U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5200 Auth Road
Camp Springs, MD 20746

Office Note 465

Estimating Observation Impact Signals in NCEP GSI using the Lanczos Method

Mozheng Wei* and Manuel S. F. V. De Pondeca
IMSG at NOAA/NWS/NCEP, Environmental Modeling Center, Camp Springs, Maryland

Zoltan Toth
NOAA/OAR/ESRL/GSD, Forecast Applications Branch, Boulder, Colorado

David Parrish
NOAA/NWS/NCEP, Environmental Modeling Center, Camp Springs, Maryland

October 25, 2010

* Corresponding author.

E-mail: Mozheng.Wei@noaa.gov

ABSTRACT

Despite the tremendous progress that has been made in data assimilation (DA) methodology, observing systems which reduce observation errors, and model improvements which reduce background errors, the analyses produced by the best available DA systems are still far from the truth. Analysis error and error covariance are important since they describe the accuracy of the analyses, and directed related to the future forecast errors, i.e. the forecast quality. In addition, analysis error covariance is critically important in building an efficient ensemble forecast system (EFS).

Estimating analysis error covariance in an ensemble based Kalman filter DA is easier in theory, but it is always challenging in variational DA systems which have been in operation at most NWP centers. In this study, we use the Lanczos method in the NCEP Gridpoint Statistical Interpolation (GSI) DA system to look into other important aspects and properties of this method which were not exploited before. We apply this method to estimate the observation impact signals (OIS) which are directly related to the analysis error variances. It is found that the smallest eigenvalue of the transformed Hessian matrix converges to one as the number of minimization iterations increases. When more observations are assimilated, the converging speed becomes slower and more eigenvectors are needed to retrieve the observation impacts. It is also found that the OIS over data-rich regions can be picked up easily by the eigenvectors with dominant eigenvalues.

Since only a limited number of eigenvectors can be computed due to computational expense, the OIS is severely underestimated, and the analysis error variance is consequently overestimated. It is found that the mean OIS values for temperature and wind components at typical model levels are increased by about 1.5 times when the number of eigenvectors is doubled. We have proposed four different calibration schemes to compensate for the missing trailing eigenvectors. Results show that the method with calibration for smaller number of eigenvectors cannot pick up the observation impacts over the regions with less observations as well as an “ideal case” with a large number of eigenvectors. But proper calibrations do enhance and improve the impact signals over regions with more data.

When compared with the observation locations, the method generally captures the OIS over regions with more observation data, including satellite data over the southern oceans. Over the tropics, some observation impacts may be missed due to the smaller background errors specified in the GSI, which is not related to the method. It is found that a lot more eigenvectors are needed to retrieve impact signals that resemble the banded structures from satellite observations, particularly over the tropics. Another benefit from the Lanczos method is that the dominant eigenvectors can be used in preconditioning the conjugate gradient algorithm in the GSI to speed up the convergence.

1. Introduction

In recent years, there have been many active research and developments in data assimilation (DA) method, such as 4D-Var (Derber 1987; Rabier *et al.* 2000) and ensemble Kalman filters (Bishop *et al.* 2001; Anderson 2001; Whitaker and Hamill 2002, Tippett *et al.* 2003; Zupanski 2005; Whitaker *et al.* 2007; Kalnay *et al.* 2007; Szunyogh *et al.*, 2008). More observations, more accurate observing systems, and the improved DA systems have played key roles in providing more accurate initial conditions for numerical weather prediction (NWP) models. These developments have improved the weather forecasts significantly, particularly over the short and medium ranges. In addition, there has also been tremendous progress in NWP model development, with more accurate physics parameterization schemes and increased computing power, which permits the use of higher resolution forecast models. In spite of all this progress in observing systems which reduce the

observation errors, and in model improvement which reduces the background errors, the analyses produced by the best available DA systems are still quite different from the true state.

Analysis error and error covariance are important for any DA systems since they describe the accuracies of the analysis fields generated by the DA systems. The analysis fields from DA are supposed to be the best possible estimates of nature. They are used as the initial states for NWP weather forecasts. The associated error and error covariance are also important because they are related to the future forecast error and error covariance as they evolve during the forecast time interval. Therefore, analysis error and error covariance directly determine the forecast errors and error covariance, i.e. the quality of forecast.

In addition, analysis error covariance is critically important in building an efficient ensemble forecast system (EFS). In the past decade, with the advances of new development and implementation of the EFS at some major NWP centers (Toth and Kalnay 1993, 1997; Molteni *et al*. 1996; Houtekamer *et al.* 1996), the forecasting capability has been improved to a new level compared with the traditional single deterministic forecast. These centers include the National Centers for Environmental Prediction (NCEP), the European Centre for Medium-Range Weather Forecasting (ECMWF), the Canadian Meteorological Center (CMC), the United Kingdom Meteorological Office (UKMO) and the Fleet Numerical Meteorological and Oceanography Center (FNMOC). Different ensemble systems and their performances have been evaluated and reviewed by various authors, e.g. Hamill *et al.* (2000), Wei and Toth (2003), Buizza *et al.* (2005), Bowler (2006), Wei *et al.* (2006, 2008), Leutbecher and Palmer (2008) and Park *et al.* (2008).

In ensemble forecasting, a limited number of different numerical forecasts are generated to represent the variability of our knowledge about the possible evolution of a dynamical system. A consensus in the scientific community is that the initial ensemble perturbations should sample the probability density function (PDF) that is represented by the analysis error covariance. Thus, in an operational environment at a NWP center, the analysis error covariance of the DA system that produces the initial analysis field for the forecasts should play a key role in generating the initial perturbations. So far, the analysis error variance/covariance has been used to a certain extent only by a few different ensemble methods at NWP centers. A recent description and comparison about how analysis error covariance is being used in ensemble initial perturbation techniques are given in Tables 1 and 2 in Wei *et al.* (2008).

There have been several efforts on estimating analysis error variance and covariance. For example, Buizza *et al.* (2005) suggested that the spread of initial states of three centers (NCEP, ECMWF and CMC) could be considered as a crude lower-bound estimate of the analysis error variance. Swanson and Roebber (2008) used the NCEP and ECMWF reanalysis data, and suggested that the reanalysis difference could be considered as a "shadow" of the analysis error. They found that the analysis difference contains certain aspects of the true flow-dependent analysis error and has significant impact on the short-time forecast skills in downstream regions. Similarly, Langland *et al.* (2008) looked at the differences between the NCEP and FNMOC analyses from January 1 to June 30, 2007. The authors found that the difference and root mean of the squared daily differences in 500hPa temperature are closely related to the distribution of radiosonde observations. The large differences between the two analyses were found to be associated with the regions with mostly satellite observations. Park *et al.* (2008) studied the ensemble performance from TIGGE (the THORPEX Interactive Grand Global Ensemble) data. They argued that the mean analysis from different centers will probably be the best to be used as a reference analysis in comparing the performance of an ensemble from each center. The analysis error could be estimated from the deviation between that analysis and the mean of centers. Bowler *et al.* (2008) also argued that the mean of analyses from multi-centers is generally better than the analysis from any one center.

By using the analysis data from NCEP, ECMWF, UKMO, CMC and FNMOC, Wei *et al.* (2009) introduced a new method for estimating the analysis error variance. The method computes the anomaly of each center's analysis by removing the long term mean using a recursive filter. The spread over the average anomaly (SPA) from different centers is then computed. These authors found that the time averaged distribution of SPA is even more related to the observation network density, compared with the spread around the center mean analysis. Furthermore, the typical systematic errors that appear in the spread around the center mean over high altitude regions are completely removed. The instantaneous values of SPA at any cycle for various variables bear a strong resemblance to the elusive analysis error variance.

While analysis error covariance in an ensemble based Kalman filter is readily available (Bishop *et al*. 2001; Anderson 2001; Whitaker and Hamill 2002, Tippett *et al*. 2003; Zupanski 2005; Whitaker *et al*. 2007; Kalnay *et al*. 2007; Szunyogh *et al*., 2008), it is not straightforward to get in 3D/4D-Var systems which have been in operation at most major NWP centers. In the NCEP global EFS, an ensemble transform with rescaling (ETR) has been used to generate the initial perturbations as described in detail in Wei *et al.* (2005, 2008). In the ETR method, the initial perturbations depend on the accurate analysis error covariance which should be provided by the DA system in operation. At NCEP, the operational DA system is the Gridpoint Statistical Interpolation (GSI), a three-dimensional variational analysis (3D-VAR) system (Derber *et al.* 1991; Parrish and Derber 1992; Wu *et al.* 2002; Derber *et al.* 2003; Kleist *et al.* 2009). In the variational analysis system, the analysis is found by minimizing the cost function, written in terms of the background fields, the observations, and their respective error covariance matrices. The analysis error covariance matrix in 3D/4D-VAR is determined by the background and observation error covariance matrices, and it can not be computed directly due to its huge memory demand.

Unlike the DA systems at NCEP and ECMWF, which are formulated in the model space, the Naval Research Laboratory Atmospheric Variational Data Assimilation System (NAVDAS) system at the US Naval Research Laboratory (NRL) is formulated on the observation space (Daley and Barker 2001; Xu *et al.* 2005). Daley and Barker (2001) proposed a local approximation in their NAVDAS to take advantage of the block-diagonal pre-conditioner and Cholesky decomposition of the diagonal blocks. The method produces an estimate of the analysis error variance at any location based on the observations and background within the observation prism in which the location is contained. It has been implemented successfully at FNMOC and NRL, and it generates the analysis error variance estimate from the NAVDAS for both global and regional ensemble forecast systems at FNMOC (McLay *et al*. 2007, 2008; Reynolds *et al.* 2008; McLay and Reynolds 2009; Bishop *et al*. 2009). Similar to the NCEP global EFS, the initial perturbations at FNMOC are generated by using the ET method.

For the 3D/4D-VAR systems in model space which have been implemented at most NWP centers, Fisher and Courtier (1995) proposed three approximate methods to estimate the analysis error variance. The most promising among them is the Lanczos method which was implemented in the ECMWF DA system. This method produces the analysis error variance estimates by computing the leading eigenvectors of the Hessian matrix. It takes advantage of the close link between the Lanczos method and the conjugate gradient method used in the minimization procedure. The authors carried out experiments using a simple univariate 3d-Var on a cyclic one-dimensional domain with 256 equally-spaced grid points. Some testing was also done in ECMWF 3D-Var system with 52 eigenvectors included.

The Lanczos method is already used in the NCEP Real-Time Mesoscale Analysis (RTMA) to estimate the analysis errors (Pondeca and Manikin 2009). The RTMA runs the GSI in 2D-Var mode to analyze near-surface observations over the continental USA and domains in Alaska, Hawaii, Puerto Rico and Guam (Pondeca and Manikin 2009; Manikin and Pondeca 2009). In the RTMA application, only the observations

near the surface are assimilated and only some of the surface variables are estimated, such as temperature, dew point, surface humidity at 2 meters, in addition to the 10-meter wind etc. In comparison to the application to the global 3D-Var GSI which is presented in this study, another advantage in the RTMA is that the analysis variables in the regional GSI are directly temperature and wind components, thus the analysis errors of these variables can be estimated directly, while the analysis variables in the global GSI are the stream function, unbalanced velocity potential etc, which makes the estimates of variances of wind components complicated. Furthermore, the RTMA is focusing on small regions, this avoids the pole problem during the transformations between different variables.

In this paper, we use the Lanczos method in the NCEP global 3D-Var GSI DA system to study some of its aspects and properties which were not exploited in Fisher and Courtier (1995) and Pondeca and Manikin (2009). The properties in question are very important not only in understanding the method but also in practical applications. In particular, we apply this method to estimate the observation impact signals (OIS) which are the square-root of the difference between the background and analysis error variances. In this method, only a small number of eigenvectors can be computed due to the computational expenses. Thus the error reduction is severely underestimated, and the analysis error variance is overestimated. In our study, we propose and compare four different calibration schemes to compensate for the missing trailing singular vectors. Without proper calibrations, the observation impacts computed using this method may be far away from reality. In addition, we study the sensitivity of the OIS to the number of observations employed in the GSI system. Also studied in this paper are the correlations between the observation locations and the OIS for different variables in an operational environment.

Section 2 provides a brief description and formulation of the analysis error variance and OIS. The dominant eigenvectors and eigenvalues of the transformed Hessian matrix are analyzed in detail in Section 3. Section 4 presents four different calibration schemes and their results, while the correlations between the observations and the OIS are exploited in section 5. Finally, discussion and conclusions are given in Section 6.

