var** experiment with LETKF scheme (Fig. 1b). The improvement in the **passiveq** experiment is due to the positive impact from the observations of the other variables. The specific humidity can extract useful information from the other variables because of the accurate estimation of the covariance between specific humidity and the other variables. The improvement in the **full-var** is from the analysis accuracy of the other variables, which in turn create more accurate specific humidity analysis. Contrary to 3D-Var, the LETKF analysis errors are still decreasing after one month integration.

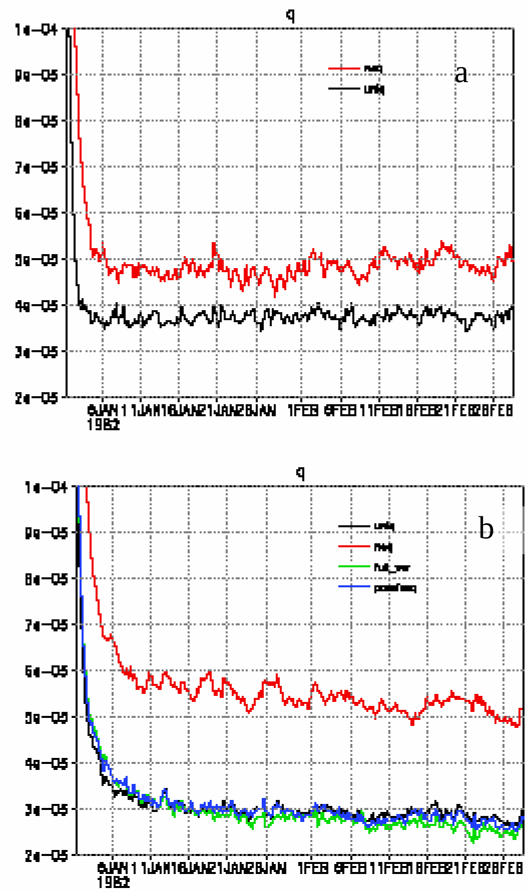

Fig. 1 500hPa specific humidity (kg/kg) RMS error evolved with time for 3D-Var (top panel) and LETKF (bottom panel). The red line is the result of **noq** experiment, the black line is the result of **uniq** experiment, the blue line is from **passiveq**, and the green line is for **full-var** experiment.

5.2 The impact of specific humidity observations on other variables

Though specific humidity analysis is improved from the assimilation of specific humidity observations (Fig. 1a) in 3D-Var, the other variables, such as zonal wind (Fig. 2a), meridional wind and temperature (not shown) are only marginally improved.

In contrast to 3D-Var, the assimilation of specific humidity observations in LETKF improves substantially the analysis of other variables as well (Fig. 2b). Compared with **noq** experiment, the zonal wind analysis is greatly improved in the **uniq** experiment. Since the accuracy of specific humidity in **passiveq** experiment is better than that of **uniq** experiment, the analysis of zonal wind is also further improved due to the positive impact of specific humidity analysis during forecast. The specific humidity can improve the analysis of other variables not only through forecast, but also through analysis, as the results of **full-var** experiment show. The accuracy of zonal wind analysis is best in the **full-var** experiment, in which the specific humidity observations affect the analysis of other variables through the covariance between specific humidity and the other variables.

