Nonlin. Processes Geophys., 19, 541–557, 2012
www.nonlin-processes-geophys.net/19/541/2012/
doi:10.5194/npg-19-541-2012
© Author(s) 2012. CC Attribution 3.0 License.

Nonlinear Processes in Geophysics

Estimation and calibration of observation impact signals using the Lanczos method in NOAA/NCEP data assimilation system

M. Wei[1], M. S. F. V. De Pondeca[2], Z. Toth[3], and D. Parrish[2]

[1]Naval Research Laboratory, Stennis Space Center, MS 39529, USA
[2]NOAA/NWS/NCEP/EMC, Camp Springs, MD 20746, USA
[3]NOAA/OAR/ESRL/GSD, Forecast Applications Branch, Boulder, Colorado, USA

Correspondence to: M. Wei (mozheng.wei@nrlssc.navy.mil)

Received: 4 May 2012 – Revised: 18 July 2012 – Accepted: 27 August 2012 – Published: 25 September 2012

Abstract. Despite the tremendous progress that has been made in data assimilation (DA) methodology, observing systems that reduce observation errors, and model improvements that reduce background errors, the analyses produced by the best available DA systems are still different from the truth. Analysis error and error covariance are important since they describe the accuracy of the analyses, and are directly 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 straightforward, but it is challenging in variational DA systems, which have been in operation at most NWP (Numerical Weather Prediction) centers. In this study, we use the Lanczos method in the NCEP (the National Centers for Environmental Prediction) Gridpoint Statistical Interpolation (GSI) DA system to look into other important aspects and properties of this method that 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 convergence 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 represented 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 a small number of eigenvectors cannot pick up the observation impacts over the regions with fewer observations as well as a benchmark 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 large number of 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 4-D-Var (Derber, 1987; Rabier et al., 2000) and ensemble Kalman filters (EnKFs) (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 that reduces the observation errors and in model improvement that 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 system since they describe the accuracies of the analysis fields generated by the DA system. 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 years, 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 to estimate analysis error variance and covariance outside EnKF. 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 1 January to 30 June 2007. The authors found that the differences and root mean of the squared daily differences in 500 hPa 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[1] 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. (2010) 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 obtain in 3-D/4-D-Var systems, which have been in operation at most

[1] THORPEX: The Observing System Research and Predictability Experiment.

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 (3-D-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 3-D/4-D-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 (Ensemble Transform) method.

For the 3-D/4-D-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 3-D-Var on a cyclic one-dimensional domain with 256 equally-spaced grid points. Some testing was also done in ECMWF 3-D-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 and Pondeca et al., 2011). The RTMA runs the GSI in 2-D-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 m, in addition to the 10-m wind, etc. Another advantage of the RTMA, in comparison to the global 3-D-Var GSI presented in this paper, is that the analysis variables in the regional GSI are directly temperature and wind components, and therefore their analysis errors can be estimated directly, whereas the global GSI analysis variables are the stream function, unbalanced velocity potential, etc., which make the estimates of variances of wind components complicated. Furthermore, the RTMA focuses 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 3-D-Var GSI DA system to study some of its aspects and properties that 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 very inaccurate. 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 Sect. 3. Section 4 presents four different calibration schemes and their results, while the correlations between the observations and the OIS are exploited in Sect. 5. Finally, discussion and conclusions are given in Sect. 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

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

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

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

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

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

where $\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ in Eqs. (2) and (3) is commonly called the Kalman gain matrix. Fisher and Courtier (1995) proposed three approximate methods to estimate analysis error variance in 3-D/4-D-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 that is widely used in 3-D/4-D-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 Eq. (3). This method was implemented in the ECMWF 4-D-Var system to estimate the analysis error covariance (M. Fisher, personal communication, 2007).

The basic formulations of Lanczos algorithm is described in Appendix A. The final analysis error covariance matrix can be expressed as in Eq. (A10).

Let

$$\boldsymbol{C} = \mathbf{B} - \boldsymbol{A} = \sum_{k=1}^{m}(1 - \lambda_k^{-1})(\boldsymbol{Q}_k^n\boldsymbol{v}_k)(\boldsymbol{Q}_k^n\boldsymbol{v}_k)^T, \quad (4)$$

where λ_k and $\boldsymbol{e}_k = \boldsymbol{Q}_k^n\boldsymbol{v}_k$ are the dominant eigenvalues and eigenvectors, respectively, of the transformed Hessian matrix of the cost function in Eq. (1).

It is clear that $\boldsymbol{C}$ is the analysis error covariance reduction from the background due to the observations. Since it is impossible to compute all the eigenvectors of the transformed Hessian matrix in a real application such as GSI due to computing resource limit, we can only compute a very limited number (K) of dominant eigenvalues and their corresponding eigenvectors. As a result, many less dominant eigenvectors are ignored. The error reduction of $\boldsymbol{C}$ and analysis error covariance $\boldsymbol{A}$ would be underestimated and overestimated, respectively.