2. Introduction of basic formulation.

The NCEP GSI DA system is a unified global/regional three-dimensional variational DA system (Derber *et al.* 1991; Parrish and Derber 1992; Wu *et al.* 2002; Derber *et al.* 2003; Kleist *et al.* 2009). The cost function in the GSI to be minimized can be expressed as:

$$J(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y}_0)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}_0) \quad (1)$$

where $\mathbf{x} = \mathbf{x}_a - \mathbf{x}_b$ is the analysis increment, $\mathbf{B}$ is the background error covariance matrix, $\mathbf{y}_0 = \mathbf{y} - \mathbf{H}\mathbf{x}_b$ is the innovation vector, $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{H}$ is the linearized observation operator and $\mathbf{x}_b$ is the background state field. It is well known that the analysis increment $\mathbf{x}$ can be solved through the minimization of equation (1) (Daley and Barker 2001), i.e.

$$\mathbf{x} = \mathbf{x}_a - \mathbf{x}_b = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}[\mathbf{y} - \mathbf{H}(\mathbf{x}_b)] \quad (2)$$

Let $\mathbf{A}$ be the analysis error covariance of $\mathbf{x}_a$ with respect to the truth. Then in the incremental 3D-Var, $\mathbf{A}$ can be described as:

$$\mathbf{A} = \mathbf{B} - \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B} \quad (3)$$

where $\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ in equations (2) and (3) is commonly called the Kalman gain matrix. Fisher and Courtier (1995) proposed three approximate methods to estimate analysis error variance in 3D/4D-VAR framework. The most promising

method among these three is the Lanczos method which uses the linkage between the Lanczos algorithm and conjugate gradient minimization which is widely used in 3D/4D-Var system. The Lanczos method uses the relationship between the Hessian matrix and the analysis error covariance. Thus a limited number of dominant eigenvectors can be estimated and used to approximate the second right term in equation (3). This method was implemented in the ECMWF 4D-Var system to estimate the analysis error covariance (Fisher 2007, personal communications).

Since different preconditioning strategies are used in ECMWF and NCEP DA systems, the equations and derivations of the analysis error covariance are different. In the GSI, let

$$\mathbf{z} = \mathbf{B}^{-1}\mathbf{x} \tag{4}$$

The gradient of cost function with respect to $\mathbf{x}$ is

$$\mathbf{g} = \nabla_{\mathbf{x}} J(\mathbf{x}) = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{x} - \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_0 = \mathbf{M}\mathbf{x} - \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_0 \tag{5}$$

and the Hessian matrix $\mathbf{M}$ can be written as

$$\mathbf{M} = \frac{\partial^2 J(\mathbf{x})}{\partial \mathbf{x}^2} = \mathbf{B}^{-1} + \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H} = \mathbf{A}^{-1} \tag{6}$$

Eq. (5) is equivalent to

$$\mathbf{M}\mathbf{x} = \mathbf{g} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_0 \tag{7}$$

The gradient with respect to $\mathbf{z}$ is

$$\mathbf{h} = \nabla_{\mathbf{z}} J(\mathbf{x}) = \mathbf{B}\nabla_{\mathbf{x}} J(\mathbf{x}) = \mathbf{x} + \mathbf{B}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}_0) \tag{8}$$

The preconditioned conjugate gradient method used to minimize the cost function defined in eq. (1) in GSI can be expressed as (Derber and Rosati 1989)

$$\begin{aligned}
&\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \\
&\mathbf{g}_{k+1} = \nabla_{\mathbf{x}} J(\mathbf{x}_{k+1}) \\
&\mathbf{h}_k = \mathbf{B}\mathbf{g}_{k+1} \\
&\mathbf{d}_{k+1} = -\mathbf{h}_{k+1} + \beta_k \mathbf{d}_k \\
&\mathbf{e}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{e}_k
\end{aligned} \tag{9}$$

where α_k is the step size, $\mathbf{d}_k$ and $\mathbf{e}_k$ are the search directions in $\mathbf{x}$ and $\mathbf{z}$, and the conjugate factor is

$$\beta_k = \frac{(\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{h}_{k+1}}{(\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{d}_k} \tag{10}$$

$\mathbf{g}_k$ and $\mathbf{h}_k$ can be normalized such that

$$\mathbf{g}_k{}^n = c_k \mathbf{g}_k \quad \text{and} \quad \mathbf{h}_k{}^n = c_k \mathbf{h}_k \tag{11}$$

where $c_k = \dfrac{1}{\sqrt{\mathbf{g}_k{}^T \mathbf{h}_k}}$

It can be shown that the two sets form a bi-orthogonal system such that

$$\left\langle \mathbf{g}_i{}^n,\ \mathbf{h}_j{}^n \right\rangle = \delta_{ij} \tag{12}$$

After numerous mathematical derivations and by using the relations in eq.s (5)-(12), it can be shown that the final analysis error covariance matrix can be expressed as

$$\mathbf{A} = \mathbf{B} - \sum_{k=1}^{K} \rho(k)(1 - \lambda_k^{-1})(\mathbf{Q}_k{}^n \mathbf{v}_k)(\mathbf{Q}_k{}^n \mathbf{v}_k)^T \tag{13}$$

For more details about the derivations, see Pondeca *et al.* (2010). Let

$$\mathbf{C} = \sum_{k=1}^{K} \rho(k)(1-\lambda_k^{-1})(\mathbf{Q}_k{}^n \mathbf{v}_k)(\mathbf{Q}_k{}^n \mathbf{v}_k)^T \quad (14)$$

$\mathbf{C}$ is the analysis error covariance reduction from the background due to the observations, K is the number of gradients in $\mathbf{x}$ and $\mathbf{z}$, and it is the actual number of inner iterations in the GSI, λ_k and $\mathbf{v}_k$ are the dominant eigenvalues and eigenvectors of a KxK tri-diagonal matrix $\mathbf{T}_k$ consisting of different coefficients as

$$\mathbf{T}_k = \begin{bmatrix} \delta_1 & \theta_1 & 0 & 0 \\ \eta_1 & \delta_2 & \theta_2 & 0 \\ 0 & \eta_2 & \delta_3 & \bullet \\ 0 & 0 & \bullet & \bullet \end{bmatrix} \quad (15)$$

where $\delta_{k+1} = \frac{1}{\alpha_{k+1}} + \frac{\beta_k}{\alpha_k}$, $\eta_{k+1} = -\frac{c_k}{\alpha_k c_{k+1}}$, $\theta_k = -\frac{\beta_{k-1} c_k}{\alpha_{k-1} c_{k-1}}$ and

$$\mathbf{Q}_k{}^n = [\mathbf{h}_1{}^n, \mathbf{h}_2{}^n, \mathbf{h}_3{}^n \ \ldots\ldots \mathbf{h}_k{}^n] \quad (16)$$

Finally $\rho(k)$ is the calibration factor we introduce in this paper to compensate for the loss of trailing eigenvectors. Since we can only compute the dominant eigenvalues and their corresponding eigenvectors of the matrix in eq.s (13) and (14), many less dominant eigenvectors are ignored. The error reduction of $\mathbf{C}$ and analysis error covariance $\mathbf{A}$ would be underestimated and overestimated respectively without introducing the calibration factor $\rho(k)$. In the NCEP RTMA, only observations near the surface are assimilated and surface variables are estimated (Pondeca and Manikin 2009). In our case, the system is applied to estimate the analysis errors for any model variables at any levels. It can be shown that λ_k and $\mathbf{e}_k = \mathbf{Q}_k{}^n \mathbf{v}_k$ are the dominant eigenvalues and eigenvectors of the transformed Hessian matrix.

3. Eigenvalues and eigenvectors

The diagonal part of $\mathbf{C}$ in equation (14) represents the reduction of error variance due to observations in DA. The following experiments are carried out with GSI at T62L64 resolution. The number of dominant eigenvectors K computed depends on the number of inner loops in the GSI minimization. The values of K that we have tested are $K = 30,\ 60,\ 100,\ 116$. The observations and background fields are for the 00Z cycle on April 10, 2007. The observations that we chose were based on the data used in the NCEP operational GSI. There were a total of $ndat = 60$ data sets which covered conventional, aircraft, GPS observations as well as radiances from different satellites. All the operational data were used in our experiments. To study the sensitivity of error reductions to the observations, we also experimented with only conventional observations. In this case, $ndat = 6$. This includes surface pressure, temperature, specific humidity, winds, sea-surface temperature, and precipitable water from Rawinsonde. The conventional data also contain the satellite derived winds such as those below 850mb from satellites JMA IR and EUMETSAT.

Fig. 1 shows the eigenvalue distribution as a function of the eigenvalue number for different numbers of observation data sets. Shown in the left panels

are the eigenvalue distributions with 6 observation data sets for $K = 30,\ 60,\ 100,\ 116$ respectively. On the right hand panels, the same eigenvalue distributions are displayed for the 60 observation data sets. Since all the eigenvalues λ_k are larger than one and $diag[(\mathbf{Q}_k{}^n\mathbf{v}_k)(\mathbf{Q}_k{}^n\mathbf{v}_k)^T] > 0.0$, it follows that the diagonal elements of $\mathbf{C}$ are always positive, which means

$$diag(\mathbf{A}) < diag(\mathbf{B})$$

This indicates that the analysis error variance is always less than the background error variance due to the observations assimilated. From equation (14), it is easy to see that the larger the eigenvalue, the greater the observation impact (and more reduction of error variance). Thus, more dominant eigenvectors with larger eigenvalues have greater impacts than the trailing eigenvectors with smaller eigenvalues. As the trailing eigenvalue is converging to one, the contribution to error reduction from the eigenvector is approaching to zero and the impact is negligible. The smallest eigenvalues for $ndat = 6$ shown on the left panel of Fig. 1 for $K = 30,\ 60,\ 100,\ 116$ are displayed in Fig.2a, while the smallest eigenvalues for $ndat = 60$ are displayed in Fig.2b. The minimum eigenvalues in both cases should approach one (shown in dotted lines) as K increases. It is clear that the convergence of the eigenvalue with increasing number of inner loops in the GSI is much slower in the case with more observation data than in the case with smaller amount of observation data. Thus, one conclusion from this result is that the more observations we have, the larger the number of eigenvectors should be included in order to minimize the loss of information from those observations. When only 6 observation data sets are used as in Figs. 1a-1d, the eigenvalues decrease faster than when there are more observations as in Figs. 1e-1h. In fact, when all the observations are used with $ndat = 60$, the eigenvalues decrease at a uniform pace with increasing number of inner loops. The maximum eigenvalue is very similar in all cases with different values of K,

As an example of eigenvector structure, we plotted in Figs. 3 and 4 the top five normalized eigenvectors ($\mathbf{Q}_k{}^n\mathbf{v}_k$) for temperature t and zonal wind component u for $K = 30,\ 60,\ 100,\ 116$ at 500hPa. The number of data sets is $ndat = 6$. From left to right, fig. 3 shows the top eigenvectors of t with different values of K, while the top 5 eigenvectors of t with the same value of K are shown from the top to the bottom respectively. On the top panel, the 1st eigenvector for t has a similar dipole structure over the North America (NA) for $K = 30,\ 60$, while for $K = 100,\ 116$, their structures are also similar, but different from when $K = 30,\ 60$. This is also true for the other dominant eigenvectors $2,\ 3,\ 4,\ 5$ shown in the following rows. When $K = 30$, eigenvectors $1,\ 2,\ 3$ have high values over the NA, but in slightly shifted positions, while eigenvectors $4,\ 5$ have larger values over the European region. All of these reflect the fact that there are relatively more conventional data coverage in NA and European regions. When $K = 60$, the top 5 eigenvectors in column 2 show similar structures as when $K = 30$ in column 1. When the number of inner loops K is increased to 100 and 116 as shown in columns 4 and 5, the larger amplitudes of the top 5 eigenvectors are mostly over the NA area. This is consistent with the conventional observations at this level which will be shown later in the paper.

Fig. 4 shows the same as fig. 3, but for u. Overall, the top 5 eigenvectors of u also demonstrate the larger presence of conventional observations over the

NA. Like the eigenvectors of t, eigenvectors 4 and 5 shows the largest amplitudes over the European region when $K = 30,\ 60$.