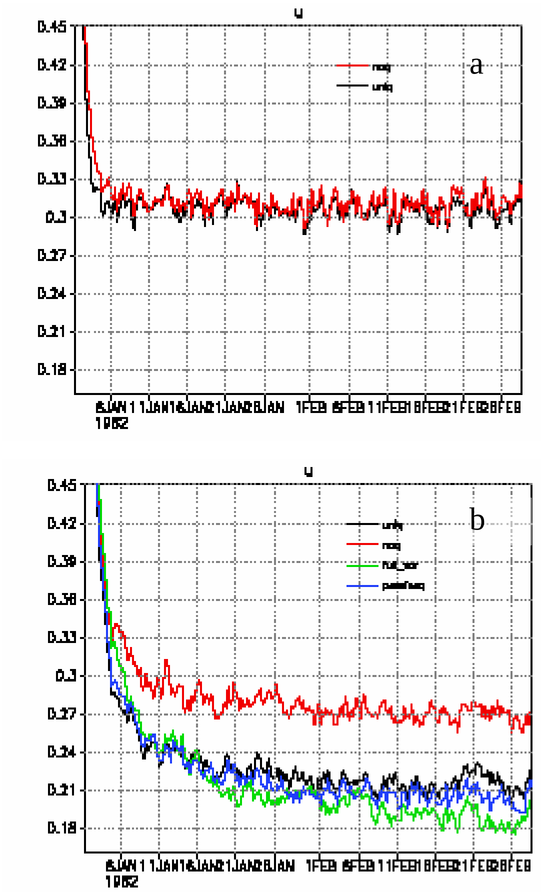

Fig.2 500hPa zonal wind (m/s) RMS error evolved with time for 3D-Var (top panel) and LETKF (bottom panel). The red line is the result of **noq** experiment, the black line is the result of **uniq** experiment, the blue line is from **passiveq**, and the green line is for **full-var** experiment.

5.3 The mechanism behind the improvement due to specific humidity data assimilation

In this section, we will try to isolate the factors influencing the analysis results by comparing the RMS error structure difference between different experiments.

Fig. 3a shows zonal mean of time average specific humidity RMS error, which reveals that the specific humidity has large errors over the Tropics and lower levels. The error gradient is similar with the gradient of the true field. Where the magnitude of specific humidity is larger, the error is also larger. Another characteristic is that the specific humidity has relatively larger RMS error around 200hPa, where the jet stream lies (Fig. 4a) and the wind speed and gradient is larger. Since specific humidity acts like a tracer gas, it is affected by the wind speed, especially when the wind speed is very large. While we have not considered the wind effect in **uniq** experiment, the RMS error is relatively large over that region. The wind effect is considered in the **passiveq** experiment, so the specific humidity analysis is much more accurate in the **passiveq** experiment (Fig. 3b) compared with **uniq** experiment, especially at the level where the jet stream lies. In turn, the improvement of the specific humidity analysis improves the wind analysis by improving the wind forecast field, especially over the jet stream and the large wind speed region (Fig. 4c). Though temperature is also used to update the specific humidity analysis in the **passiveq** experiment, we believe that the impact from temperature is smaller than the wind because the correlation between specific humidity and temperature is very small (Dee and Da silva, 2003). Compared with **noq** experiment, the specific humidity in the **uniq** experiment is improved everywhere (not shown), so does the wind field (Fig. 4b). The possible reason is that the introduction of the specific humidity observation into the analysis system in the **uniq** experiment makes the combination of the ensemble member more optimal. Therefore, it produces more accurate analysis for other variables compared with **noq** experiment.

Compared with **uniq** experiment, the **full-var** experiment has large improvement not only over the place where the **passiveq** experiment shows improvement, but also the other regions, especially below 200hPa (Fig. 3c, Fig. 4d). Compared with **passiveq** experiment, the zonal wind analysis is better, which is due to the impact of specific humidity observation on the zonal wind analysis during data assimilation. The result reveals that the error covariance estimated from the ensembles represent the true covariance between specific humidity and zonal wind, which further support the possibility of improving the wind analysis field from specific humidity observations. In turn, the improvement of zonal wind improves the specific humidity field (Fig. 3c).