In order to compensate for the loss of trailing eigenvectors, we introduce a calibration factor $\rho(k)$ in this paper so that

$$\boldsymbol{C} \approx \sum_{k=1}^{K}\rho(k)(1 - \lambda_k^{-1})(\boldsymbol{Q}_k^n\boldsymbol{v}_k)(\boldsymbol{Q}_k^n\boldsymbol{v}_k)^T. \quad (5)$$

Eq. (5) can be applied to estimate the analysis errors for any model variables at any levels. Different calibration schemes are discussed in the following sections.

3 Eigenvalues and eigenvectors

The diagonal part of $\boldsymbol{C}$ in Eq. (5) represents the reduction of error variance due to observations in DA. The following experiments were 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 $K = 30, 60, 100$, and 116. The observations and background fields are for the 00:00 Z cycle on 10 April 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 850 mb from satellites JMA IR (Japan Meteorological Agency Infrared) and EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites).

Figure 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 $K = 30, 60, 100$, and 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 $\mathrm{diag}[(\boldsymbol{Q}_k^n\boldsymbol{v}_k)(\boldsymbol{Q}_k^n\boldsymbol{v}_k)^T] > \mathbf{0}.0$, it follows that the diagonal elements of $\boldsymbol{C}$ are always positive, which means

$$\mathrm{diag}(\boldsymbol{A}) < \mathrm{diag}(\mathbf{B}).$$

This indicates that the analysis error variance is always less than the background error variance due to the observations assimilated. From Eq. (5), 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 zero, and the

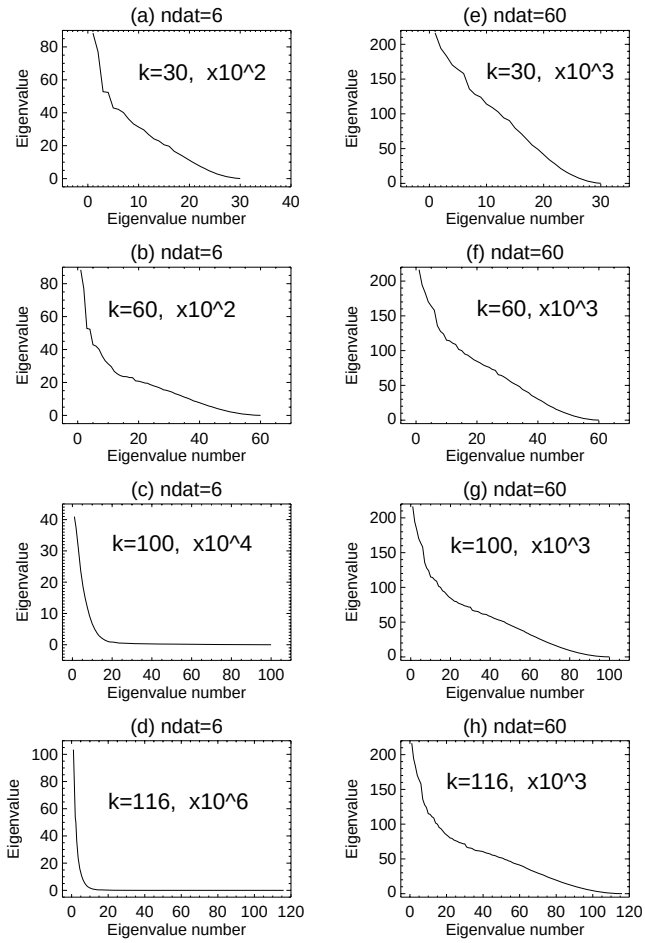

Fig. 1. The eigenvalue distribution: **(a)** ndat = 6, $K = 30$, **(b)** ndat = 6, $K = 60$, **(c)** ndat = 6, $K = 100$, **(d)** ndat = 6, $K = 116$, **(e)** ndat = 60, $K = 30$, **(f)** ndat = 60, $K = 60$, **(g)** ndat = 60, $K = 100$, and **(h)** ndat = 60, $K = 116$.

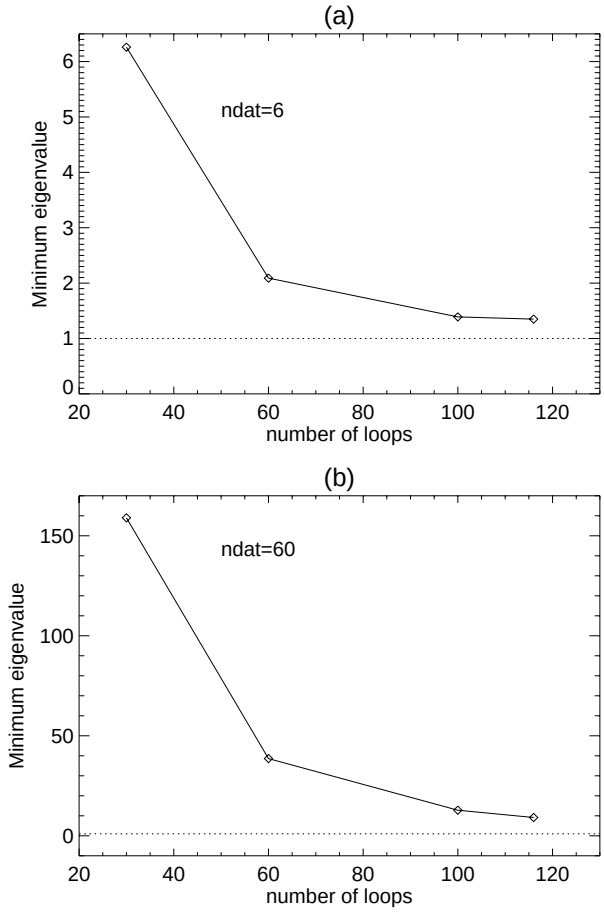

Fig. 2. The smallest eigenvalues as a function of the number of loops **(a)** for ndat = 6 and **(b)** for ndat = 60.