Next, we look at the overall observation impacts due to observations in the GSI. To see the sensitivity of error reduction to the number of observations, we have computed the error reduction using two different numbers of observation data sets: $ndat = 6$ and $ndat = 60$. From equation (13), the analysis error variances depend on the background error variances which are static and pre-defined in most 3D/4D-Var DA systems. In the NCEP GSI, they are pre-computed using the NMC method (Parrish and Derber 1992). The diagonal components of $\mathbf{C}$ are the direct impacts from the observations assimilated by the DA. Therefore, unless we have sensible situation-dependent background error variances, it is more meaningful to look at the diagonal components of $\mathbf{C}$ (instead of $\mathbf{A}$) in order to assess the direct impacts from the observations. In the following, we will show the square root of error variance reduction, i.e. $\sqrt{diag(\mathbf{C})}$ (referred to as *observation impact signal* or OIS in the following) for different variables at different levels. First, let's look at the OIS without calibration, i.e. $\rho(k) = 1.0$ in equation (14). From the left to the right, Fig. 5 shows the OIS for u_{500}, t_{500}, q_{1000} (the subscripts indicate the level in hPa) with different number of inner loops $K = 30,\ 60,\ 100,\ 116$ respectively. The results show that the OIS for all variables increase as the number of eigenvectors increases. Rows 1 to 3 shows the OIS for u_{500}, t_{500}, q_{1000} with $ndat = 6$, while rows 4 to 6 show the OIS for same variables but for $ndat = 60$. The top two panels show that as K increases, the impacts of observations for u_{500}, t_{500} increase, particularly over the conventional data-rich regions such as NA. For cases with smaller numbers of eigenvectors ($K = 30,\ 60$), the observation impacts over Europe, Asia and areas near Australia are less clear. The impact signals are much more pronounced for $K = 100,\ 116$. Panels in row 3 show the OIS for relative humidity near the surface q_{1000}. Again, as more eigenvectors are included, stronger impact signals are observed in NA, Europe, Asia and areas near Australia, where there are more observations. The observation locations will be shown later in Figs. 8-11.

Panels in rows 4 to 6 show the OIS values for the same variables as the top 3 rows, but with $ndat = 60$. The above conclusions for $ndat = 6$ still hold when many more observations are included in the analysis. In this case, there are a lot more observations in the Southern Hemisphere (SH) that impact the OIS for three variables. In the tropic oceans, there are many moisture observations in the lower atmosphere, thus also resulting in very strong OIS.

4. Calibration of observation impacts

Results shown in Fig. 5 are the values of OIS for some variables at specific levels without calibration, i.e. $\rho(k) = 1.0$. As demonstrated in the above section, the OIS is underestimated due to the missing trailing eigenvectors. However, to what extent the OIS is underestimated is not clear from Fig. 5. To gain a quantitative assessment of underestimation, we have computed the mean OIS values of u, t, q for $K = 30,\ 60,\ 100,\ 116$ at three typical model levels, namely $L = 1$ (lowest level, about 1000mb), $L = 25$ (about 500mb) and $L = 64$ (highest model level, about 0.27mb). For each of variables at each level, we calculate the ratios of the mean OIS values with $K = 30,\ 60,\ 100,\ 116$ to the mean OIS value with $K = 30$. These ratios are shown as a function of the number of loops in Fig.

6. The left panels of Figs. 6a-6c show these ratios for u, t, q respectively with $ndat = 6$, while Figs. 6d-6f show the same, but with $ndat = 60$.

With the smaller number of observations ($ndat = 6$), Fig. 6a and 6b show that the mean OIS values of u and t at all three levels for $K = 60,\ 100,\ 116$ are about $1.5,\ 2.0,\ 2.2$ times larger than the mean OIS values for $K = 30$ respectively. This is equivalent to about 1.5 times increase of mean OIS when K is doubled. These ratios for q are even larger, especially at highest model level ($L = 64$) where values reach over 4.5 for $K = 100,\ 116$ as shown in Fig. 6c. This means that the OIS for q is even more severely underestimated compared with the other two variables when only a limited number of eigenvectors are included.

When more observations are assimilated with $ndat = 60$, as shown in Figs. 6d, 6e and 6f, the conclusions are similar, except that the ratio for u at the top model level are much larger than for the other two levels (Fig. 6d). In addition, the ratio for q at the top model level is also much larger than at other two levels (Fig. 6f), but not as large as in the case of $ndat = 6$ (Fig. 6c). All these results clearly show that the mean OIS values for all variables at these three typical model levels increase as the number of loops K increases in these two cases with two very different observation data sets. The maximum value of K that we have used is 116, which is too small in both cases. The results show no sign of converging at all even with smaller amount of conventional observations ($ndat = 6$). Since the minimization algorithm in GSI has its own built-in convergence criterion, it will stop once this criterion is satisfied. Therefore we could not test much larger K as we wished. Of course it is computationally prohibitive if K is too large in an operational system. To account for the impacts from the missing eigenvectors, calibration is inevitable.

The calibration schemes we have tested are termed $C1,\ C2,\ C3,\ C4$, and are defined as follows:

$$C1: \quad \rho(\mathrm{k}) = 1.0 + \frac{\lambda_k}{\sum_{k=1}^{K} \lambda_k}$$

$$C2: \quad \rho(\mathrm{k}) = (1.0 + \frac{\lambda_k}{\sum_{k=1}^{K} \lambda_k})^2$$

$$C3: \quad \rho(\mathrm{k}) = 1.0 + \ln(\mathrm{k})$$

$$C4: \quad \rho(\mathrm{k}) = 1.0 + \log_{10}(\mathrm{k})$$

In schemes $C1$ and $C2$, the calibration factors are functions of the eigenvalues. The formulation is such that they decay with the number of eigenvectors. Less weights are given to the less dominant eigenvectors, reflecting the fact that the trailing eigenvectors contribute less to the OIS in equation (14). Unlike these two schemes, the calibration factors defined in schemes $C3$ and $C4$ increase as a function of k. This means that more weights are given to the more trailing eigenvectors in compensating the missing eigenvectors. Note, however, that the calibration factors $C3$ and $C4$ do not depend on the eigenvalues. To test these calibration schemes, we choose the smallest number of loops from our experiments, i.e. $K = 30$. The original OIS values without calibration for u_{500}, t_{500}, q_{1000} are displayed in the first column

of Fig. 5. Since we are unable to run an ideal case with a very large number of loops to assess the effectiveness of different calibration schemes, we assume that the "ideal case" to compare with is the one with $K=116$, which is shown in the last column of Fig. 5. From the left to the right, columns in Fig. 7 show the OIS values from schemes $C1,\ C2,\ C3,\ C4$ respectively. Similar to Fig. 5, rows 1 to 3 and rows 4 to 6 in Fig. 7 are the OIS for $ndat=6$ and $ndat=60$ respectively.

By comparing with the results without calibration (first column of Fig. 5), the OIS values from calibration $C1$ in first column of Fig. 7 show a very similar pattern and magnitude for all the variables. Clearly, the magnitudes are generally smaller than those in the "ideal case". Further more, the results from $C1$ with $ndat=6$ fail to pick up much impact over regions in Europe, Asia and Australia as in the "ideal case" (last column of Fig. 5). Scheme $C2$ (column 2 of Fig. 7) also shows similar impact patterns and magnitudes as scheme $C1$. Both of these schemes can not achieve the ideal results we have hoped.

The results from scheme $C3$ (column 3 of Fig. 7) show somewhat larger maximum values than the "ideal case" for all the variables, while the magnitudes of OIS from $C4$ (column 4 of Fig. 7) are similar to those in the "ideal case". However, like $C1,\ C2$, both schemes $C3,\ C4$ do not seem to pick up the impact patterns over regions in Europe, Asia and Australia as well as the "ideal case". Overall, scheme $C3$ is probably the best among these given the fact that the "ideal case" with $K=116$ is far from converging as seen from Fig. 6. From the results of these calibration experiments, one can conclude that none of the schemes tested can produce OIS distribution that are as good as the ones from the "ideal case" which uses more eigenvectors. There are some regions where the observations are less dense. These regions will certainly need more eigenvectors to yield reasonable OIS values. Our results indicate that it is hard to use calibration factors which only increase the magnitudes of some 30 dominant eigenvectors to recover those impact signals. However, two of the calibration schemes do enhance the OIS magnitudes in the regions with dense conventional data network coverage, which in general can be gravely underestimated due to the missing eigenvectors.

5. Correlations between observations and observation impact signals

In order to further assess the impacts from observations in the GSI, we will study the OIS distributions at certain model levels and their correlations with the observation locations. In the following figures, we will plot the horizontal locations of observations which are located between the vertical levels of $p-50mb$ and $p+50mb$, and compare with the OIS values computed from equation (14) at level p. As an example, we choose $K=100$ and conventional data set only with $ndat=6$ in the following.

Fig. 8 shows the locations of temperature observations between 500mb-50mb and 500mb+50mb, and the OIS for temperature at 500mb. It is clear that the region with most conventional data is the NA, followed by Europe, Asia and regions near Australia. The method indeed produces larger OIS values over these data dense areas. However, it is also noticeable that some sparsely distributed observations around the tropics did not generate much OIS. It is expected that more eigenvectors are needed in these areas to recover some impact from observations.

Fig. 9 shows the same as Fig. 8, but for u. Similarly the OIS values are generally larger over denser observation areas. As for t, some observations around the tropics do not produce much OIS. In Fig. 10, we display the same as Fig. 9, but for surface level (around 1000mb). Around this lower level, there are many satellite derived, scatterometer type wind observations (such as from JMA IR

and EUMETSAT) in SH and the tropics as shown in Fig. 10. Our method does produce strong OIS in the SH, but not much signal can be seen in the tropics. In addition, the method does not produce a distribution that resembles the satellite data coverage across the tropics. It is expected that it will take a lot of more eigenvectors to cover these satellite data, particularly in the tropics.

Fig. 11 is the same as Fig. 10, but for relative humidity q_{1000}. The observations over Europe are equally as dense as over NA. Our method does generate some strong OIS impacts over these regions. Similarly, in Asia and Australia, the dense observations are correlated well with the strong OIS signals. Interestingly, a few sparse observations in the SH oceans are also captured by OIS. However, there are also some sparse observations around the tropics near Africa and South America (SA) which are missed by the OIS.

In conclusion, the method tested in this paper is able to capture some of the observation impacts, especially over the regions with dense conventional observations. Most satellite observations in the southern oceans can be captured. However, to capture the satellite band structures in the tropics, the number of eigenvectors we have tested is clearly not enough. All the results shown in this section indicate that many of the sparse observations in the tropics can not be captured, while the sparse observations in the southern oceans can be reflected in our OIS. This is related to the fact that the background error variances (the diagonal parts of $\mathbf{B}$) around the tropics are lower than in the extra-tropics. The background error covariance matrix in NCEP GSI is produced by using the "NMC" method (Derber *et al.* 1991; Parrish and Derber 1992; Derber *et al.* 2003). Since the observation errors do not depend on latitude in the GSI, the smaller background errors around the tropics reduce the impact of the observations, and pull the analysis closer to the background. As a result, the analysis increment is also reduced. This can also be confirmed by looking at the analysis increments in equation (2).

Corresponding to the observation locations and the OIS in Figs. 8–11, Figs. 12a–12d show the analysis increments for t_{500}, u_{500}, u_{1000}, q_{1000}. If we compare Figs. 8 and 9 with Figs. 12a and 12b, we can clearly see that there are a reasonable number of observations in the tropics, but the increments are very small. As explained, this is due to the smaller background error values used in the GSI. Thus, the OIS values are also very small around the tropics. When Figs. 10 and 11 are compared with Figs. 12a and 12d respectively, we see that areas around the tropics which do not show much OIS are also approximately the areas where the analysis increments are small.

6. Discussion and conclusions

As described in the introduction, analysis error covariance is important not only in NWP weather forecasts, but also in building a superior EFS which relies on a limited number of effective initial perturbations. An effective set of initial perturbations should be constructed from the PDF of the initial state which is the analysis field produced by DA system. (Toth and Kalnay 1993, 1997; Molteni *et al*. 1996; Houtekamer *et al.* 1996, Buizza *et al.* 2005, Bowler 2006, Wei *et al.* 2006, 2008; McLay *et al*. 2007, 2008; Reynolds *et al.* 2008; Leutbecher and Palmer 2008; McLay and Reynolds 2009; Bishop *et al*. 2009). How to use the analysis error variance to generate initial perturbations by different techniques has been reviewed and compared by Wei *et al.* (2008). At NCEP, the ETR method, which has been implemented since 2006, uses the analysis error variance information to generate the initial perturbations.