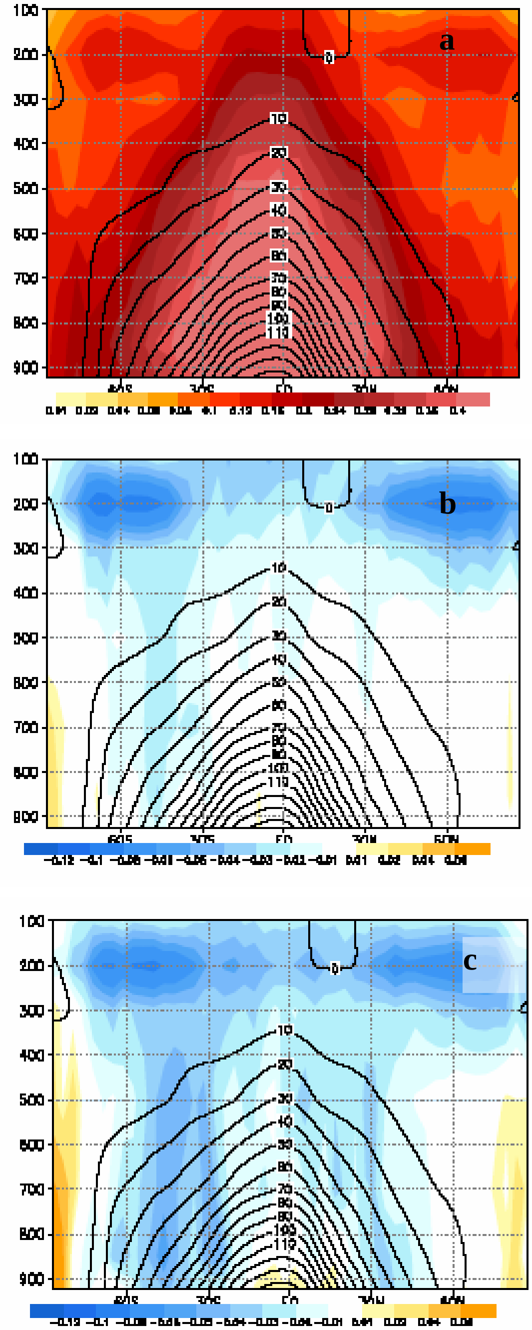

Fig. 3 a. Zonal average of the time mean specific humidity analysis error (shaded, Unit: 10^{-1}g/kg) for **uniq** experiment; b. the zonal average of the time mean specific humidity RMS error difference (shaded, Unit: 10^{-1}g/kg) between **passiveq** experiment and **uniq** experiment; c, the RMS error difference (shaded, Unit: 10^{-1}g/kg) between **full-var** experiment and **uniq** experiment. The contour is the time average true specific humidity (Unit: 10^{-1}g/kg)

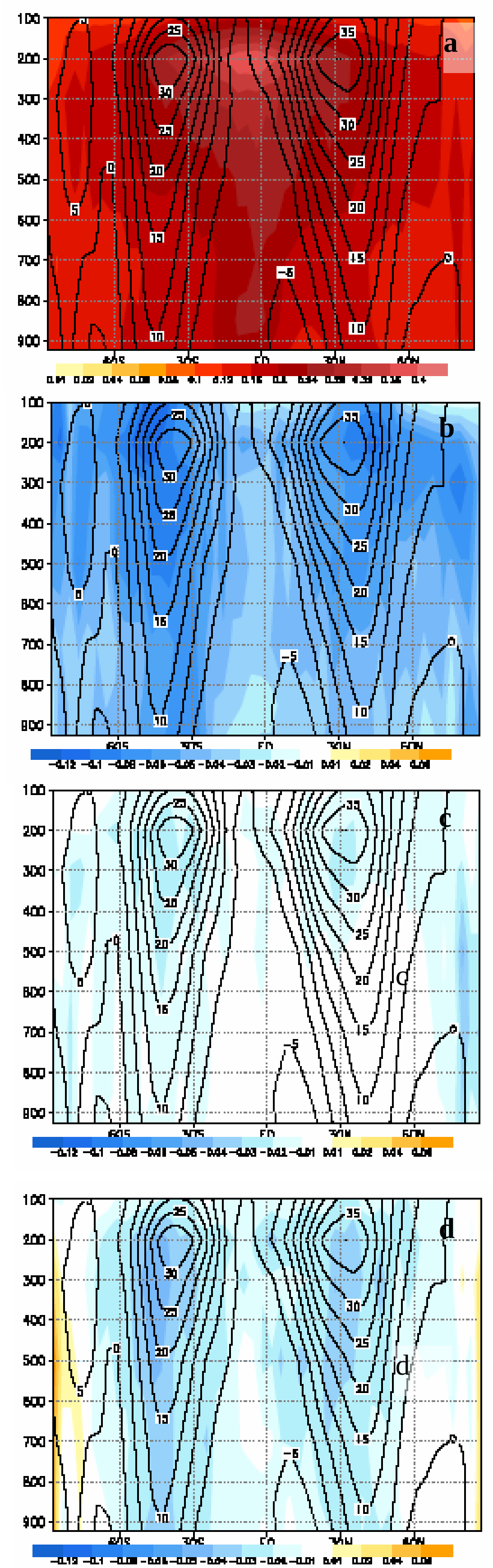

Fig. 4 a. Zonal average of the time mean zonal wind analysis error (shaded, Unit: m/s) for **uniq** experiment; b. the RMS error difference between **uniq** experiment and **noq** experiment. The contour is the time average true specific humidity (Unit: m/s); c. the RMS error difference (shaded, Unit: m/s) between **passiveq** experiment and **uniq** experiment; d, the RMS error difference (shaded, Unit: m/s) between **full-var** experiment and **uniq** experiment; The contour is the time average true specific humidity (Unit: m/s).