impact is negligible. The smallest eigenvalues for ndat = 6 shown on the left panel of Fig. 1 for $K = 30, 60, 100$, and 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 a 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 Fig. 1a–d, the eigenvalues decrease faster than when there are more observations, as in Fig. 1e–h. In fact, when all the observations are used with ndat = 60, the eigenvalues decrease at a similar rate with increasing number of inner loops. The maximum eigenvalues are very similar in all cases with different values of K, as shown in Fig. 1e–h.

As an example of eigenvector structure, we plotted in Figs. 3 and 4 the top five normalized eigenvectors ($\boldsymbol{Q}_k^n \boldsymbol{v}_k$) for temperature t and zonal wind component u for $K = 30, 60, 100$, and 116 at 500 hPa. The number of data sets is ndat = 6. From left to right, Fig. 3 shows the largest 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 North America for $K = 30$ and 60, while for $K = 100$ and 116, their structures are also similar, but different from when $K = 30$ or 60. This is also true for the other dominant eigenvectors 2, 3, 4, and 5 shown in the following rows. When $K = 30$, eigenvectors 1, 2, and 3 have high values over North America, but in slightly shifted positions, while eigenvectors 4 and 5 have larger values over the European region. All of these reflect the fact that there is relatively more conventional data coverage in North America and European regions. When $K = 60$, the top 5 eigenvectors in column 2 show similar structures as