While in ensemble Kalman filter DA systems the analysis error variance is a by-product (Bishop *et al*. 2001; Whitaker and Hamill 2002; Whitaker *et al.* 2007; Kalnay *et al.* 2007), estimating the analysis error variance from the 3D/4D-Var DA systems is not straightforward. Recently, there have been some efforts in

estimating the analysis error variance using different analysis data from different centers or sources. For example, Swanson and Roebber (2008) used the differences between NCEP and ECMWF reanalysis data. Langland *et al.* (2008) analyzed the differences between NCEP and FNMOC analyses. Errico *et al.* (2007) tried to estimate analysis error variance using an observation system simulation experiment (OSSE). Wei *et al.* (2009) used the analysis data from NCEP, ECMWF, UKMO, CMC and FNMOC, and computed the anomaly of each center's analysis by removing the long term mean using a recursive filter. This method has the advantage that the typical systematic errors that appear in the spread around the center mean (i.e. mean of analyses from different centers) over high altitude regions are completely removed. The instantaneous values of spread over average anomaly at any cycle for various variables bear a strong resemblance to the elusive analysis error variance.

This paper represents another effort at NCEP/EMC in estimating the analysis error variance, using the Lanczos method proposed by Fisher and Courtier (1995), in NCEP global 3D-Var GSI DA system. Fisher and Courtier (1995) tested in a simplified case and in the ECMWF 3D-Var system. Pondeca and Manikin (2009) used this method for estimating analysis error variances over the CONUS for 2-dimensional surface variables in the NCEP RTMA system which is based on the 2D-Var option in the GSI. In this paper, we have applied this method to the global 3D-Var GSI and studied other different aspects of this method which were not exploited in Fisher and Courtier (1995) and Pondeca and Manikin (2009). The properties of convergence and different calibration schemes we have found about this method have greatly improved our understanding of the method and its implications in practical applications in an operational environment.

When applying this method to the NCEP global GSI, our focus is on estimating the observation impact signals (OIS) which are the square root of the difference between the background and analysis error variances. This quantity is a measure of the error reduction due to the observations assimilated. The analysis error variance in most 3D/4D-Var DA systems depends on the background error covariance which is static and pre-computed using the so-called NMC method (Parrish and Derber 1992).

The OIS values for different variables at typical model levels are computed for various numbers of inner loops K in the GSI with different numbers of observation sets. As expected, the smallest eigenvalue of the transformed Hessian matrix converges to one as K increases. However, the rate of convergence depends on the number of observations assimilated. Our results show that the converging speed is faster when smaller numbers of observations are used. If more observations are used, the converging speed is slower and a larger number of eigenvectors should be included in order to minimize the loss of information from the missing eigenvectors.

The top five corresponding normalized eigenvectors are also studied. In general, the structures for the top eigenvectors when K is small show larger impacts in the regions where conventional data are dominant. If K is increased, the OIS can pop up in other areas with less observations. For the same number of K, the less dominant eigenvectors may convey the impact signals from the less dominant observation regions.

When the OIS values are computed with different number of data sets, the results show that the impact signals in the data rich regions are stronger with larger K. At the same time, more signals in the regions with less observations start to emerge as the number of inner loops increases. When the number of observations is increased, the method can clearly pick up the impact signals from the observations. As only a limited number of eigenvectors can be computed due to the computational constraint, the error reduction is severely underestimated. To estimate to what extent we are losing the information from the missing

eigenvectors, we computed the mean OIS values for u, t, q at three typical model levels (i.e. top, middle and bottom) for different values of K with two different numbers of observation data sets. It is found that the OIS values at $K = 60, 100, 116$ are about $1.5, 2.0, 2.2$ times the value of OIS at $K = 30$ respectively. This is roughly 1.5 times the increase of the mean OIS when K is doubled. These ratios are much larger for the relative humidity at the top model level.

All of our results indicate that without proper calibrations, the observation impacts computed using this method are not going to be near the reality. To overcome this difficulty, we have proposed and investigated four different calibration schemes to compensate for the missing trailing eigenvectors. Different schemes give different weights on different number of eigenvectors. Our results show that the first two schemes cannot pick up the impact signals over the regions with less conventional data in comparison with the "ideal case" which has the largest number of inner loops. It is found that scheme $C3$ performs better than other schemes and can boost the OIS values in the data rich regions to the level in the "ideal case". However, it seems that none of them can pick up the impacts in the regions with less observation data as well as the "ideal case". The benefit of calibrations lies in the fact that they do enhance the OIS magnitudes in the regions with more traditional data coverage, which would be missed without calibrations.

We also studied the correlations between the observation locations and the OIS distributions for various variables at different levels. It is found that the method generally picks up the impact signals over the regions with conventional observations, particularly over the data dense areas. It even picks up the satellite observation impacts over the southern oceans. However, with the number of inner loops we have used, the method cannot show the satellite band structure over the tropics. It would probably need a lot more eigenvectors to recover the whole satellite observation impacts. The area where the method is more likely to miss is around the tropics. This is found to be due to the fact that the background errors produced by the NMC method are generally smaller over the tropics than over the extra-tropics, and the observation errors do not change with latitude. As a result, the observation impacts over the tropics are reduced. This also leads to the smaller analysis increments over the tropics.

In conclusion, the method presented in this paper with proper calibration is capable and effective in estimating the observation impacts from the observations assimilated in the GSI, especially over those regions with more conventional data coverage. Since those gradient vectors can be generated by the operational global GSI almost at no cost, the computational expense in estimating the dominant eigenvectors is completely manageable with the current NCEP computing resources.

Another benefit of using this method is that the eigenvectors can be used in preconditioning the conjugate gradient algorithm in minimization to speed up the convergence. For example, an explicit or implicit preconditioner based on an approximation to the Hessian matrix can be chosen (Fisher 1998). In this case, the time spent on computing the dominant eigenvectors can be offset by the time saved from this preconditioning. Therefore, this method is very suitable for an operational 3D/4D-Var system to estimate the observation impacts, and it can be used as part of a routine verification package.

Acknowledgements: We thank John Derber and Jim Purser for their careful reviews of the manuscript with many helpful suggestions and comments which have improved the presentation. We are grateful to many colleagues at NCEP/EMC for their help during this work, particularly, Wan-Shu Wu, Russ Treadon, Xiujuan Su and Daryl Kleist for helpful discussions about GSI system. We thank Dr. Steve Lord for his constant support and encouragement during this work, Mary Hart for improving the presentation, and Drs. Carolyn Reynolds and Craig Bishop at NRL for helpful discussion and sharing the FNMOC analysis error estimates with us.

REFERENCES

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

Bishop, C.H., B.J. Etherton, and S. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. part I: theoretical aspects. *Mon. Wea. Rev.*, **129**, 420-436.

Bishop, C. H., T. R. Holt, J. Nachamkin, S. Chen, J. G. McLay, J. D. Doyle and W. T. Thompson, 2009: Regional ensemble forecasts using the ensemble transform technique. *Mon. Wea. Rev.*, **137**, 288-298.

Bowler, N. E., 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus,* **58A**, 538-548.

Bowler, N. E., A. Arribas, and K. Mylne, 2008: The benefits of multi-analysis and poor-man's ensemble. *Mon. Wea. Rev.*, **136**, 4113-4129.

Buizza, R., P. L. Houtekamer, Z. Toth, P. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076-1097.

Daley, R., and E. Barker, 2001: NAVDAS: Formulation and Diagnostics. *Mon. Wea. Rev.*, **129**, 869-883.

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

Derber, J. C., and A. Rosati, 1989: A global oceanic data assimilation system. *Journal of Physical Oceanography,* **19**, 1333-1347.

Derber, J. C., D. F. Parish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Weather and Forecasting*, **6**, 538-547.

Derber, J. C., R. J. Purser, W.-S. Wu, R. Tredon, M. Pondeca, D. Parish, and D. Kleist, 2003: Flow dependent Jb in global grid-point 3D-Var. ECMWF 2003 Seminar, Recent development in data assimilation for atmosphere and Ocean, 8 to 12 September 2003, 125-134.

Errico, R. M., R. Yang, M. Masutani, and J. Woollen, 2007: The estimation of analysis error characteristics using an observation system simulation experiment. *Meteorologische Zeitschrift*, **16**, 695-708.

Fisher, M., and P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF Technical Memorandum*, No. 220. August 1995.

Fisher, M., 1998: Minimization algorithm for variational data assimilation. *Proceedings of a seminar held at ECMWF on Recent developments numerical methods for atmospheric modeling,* 7-11 September 1998, 364-385.

Hamill, T. M., C. Snyder, and R. E. Morss, 2000: A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon. Wea. Rev.*, **128**, 1835-1851.

Houtekamer, P. L., L. Lefaivrem, J. Derome, and J.,Ritchie, and H. L. Mitchell, 1996: A sytem simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.

Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy, 2007: 4-D-Var

or ensemble Kalman filter?. *Tellus*, **59A**, 758–773.
Kleist, D. T., D. Parrish, J. C. Derber, and R. Treadon, 2009: Improving incremental balance in the GSI 3D VAR analysis system. *Mon. Wea. Rev.*, **137**, 1046–1060.
Langlang, R. H., R. N. Maue, and C. H. Bishop, 2008: Uncertainty in atmospheric temperature analysis. *Tellus*, **60A**, 598–603.
Leutbecher, M. and T. N. Palmer, 2008: Ensemble forecasting. *Journal of Computational Physics*, **227**, 3515–3539.
Manikin, G. S., and M. Pondeca, 2009: Challenges with the real-time mesoscale Analysis (RTMA), 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction, Omaha, NE, June 2009, American Meteorological Society.
McLay, J., C. H. Bishop, and C. A. Reynolds, 2007: The ensemble-transform scheme adapted for the generation of stochastic forecast perturbations. *Quart. J. Roy. Meteor. Soc.*, **133**, 1257–1266.
McLay, J., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093–1108.
McLay, J., and C. A. Reynolds, 2009: Two alternative implementations of the ensemble-transform (ET) analysis-perturbation scheme: The ET with extended cycling intervals, and the ET without cycling. *Quart. J. Roy. Meteor. Soc.*, **135**, 1200–1213.
Molteni, F., R. Buizza, T. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: Preliminary results on comparing and combining ensembles. *Quart. J. Roy. Met. Soc.*, **134**, 2029–2050.
Parrish, D. F. and J. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
Pondeca, M., and G. S. Manikin, 2009: Recent improvements to the real-time mesoscale analysis, 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction, Omaha, NE, June 2009, American Meteorological Society.
Pondeca, M., G. S. Manikin, G. DiMego, S. G. Benjamin, D.Parrish, J. Purser, W-S. Wu, J. Horel, Y. Ling, R.M. Aune, D. Keyser, L. Anderson, B. Colman, G. Mann and J. Vavra, 2010: The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: current status and development. (submitted to Wea. And Forecasting)
Rabier, F., H. Jarvine,E.Klinker, J.-F.Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics. *Quart. J. Roy. Met. Soc.*, **126**, 1143–1170.
Reynolds, C. A., J. Teixeira and J. G. McLay, 2008: Impact of stochastic convection on the ensemble transform. *Mon. Wea. Rev.*, **136**, 4517–4526.
Swanson, K. L.,and P. Roebber, 2008: The impact of analysis error on medium-range weather forecasts. *Mon. Wea. Rev.*, **136**, 3425–3431.
Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. York, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, **60A**, 113–130.
Tippett, M. K., J. L. Anderson, C. H. Bishop, T. Hamill, and J.S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Amer. Meteror. Soc.*, **174**, 2317–2330.
Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
Wei, M., and Z. Toth, 2003: A new measure of ensemble performance: Perturbations

versus Error Correlation Analysis (PECA). *Mon. Wea. Rev.*, **131**, 1549–1565.

Wei, M., Z. Toth, R. Wobus, Y. Zhu, and C.H. Bishop, 2005: Initial Perturbations for NCEP Ensemble Forecast System. Thorpex Symposium Proceedings for t*he First THORPEX Internal Science Symposium* 6–10 December 2004, Montreal, Canada. The Symposium Proceedings in a WMO Publication 2005, ***WMO TD No.1237,* WWRP THORPEX No. 6**, 2005. p227–230.