6. Conclusions and discussion

The humidity field is important in both weather forecast and climate studies. At the same time, the assimilation of humidity observations is a difficult problem that lags behind the other variables. In this paper, we discussed the possibility and results of specific humidity assimilation using LETKF scheme.

Unlike 3D-Var, LETKF connect the specific humdity and the other variables automatically through the estimated covariance between specific humdity and other variables. Therefore, the specific humidity can be fully coupled with the other variables. The results show that the specific humidity observation not only has big impact on the assimilation of itself, it can also improve the analysis of other variables, especially the wind fields. At the same time, acting like tracer gases, specific humidity extract useful information from the other variables as well.

Though we get very promising results here, we realize that the observation error distribution of the specific humidity is not very realistic, which is least Gaussian in reality. At the same time we also have to consider the effect of the model error. In the future, we plan to deal with the non-Gaussian observation error problem by using a new variable, pseudo-relative humidity (Dee and DaSilva, 2003), which has more Gaussian error distribution. We will try to estimate the parameter related with the precipitation process along with data assimilation to deal with model error.

References

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

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

Andersson, E., E. Holm, P. Bauer, A. Beljaars, G. A. Kelly, A. P. McNally, A. J. Simmons, J-N. the paut and A. M. Tompkins, 2006: Analysis and forecast impact of the main humidity observing systems. ECMWF Tech Memo., 493, pp 22.

Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial results. *Mon. Wea. Rev.*, **132**, 897-914.

Bengtsson, L. and K. I. Hodges, 2005: On the impact of humidity observations in numerical weather prediction, *Tellus*, 57A, 701-708.

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

Chen, M., R. B. Rood, and J. Joiner, 1999: Assimilating TOVS humidity into the GEOS-2 data assimilation system. *J. of Climate*, **12,** 2983-2995

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophy. Res*., **99**, 1043-1062.

Dee, D. P., Da Silva, A., 2003: The choice of variable for atmospheric moisture analysis. *Mon. Wea. Rev*., **131**, 155-171

Elliot, W. P., and D. J. Gaffen, 1991: On the utility of radiosonde humidity archives for climate studies. *Bull. Amer. Meteor. Soc.*, **72,** 1507-1520.

Hamill, T., Whitaker, J., and Synder, C., 2001: Distance-dependent filtering of the background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev*., **129**, 2776-2790.

Holm, E., E. Anderson, A. Beljaars, P. Lopez, J-F. Mahfouf, A.J. Simmons and J.N. Thepaut, 2002: Assimilation and modeling of the hydrological cycle: ECMWF's status and plans. ECMWF Tech. Memo., 383, pp55.

Hou, A. Y., D. Ledvina, A. da Silva, S. Zhang, J. Joiner, R. Atlas, G. Huffman, and C. Kummerow, 2000a: Assimilation of SSM/I-derived surface rainfall and total precipitable water for improving the GEOS analysis for climate studies. *Mon. Wea. Rev.*, **128**, 509-537.

Houtekamer, P.L. and Mitchell, H.L., 2001: A sequential Ensemble Kalman Filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

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

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., 2003: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 341pp.

Liu, J., E. Fertig, H. Li, I. Szunyogh, B. R. Hunt, E. Kalnay, E. Kostelich, and R. Todling, 2005: Application of Local Ensemble Transform Kalman Filter: perfect model experiments with NASA fvGCM model. 86th AMS Annual Meeting. Jan26-Feb2, 2006, Atlanta. (Peer reviewed manuscript in preparation).

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Climate Dyn.*, **20**, 175-191.

Miyoshi, T. 2005: Ensemble Kalman filter experiments with a primitive-equation global model, PhD Thesis at the University of Maryland.

Ott, W. R., 1995: *Environmental Statistics and Data Analysis*. Lewis Publishers, 313pp

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.*, 120, 1747-1763.

Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Undén, J. Eyre, A. Hollingsworth and F. Bouttier, 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part 2 : Struture functions. *Q. J. R. Meteorol. Soc.*, **124**, 1809-1829

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

Tippett, M., Anderson, J., Bishop, C., Hamill, T., and Whitaker, J., 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485-1490.

Whitaker, J. S., and Hamill, T. H., 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913-1924.

Whitaker, J. S., Compo, G. P., Wei, X., and Hamill, T. H., 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190-1200.

Whitaker, J. S., Hamill, T. H., Wei, X., Song, Y. C., and Toth, Z., 2006: Ensemble data assimilation with NCEP global forecast system. *Submitted.*