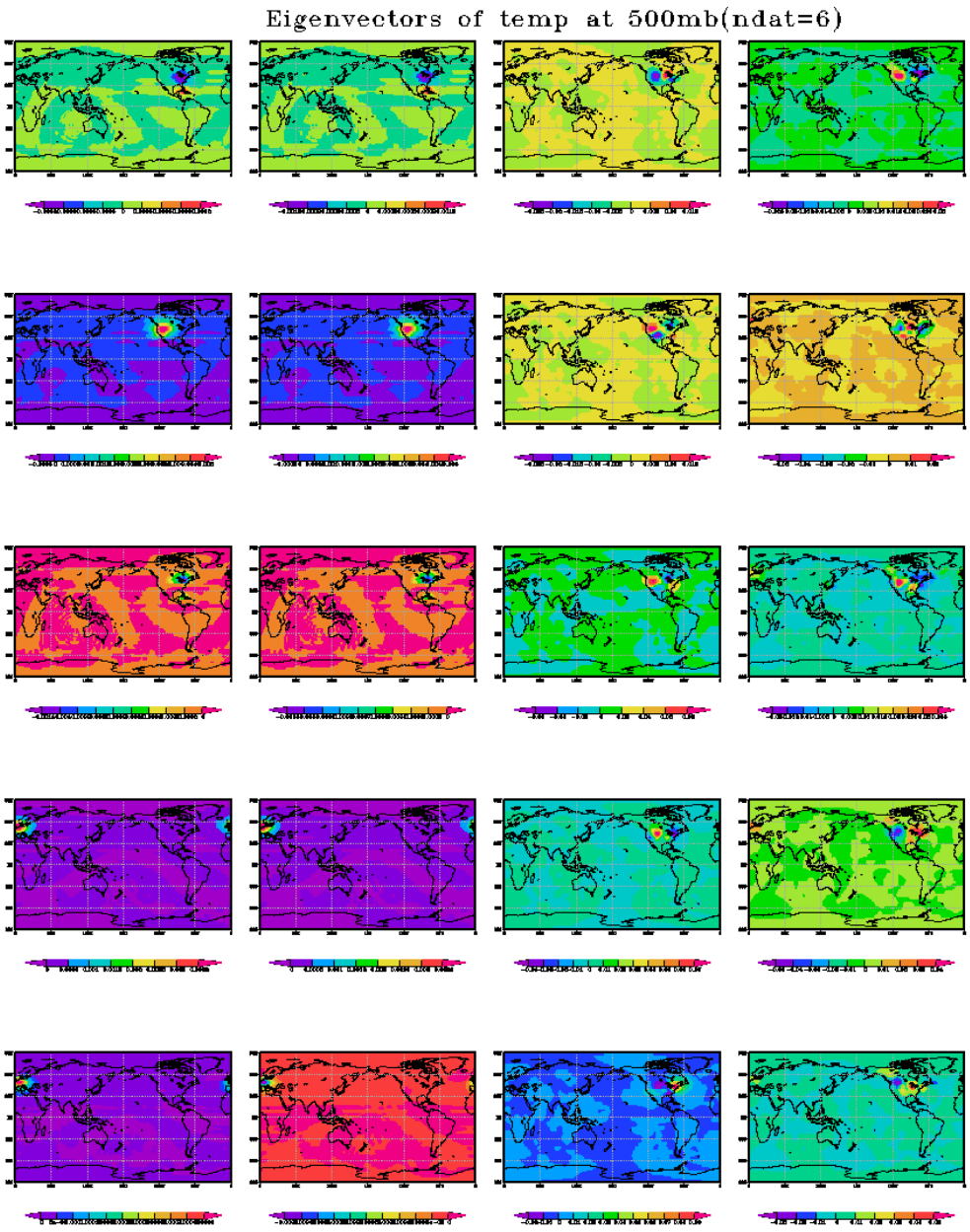

Fig. 3. Top five normalized eigenvectors of the transformed Hessian with respect to temperature at 500 hPa with ndat = 6. From the left to the right on each row: top eigenvectors for $K = 30, 60, 100,$ and 116. Top to the bottom on each column: eigenvectors 1, 2, 3, 4, and 5 for each value of K.

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 North America area. This is consistent with the conventional observations at this level, which will be shown later in the paper.

Figure 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 North America. Like the eigenvectors of t, eigenvectors 4 and 5 show the largest amplitudes over the European region when $K = 30$ or 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 Eq. (4), the analysis error variances depend on the background error variances, which are static and pre-defined in most 3-D/4-D-Var

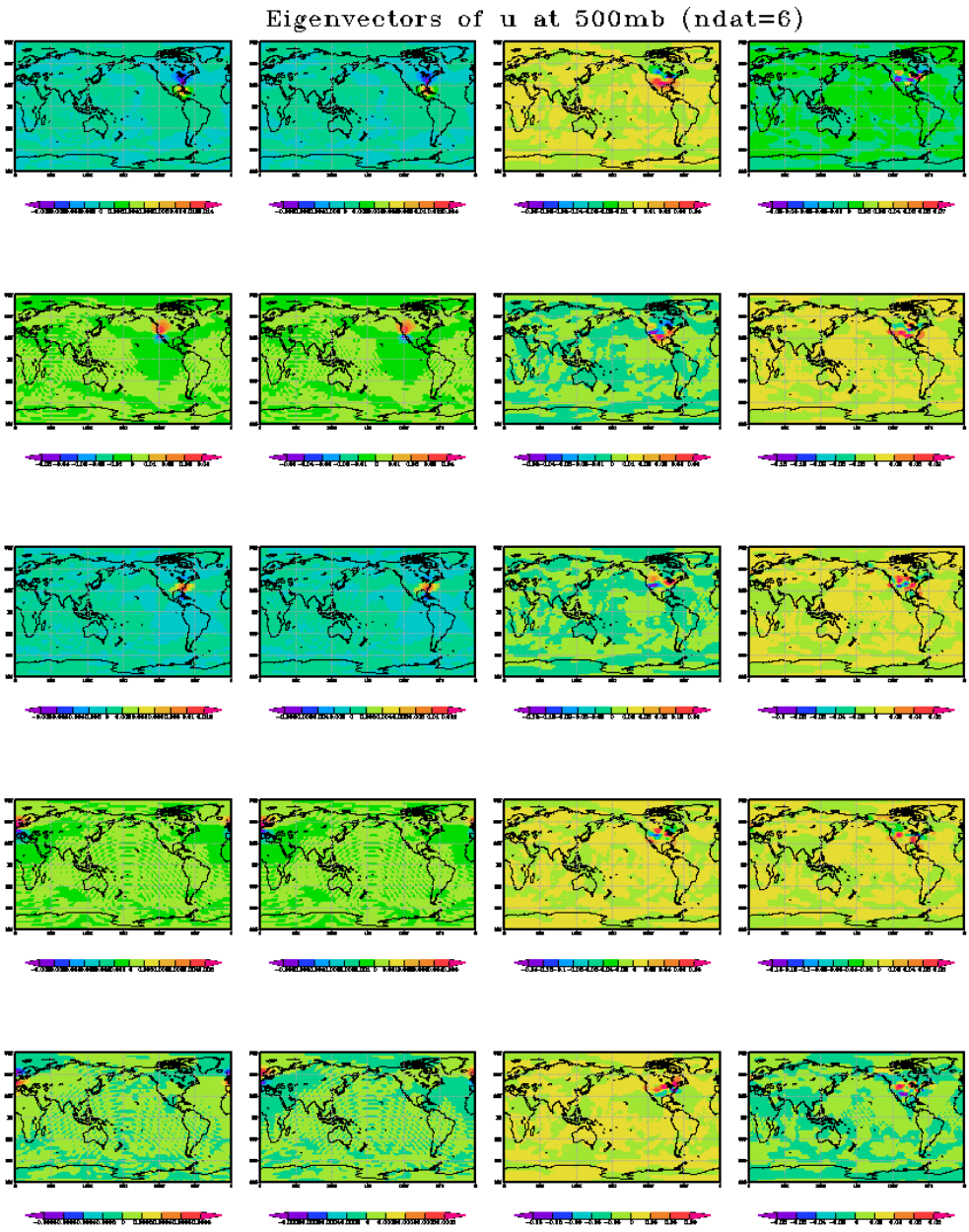

Fig. 4. The same as Fig. 3, but for u.