Wei, M., Z. Toth, R. Wobus, Y. Zhu, C. H. Bishop, and X. Wang, 2006: Ensemble Transform Kalman Filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus,* **58A**, 28–44.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the Ensemble Transform (ET) technique in the NCEP global operational forecast system. *Tellus* **60A**, 62–79.

Wei, M., Z. Toth, and Y. Zhu, 2009: Analysis differences and error variance estimates from multi-center analysis data. *The Australian Meteorological & Oceanographic Journal* (in press).

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

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2007: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**, 463–482.

Wu, W.-S., R.J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916.

Xu, L., T. Rosmond, and R. Daley, 2005: Development of NAVDAS-AR: Formulation and initial tests of the linear problem. *Tellus*, **57A**, 546–559.

Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.*, **133**, 1710–1726.

Figure Captions

Fig.1. The eigenvalue distribution, (a) *ndat=6, K=30*, (b)*ndat=6, K=60*, (c) *ndat=6, K=100*, (d) *ndat=6, K=116*, (e) *ndat=60, K=30*, (f) *ndat=60, K=60*, (g) *ndat=60, K=100*, (h) *ndat=60, K=116*.

Fig.2. The smallest eigenvalues as a function of the number of loops (a) for $ndat = 6$, (b) for $ndat = 60$.

Fig. 3. Top five normalized eigenvectors of the transformed Hessian with respect to temperature at 500hPa with ndat=6. From the left to the right on each row: top eigenvectors for *K=30, 60, 100, 116* respectively. Top to the bottom on each column: eigenvectors 1, 2, 3, 4, 5 for each value of *K*.

Fig. 4. The same as Fig. 3, but for u.

Fig.5. Observation impact signal (OIS) for u_{500}, t_{500}, q_{1000}. From the left to the right on each row: $K = 30,\ 60,\ 100,\ 116$ respectively. Rows 1 to 3 for u_{500}, t_{500}, q_{1000} with $ndat = 6$, and rows 4 to 6 for $ndat = 60$.

Fig.6. The ratio of the mean OIS values using $K = 30,\ 60,\ 100,\ 116$ to the mean OIS value using $K = 30$ as a function of the number of loops at 1000hPa, 500hPa and 0.27hPa for. (a) *u* with *ndat=6*, (b) *t* with *ndat=6*, (c) *q* with *ndat=6*, d) *u* with *ndat=60*, (e) *t* with *ndat=60*, (f) *q* with *ndat=60*.

Fig. 7. OIS with four calibration schemes for u_{500}, t_{500}, q_{1000}. From the left to the right on each column: OIS values from schemes $C1,\ C2,\ C3,\ C4$ respectively. Rows 1 to 3 on each column: for u_{500}, t_{500}, q_{1000} with $ndat = 6$. Rows 4 to 6 on each column: the same variables for $ndat = 60$.

Fig.8. Circles represent the locations of temperature observations between 500mb-50mb and 500mb+50mb, and contour lines the OIS for temperature at 500mb.

Fig. 9. The same as Fig. 8, but for u.

Fig. 10. The same as Fig. 9, but for surface level (around 1000mb).

Fig. 11. The same as Fig. 10, but for relative humidity q.

Fig. 12. The analysis increments of (a) t_{500}, (b) u_{500}, (c) u_{1000}, (d) q_{1000}.

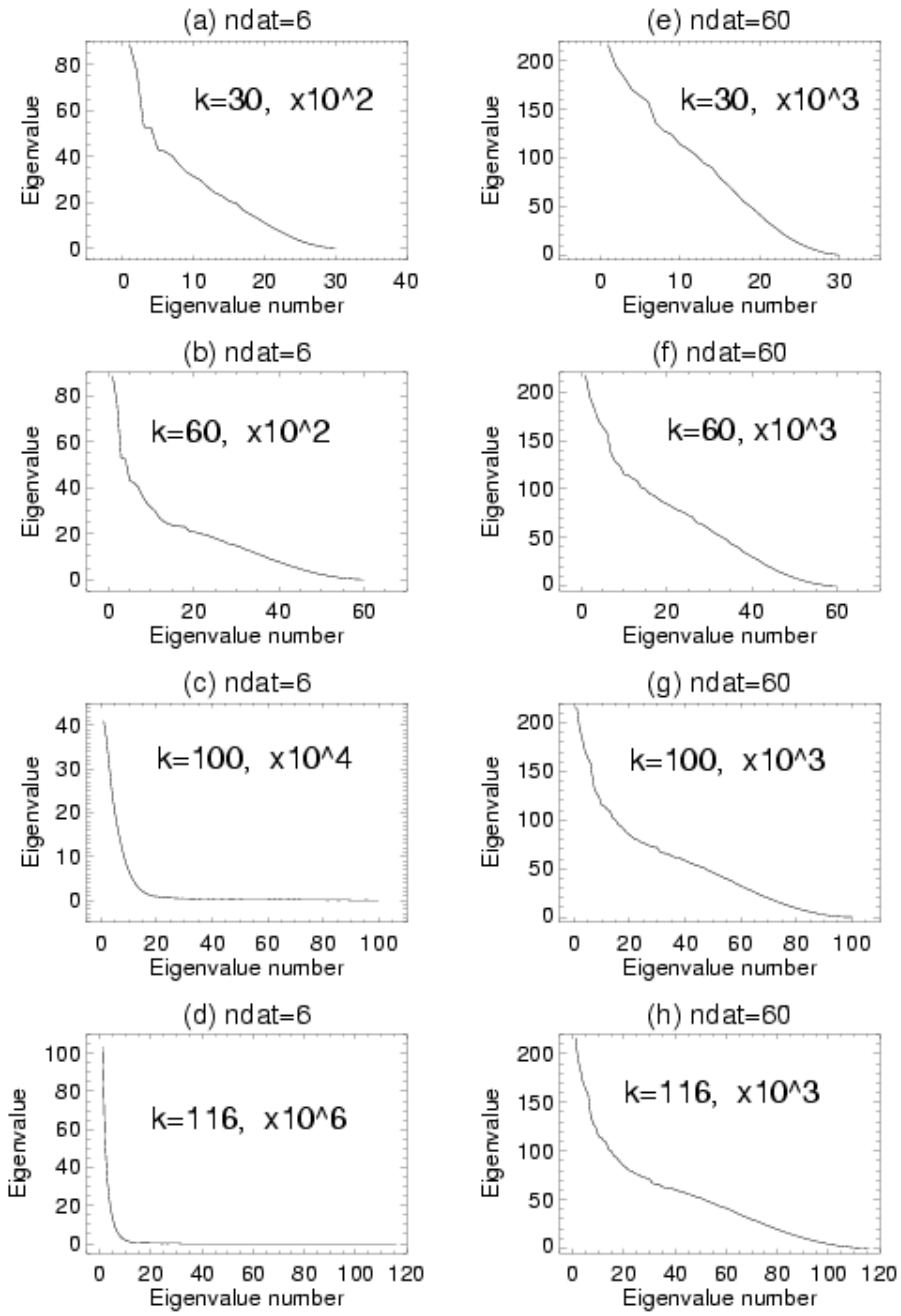

Fig.1. The eigenvalue distribution, (a) *ndat=6, K=30*, (b) *ndat=6, K=60*, (c) *ndat=6, K=100*, (d) *ndat=6, K=116*, (e) *ndat=60, K=30*, (f) *ndat=60, K=60*, (g) *ndat=60, K=100*, (h) *ndat=6, K=116*.

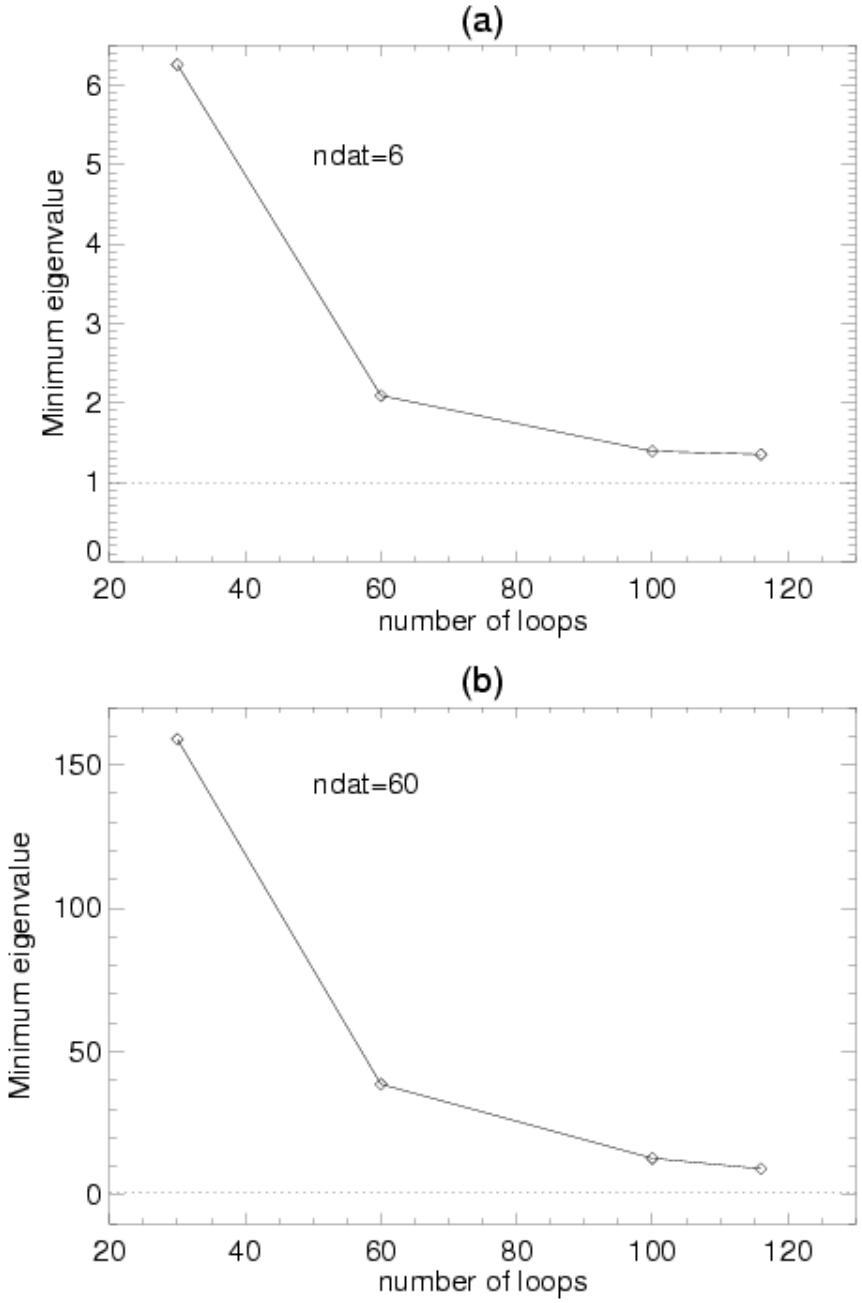

Fig.2. The smallest eigenvalues as a function of the number of loops (a) for $ndat = 6$, (b) for $ndat = 60$.

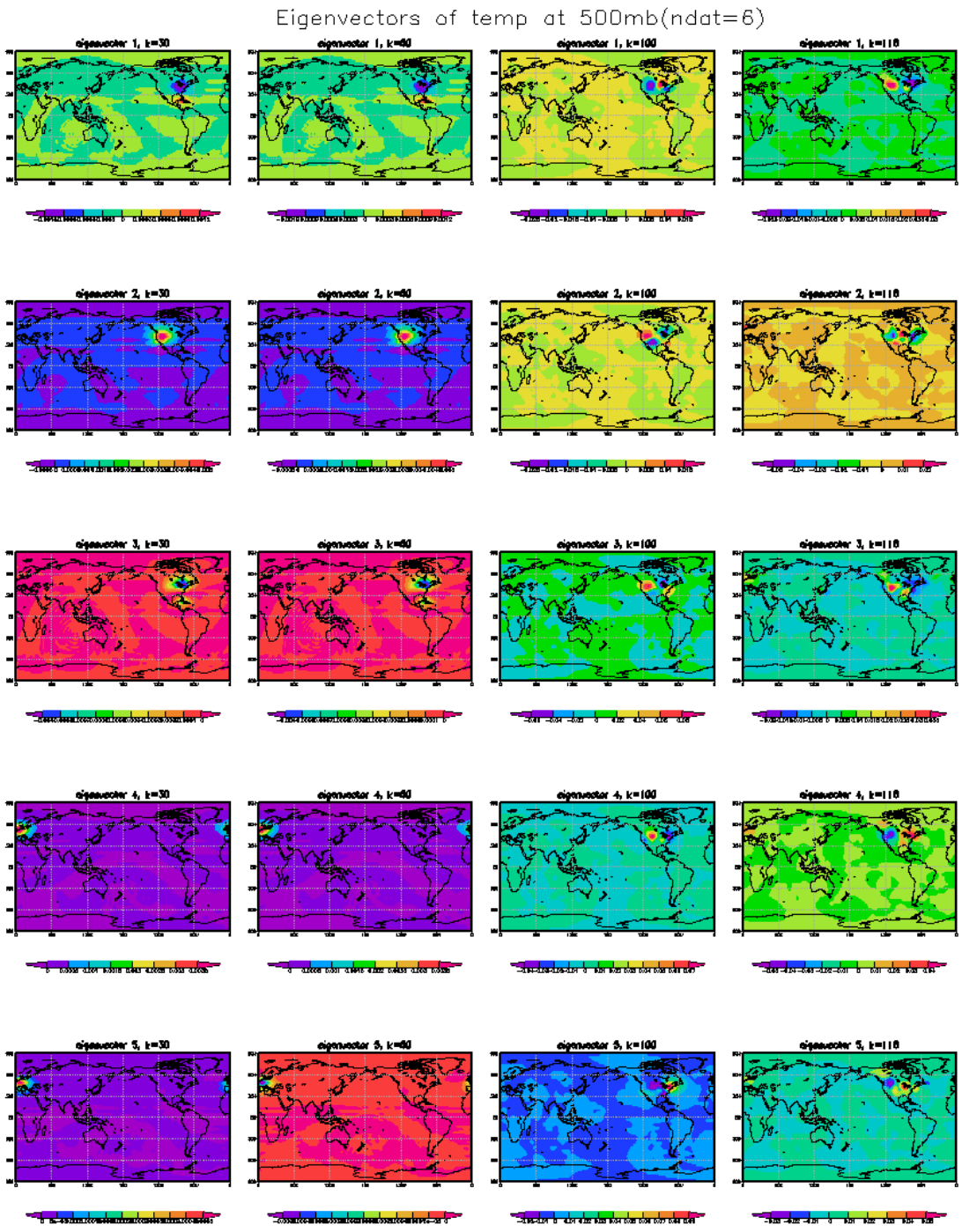

Fig.3. Top five normalized eigenvectors of the transformed Hessian with respect to temperature at 500hPa with ndat=6. From the left to the right on each row: top eigenvectors for *K=30, 60, 100, 116* respectively. Top to the bottom on each column: eigenvectors 1, 2, 3, 4, 5 for each value of *K*.

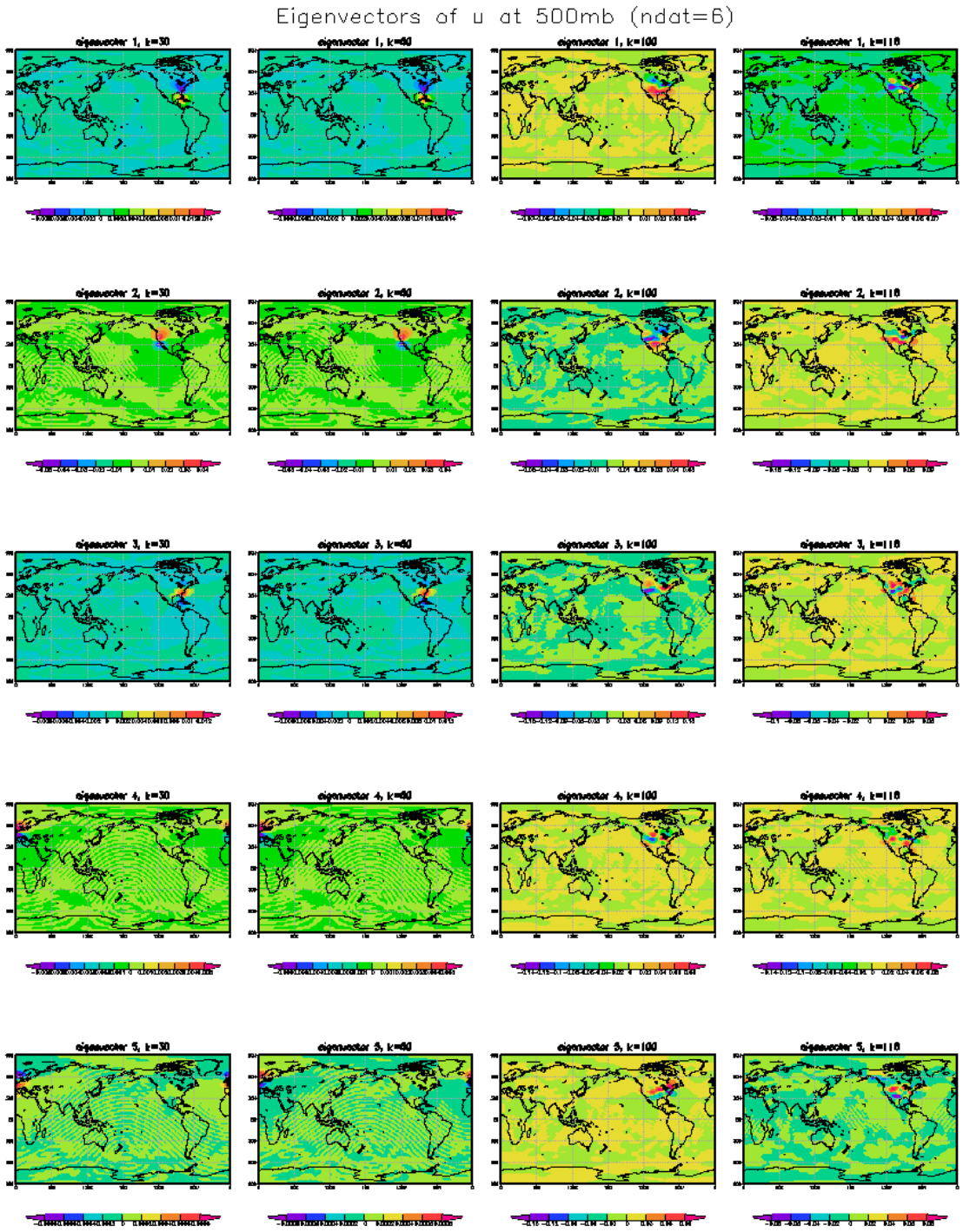

Fig. 4. The same as Fig. 3, but for u.

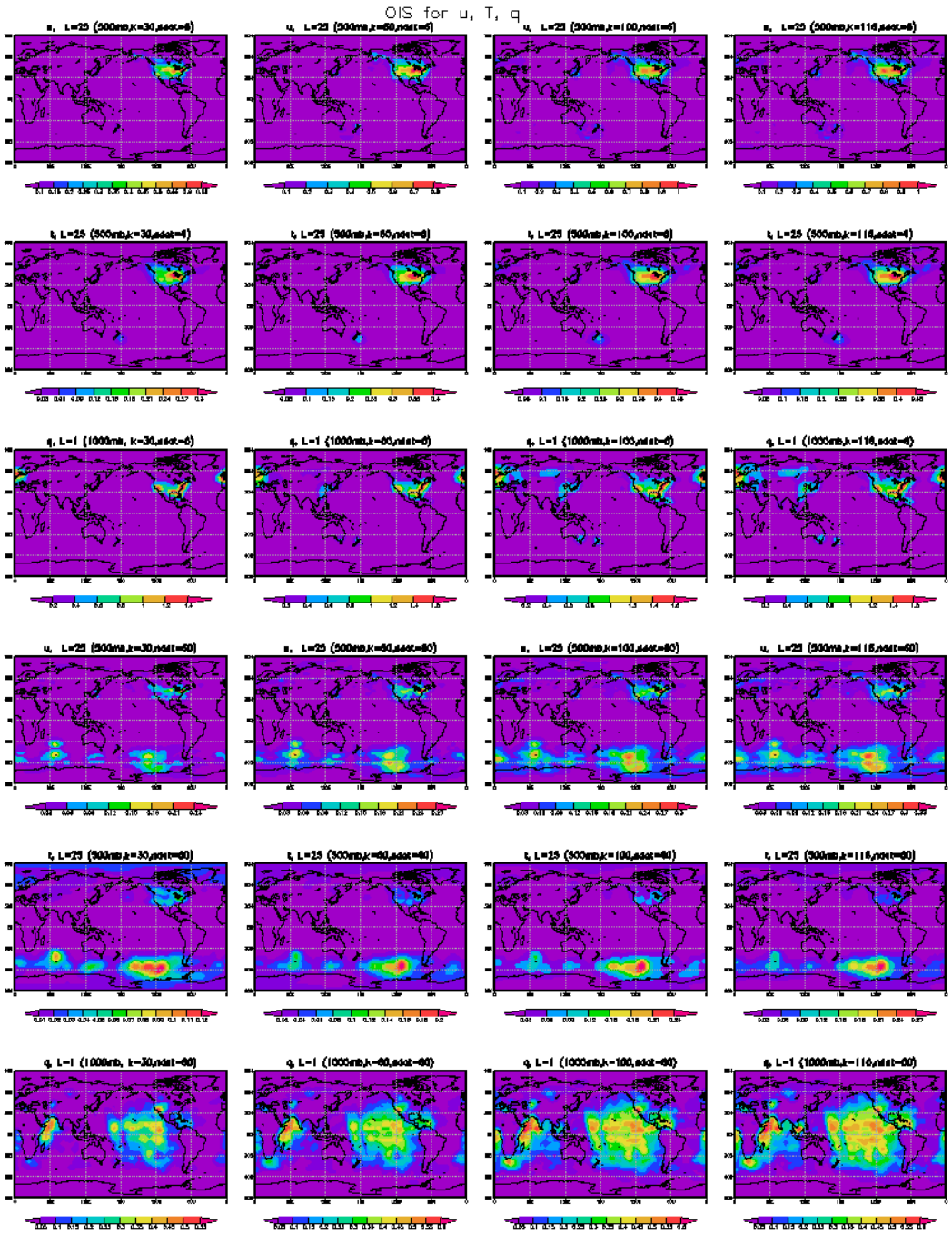

Fig.5. Observation impact signal (OIS) for u_{500}, t_{500}, q_{1000}. From the left to the right on each row: $K = 30,\ 60,\ 100,\ 116$ respectively. Rows 1 to 3 for u_{500}, t_{500}, q_{1000} with $ndat = 6$, and rows 4 to 6 for $ndat = 60$.