DA systems. In the NCEP GSI, they are pre-computed using the NMC (National Meteorological Center) method (Parrish and Derber, 1992). The diagonal components of $\boldsymbol{C}$ 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 $\boldsymbol{C}$ (instead of $\boldsymbol{A}$) 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., $\sqrt{\mathrm{diag}(\boldsymbol{C})}$ (referred to as *observation impact signal* or OIS in the following) for different variables at different levels. Similar quantities, such as information content, relative entropy, degrees of freedom for signal, and mutual information, for quantifying the impacts of observations in a DA system have been introduced and studied by Rodgers (2000), Xu (2007), Zupanski et al. (2007) and Fowler and van Leeuwen (2012).

First, we consider the OIS without calibration, i.e., $\rho(k) = 1.0$ in Eq. (5). From the left to the right, Fig. 5 shows the OIS for u_{500}, t_{500}, and q_{1000} (the subscripts indicate the level in hPa) with different numbers of inner loops, $K = 30, 60, 100$, or 116. 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}, and q_{1000} with ndat = 6, while rows 4 to 6 show the OIS for the same variables but for

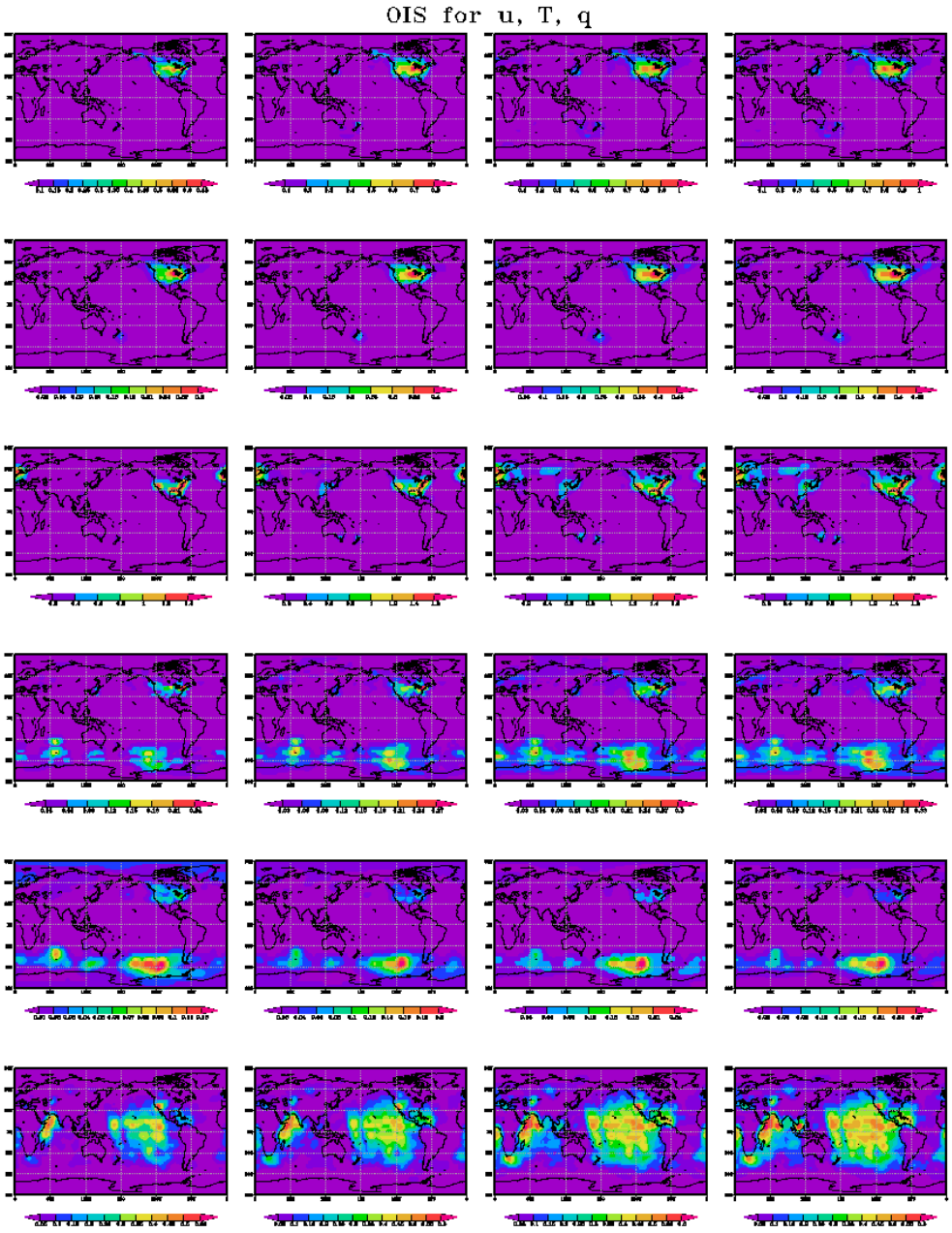

Fig. 5. Observation impact signal (OIS) for u_{500}, t_{500}, and q_{1000}. From the left to the right on each row, $K = 30, 60, 100$, and 116. Rows 1 to 3 for u_{500}, t_{500}, and q_{1000} with ndat = 6, and rows 4 to 6 for ndat = 60.

ndat = 60. The top two panels show that as K increases, the impacts of observations for u_{500} and t_{500} increase, particularly over the conventional data-rich regions such as North America. 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 $K = 100$ and 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 North America, Europe, Asia and areas near Australia, where there are more observations. The observation locations will be shown in Figs. 9–12.