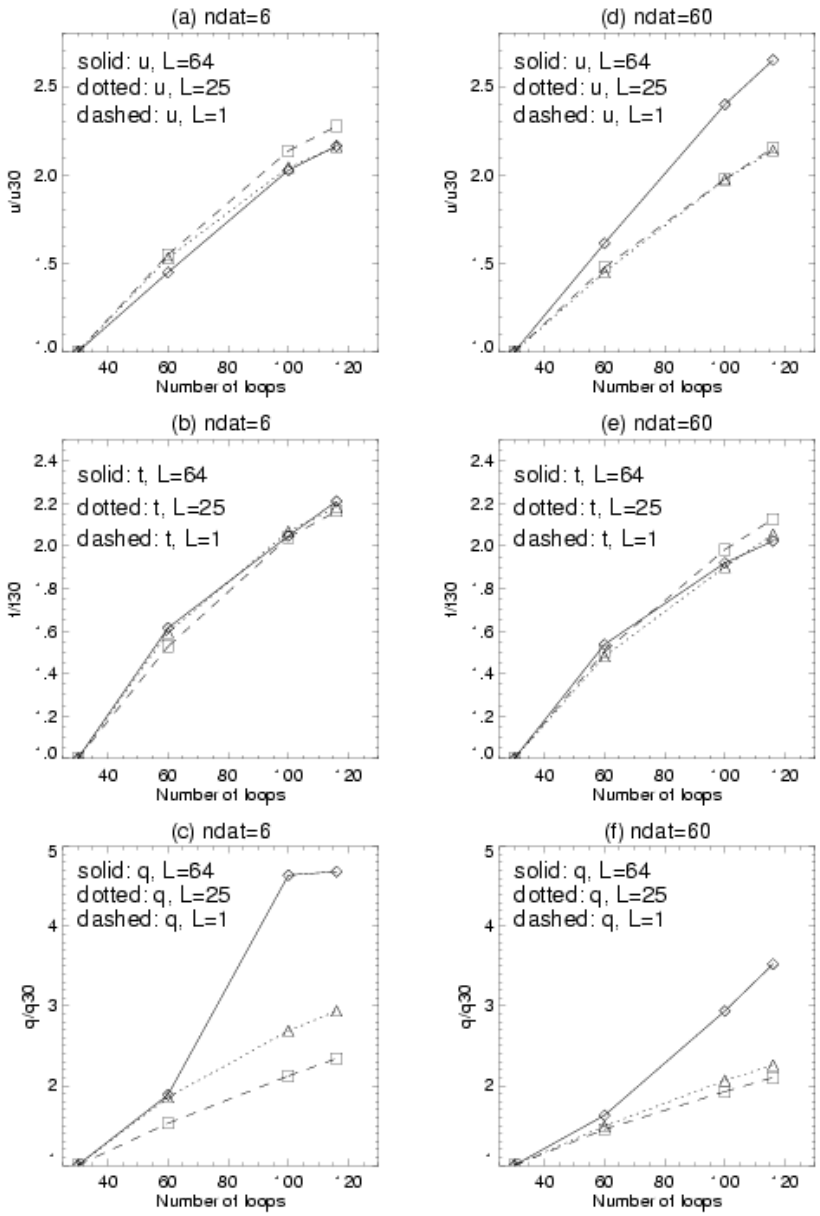

Fig. 6. The ratio of the mean OIS values using $K = 30,\ 60,\ 100,\ 116$ to the mean OIS value using $K = 30$ as a function of the number of loops at 1000hPa, 500hPa and 0.27hPa for. (a) *u* with *ndat=6*, (b) *t* with *ndat=6*, (c) *q* with *ndat=6*, d) *u* with *ndat=60*, (e) *t* with *ndat=60*, (f) *q* with *ndat=60*.

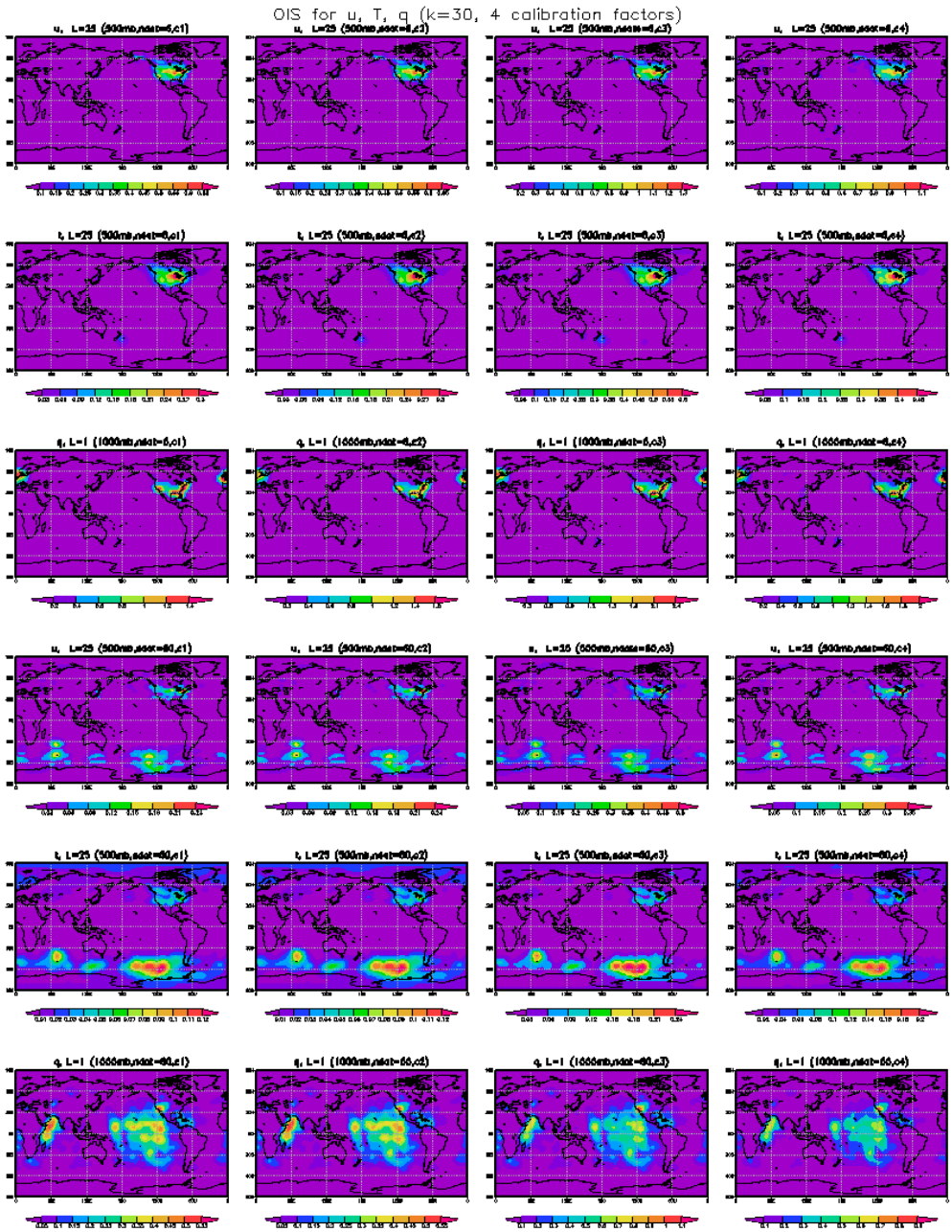

Fig.7. OIS with four calibration schemes for u_{500}, t_{500}, q_{1000}. From the left to the right on each column: OIS values from schemes $C1$, $C2$, $C3$, $C4$ respectively. Rows 1 to 3 on each column: for u_{500}, t_{500}, q_{1000} with $ndat = 6$. Rows 4 to 6 on each column: the same variables for $ndat = 60$.

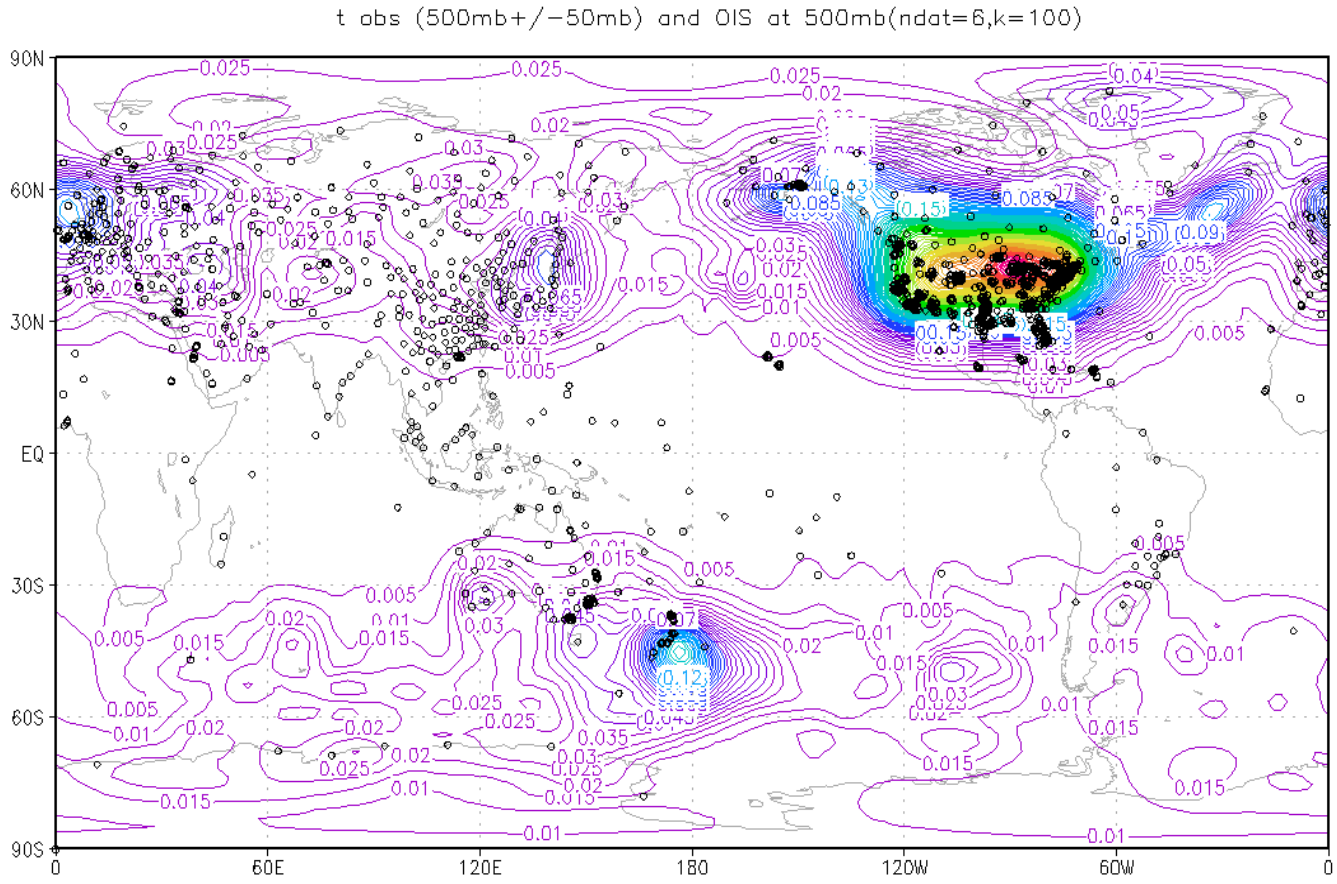

Fig. 8. Circles represent the locations of temperature observations between 500mb-50mb and 500mb+50mb, and contour lines the OIS for temperature at 500mb.

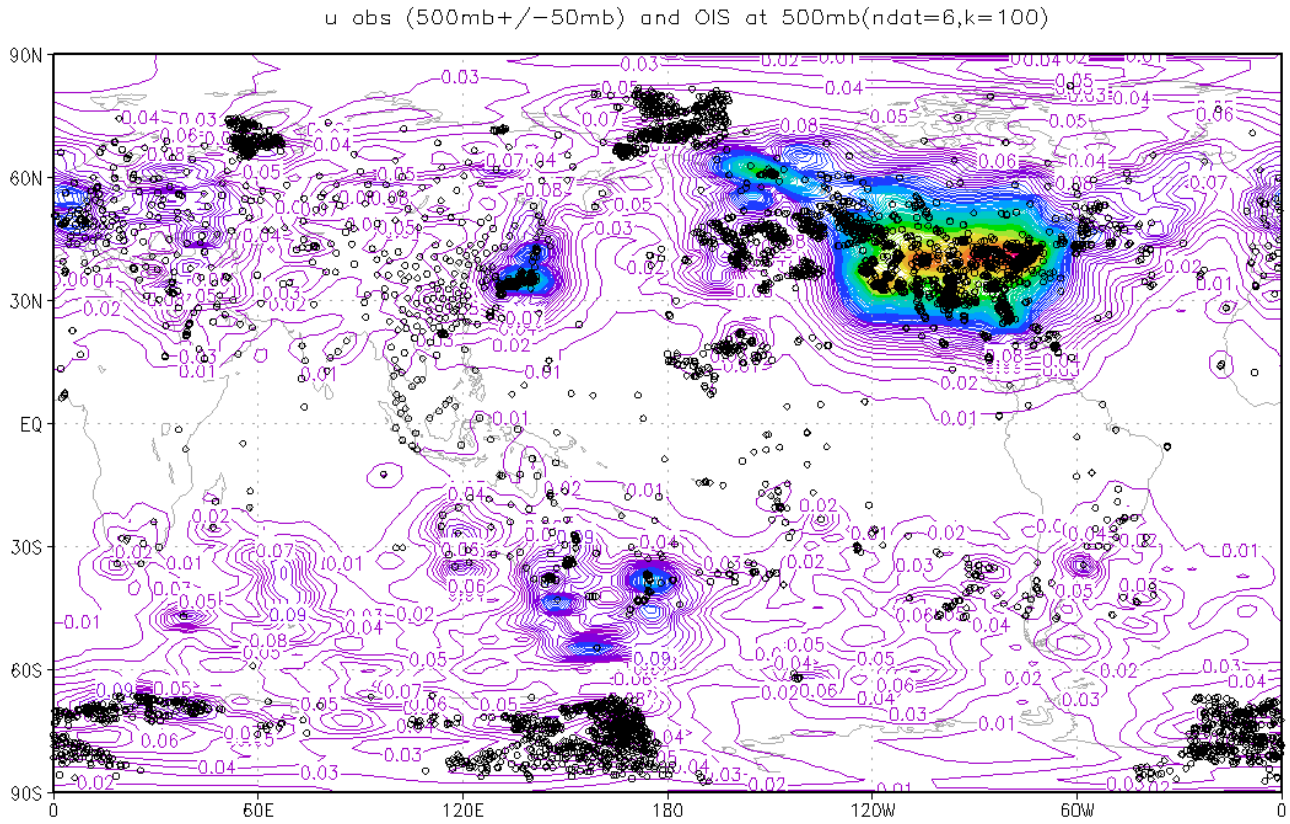

Fig. 9. The same as Fig. 8, but for u.