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 the three variables. In ocean areas in the tropics, 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, and q for $K = 30, 60, 100$, and 116 at three typical model levels, namely $L = 1$ (lowest level, about 1000 mb), $L = 25$ (about 500 mb) and $L = 64$ (highest model level, about 0.27 mb). For each of variables x at each level, we calculate the ratios of the mean OIS values with $K = 30, 60, 100$, and 116 to the mean OIS value with $K = 30$, i.e., $\frac{\overline{\mathrm{OIS}(x,L,K)}}{\overline{\mathrm{OIS}(x,L,30)}}$. These ratios are shown as a function of the number of loops in Fig. 6. The left panels of Fig. 6a–c shows these ratios for u, t, and q, respectively with ndat = 6, while Fig. 6d–f show the same, but with ndat = 60.

With the smaller number of observations (ndat = 6), Fig. 6a and b shows that the mean OIS values of u and t at all three levels for $K = 60, 100$, and 116 are about 1.5, 2.0, and 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 the highest model level ($L = 64$) where values reach over 4.5 for $K = 100$ and 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 Fig. 6d, e and f, the conclusions are similar, except that the ratio for u at the top model level is 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 the 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 OIS values have not converged when $K = 116$, even with the 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. To test larger K was computationally prohibitive; it would be possible, but it is hard to bypass the GSI criterion. In addition, computational cost is always a concern in an operational system like GSI. To account for the impacts from the missing eigenvectors, calibration is inevitable.

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

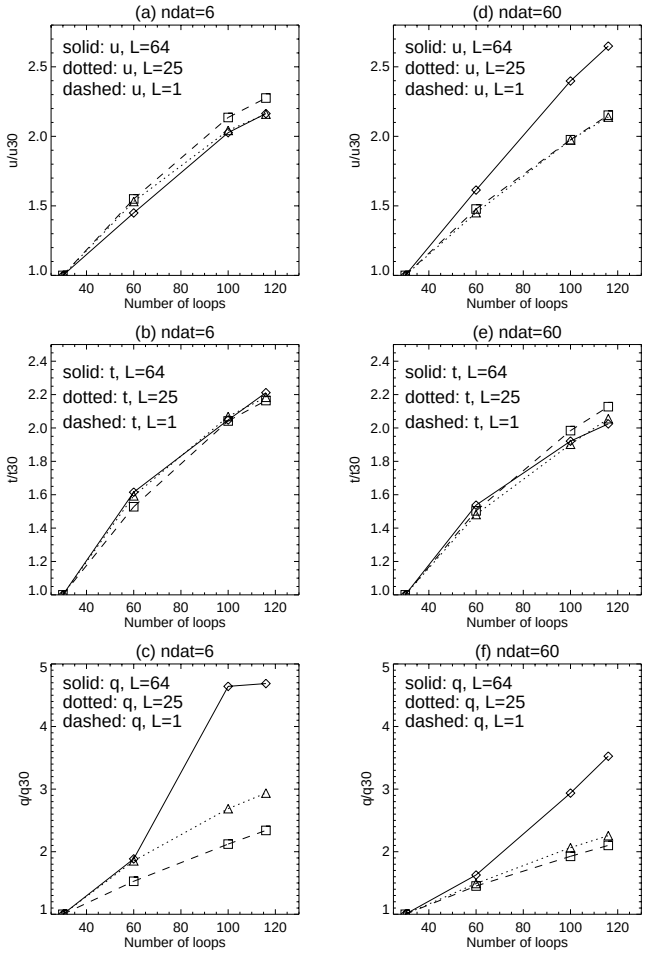

Fig. 6. The ratio of the mean OIS values using $K = 30, 60, 100$, and 116 to the mean OIS value using $K = 30$ as a function of the number of loops at 1000 hPa, 500 hPa and 0.27 hPa for: **(a)** u with ndat = 6, **(b)** t with ndat = 6, **(c)** q with ndat = 6, **(d)** u with ndat = 60, **(e)** t with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

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

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

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

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

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 weight is given to the less dominant eigenvectors, reflecting the fact that the trailing eigenvectors contribute less to the OIS in Eq. (5).

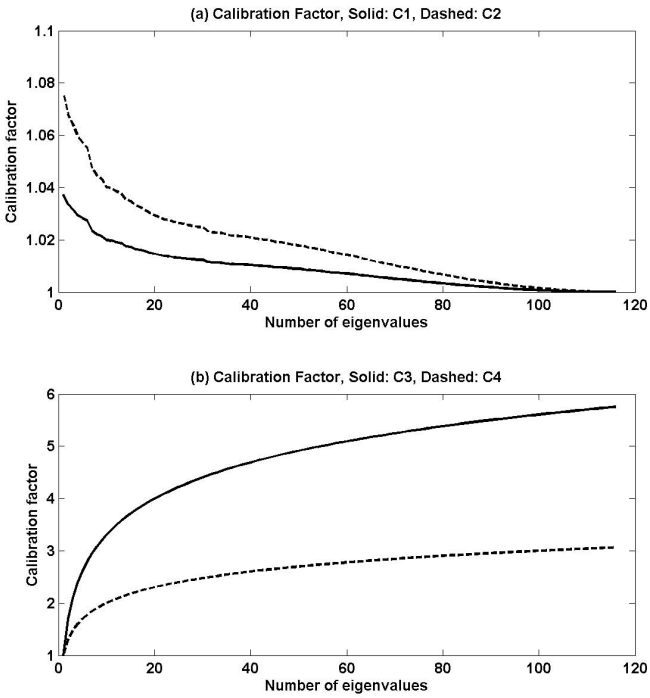

Fig. 7. Four calibration factors as functions of number of eigenvalues for $C1$ (solid) and $C2$ (dashed) in **(a)** and $C3$ (solid) and $C4$ (dashed) in **(b)** for $K = 116$ and ndat $= 60$.

Unlike these two schemes, the calibration factors defined in schemes $C3$ and $C4$ increase as a function of k. This means that greater weight is 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. The calibration factors in these four schemes are shown in Fig. 7 as functions of number of eigenvalues for $K = 116$ and ndat $= 60$. 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}, and 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 benchmark 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. 8 show the OIS values from schemes $C1, C2, C3$, and $C4$, respectively. Similar to Fig. 5, rows 1 to 3 and rows 4 to 6 in Fig. 8 are the OIS for ndat $= 6$ and ndat $= 60$, respectively.

By comparison with the results without calibration (first column of Fig. 5), the OIS values from calibration $C1$ in first column of Fig. 8 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 benchmark (last column of Fig. 5). Scheme $C2$ (column 2 of Fig. 8) also shows similar impact patterns and magnitudes as scheme $C1$. Both of these schemes can not achieve the ideal results we have hoped for.

The results from scheme $C3$ (column 3 of Fig. 8) show somewhat larger maximum values than the benchmark for all the variables, while the magnitudes of OIS from $C4$ (column 4 of Fig. 8) are similar to those in the benchmark. However, like $C1$ and $C2$, both schemes $C3$ and $C4$ do not seem to pick up the impact patterns over regions in Europe, Asia and Australia as well as the benchmark. Overall, scheme $C3$ is probably the best among these given the fact that the benchmark 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 benchmark 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 greatly 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 that are located between the vertical levels of $p - 50$ mb and $p + 50$ mb, and compare them with the OIS values computed from Eq. (5) at level p. As an example, we choose $K = 100$ and conventional data set only with ndat $= 6$ in the following.

Figure 9 shows the locations of temperature observations between 500 mb plus and minus 50 mb, and the OIS for temperature at 500 mb. It is clear that the region with most conventional data is North America, 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.

Figure 10 shows the same as Fig. 9, 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. 11, we display the same as Fig. 10, but for surface level (around 1000 mb). Around this lower level, there are many satellite-derived,