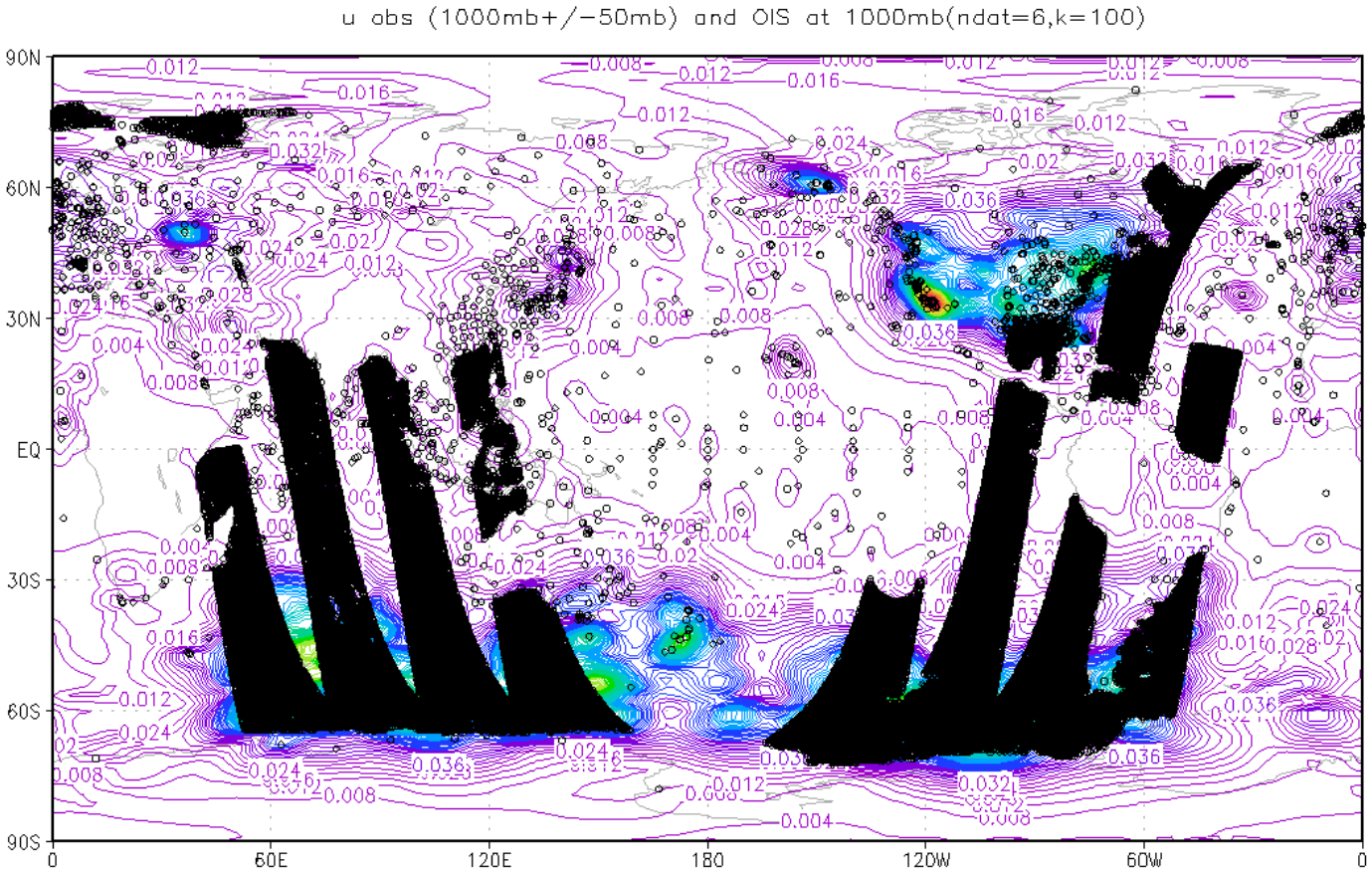

Fig. 10. The same as Fig. 9, but for surface level (around 1000mb).

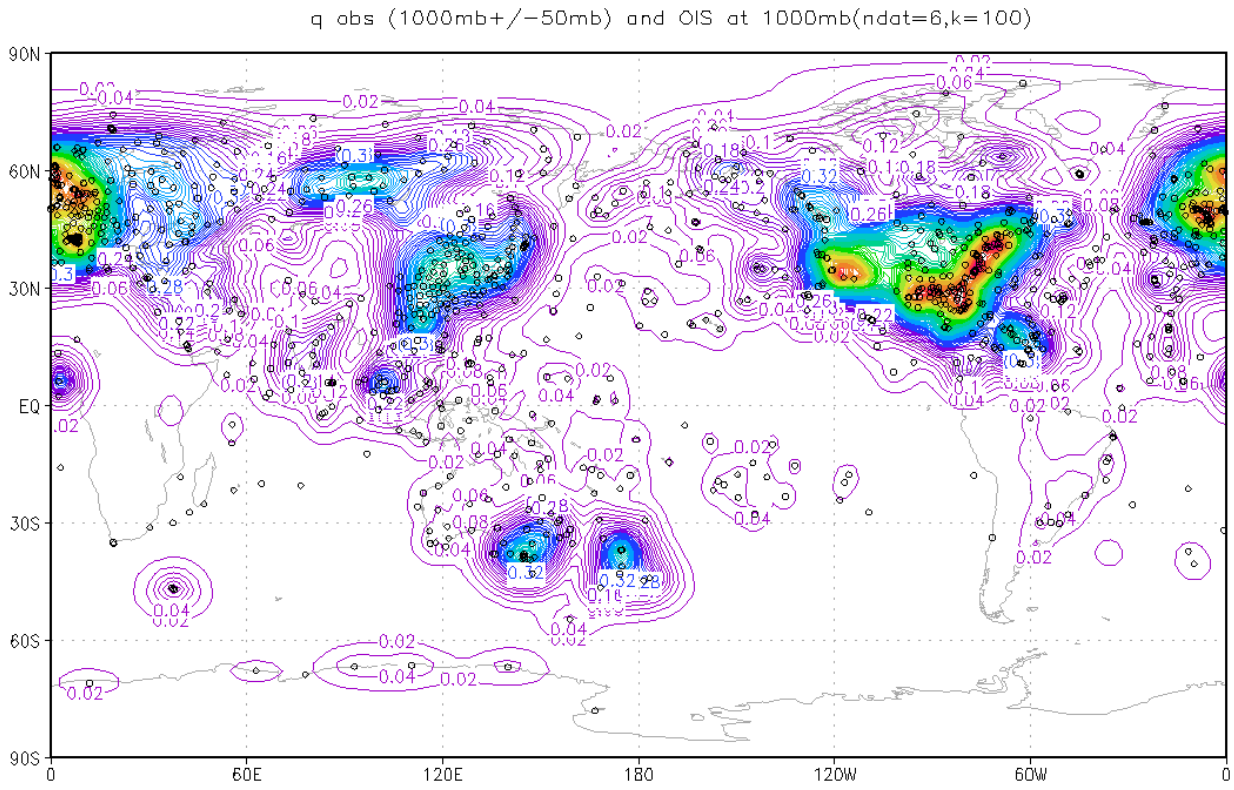

Fig. 11. The same as Fig. 10, but for relative humidity q.

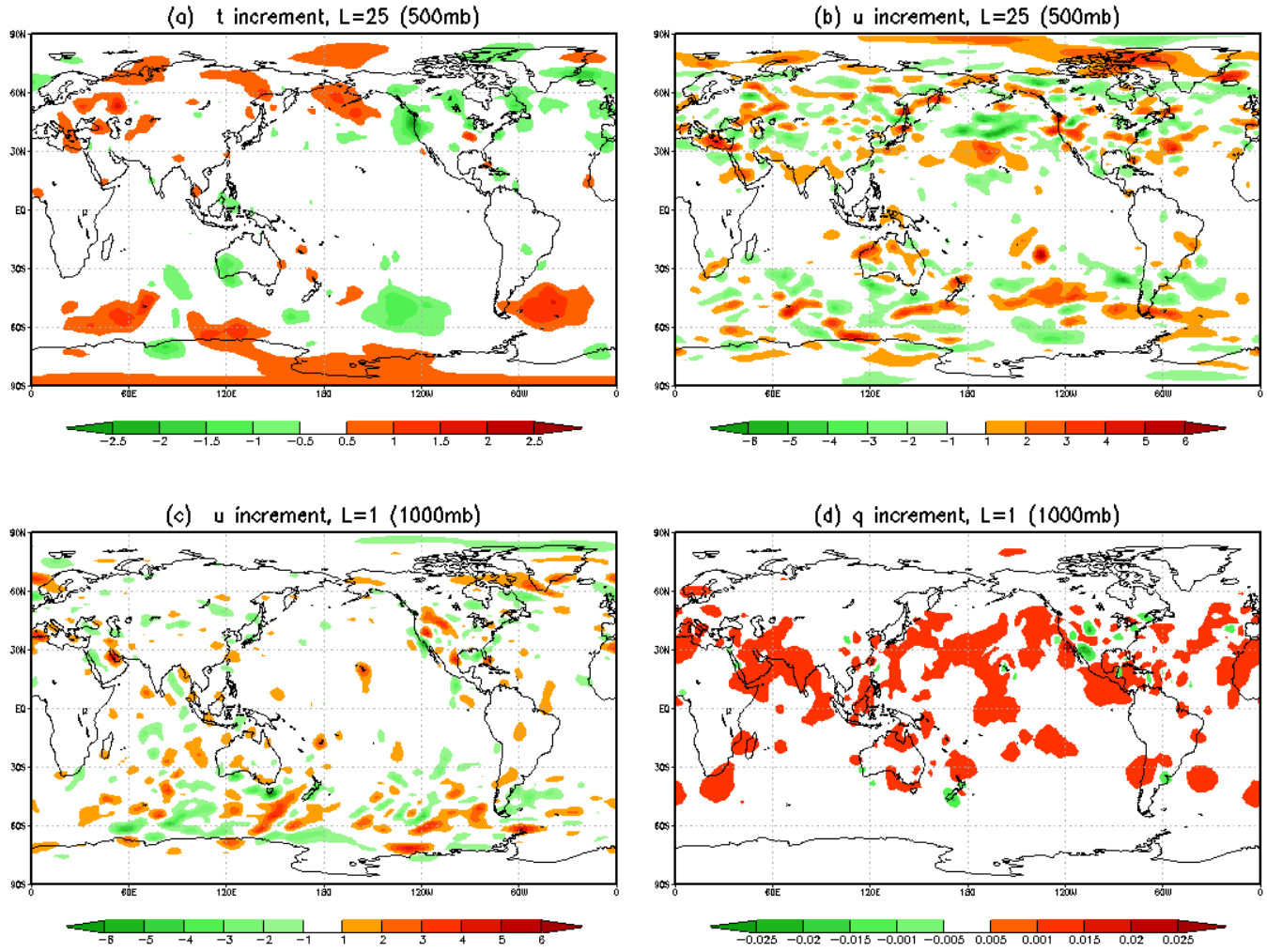

Fig. 12. The analysis increments of (a) t_{500}, (b) u_{500}, (c) u_{1000}, (d) q_{1000}.