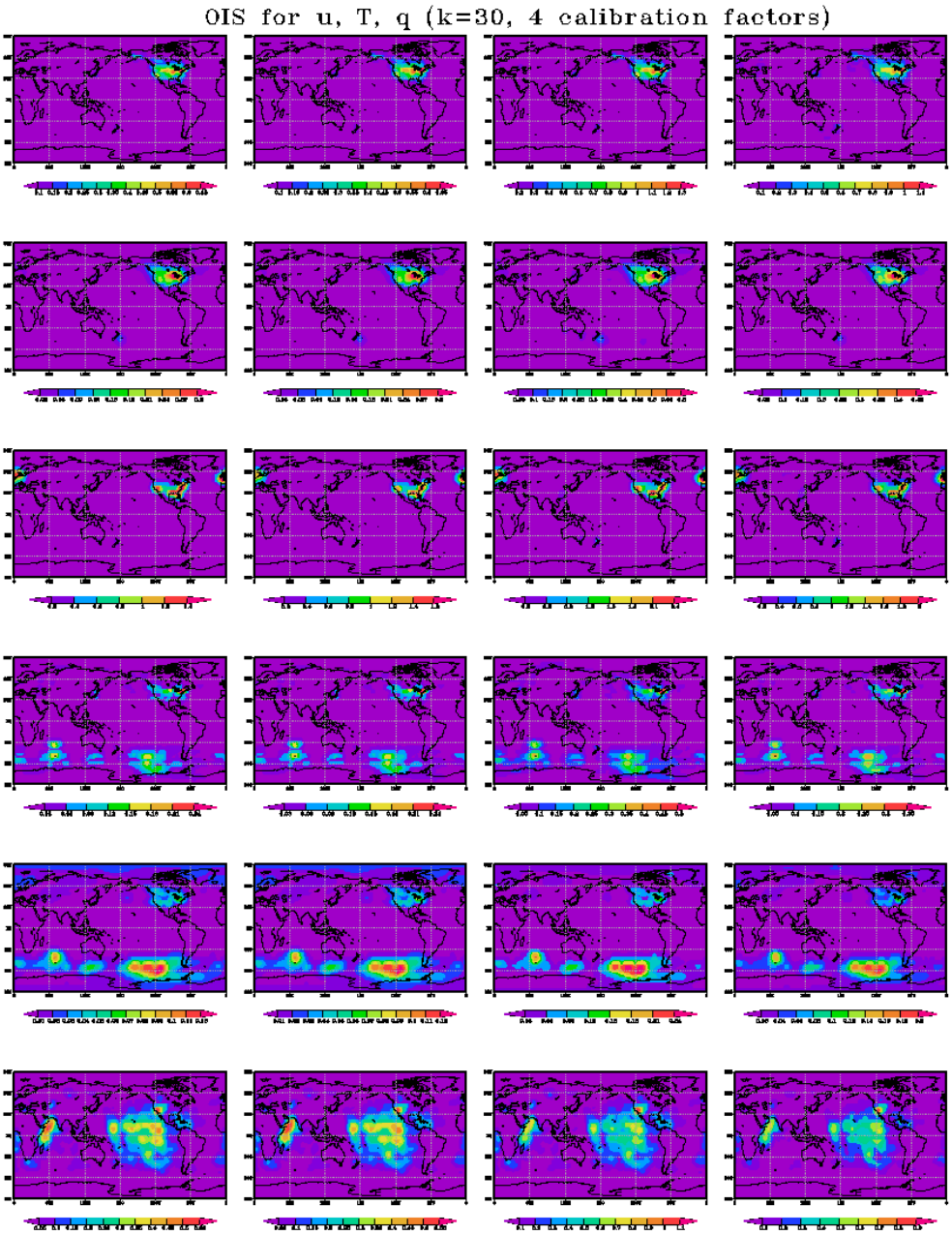

Fig. 8. OIS with four calibration schemes for u_{500}, t_{500}, and q_{1000}. From the left to the right on each column: OIS values from schemes $C1, C2, C3$, and $C4$, respectively. Rows 1 to 3 on each column: for u_{500}, t_{500}, and q_{1000} with ndat = 6. Rows 4 to 6 on each column: the same variables for ndat = 60.

scatterometer-type wind observations (such as from JMA IR and EUMETSAT) in SH and the tropics as shown in Fig. 11. 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 more eigenvectors to cover these satellite data, particularly in the tropics.

Figure 12 is the same as Fig. 11, but for relative humidity q_{1000}. The observations over Europe are equally as dense as over North America. 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) that are missed by the OIS.

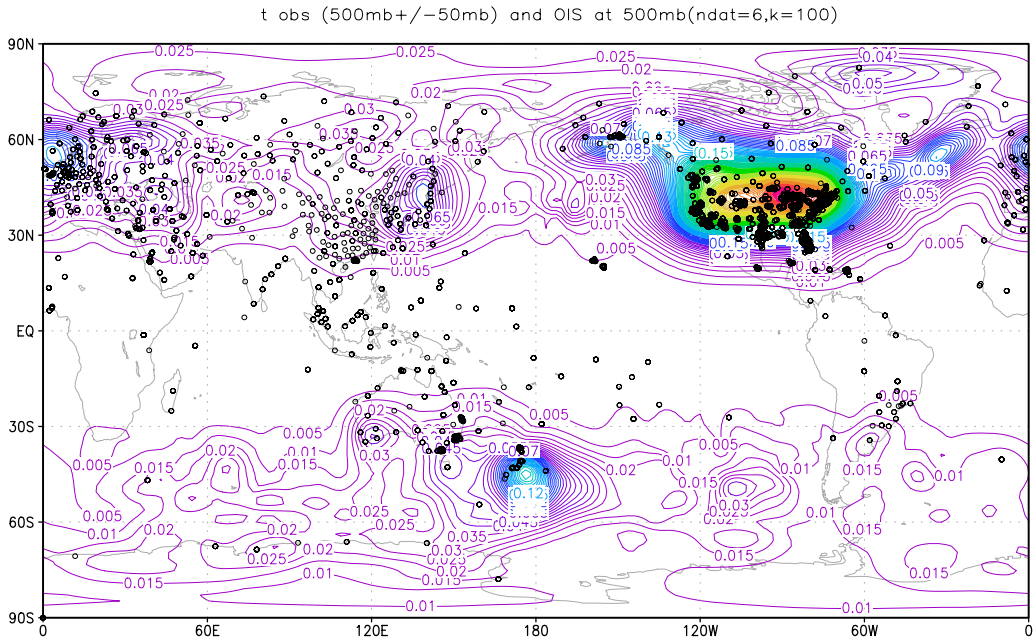

Fig. 9. Circles represent the locations of temperature observations between 500 mb − 50 mb and 500 mb + 50 mb, and contour lines represent the OIS for temperature at 500 mb.

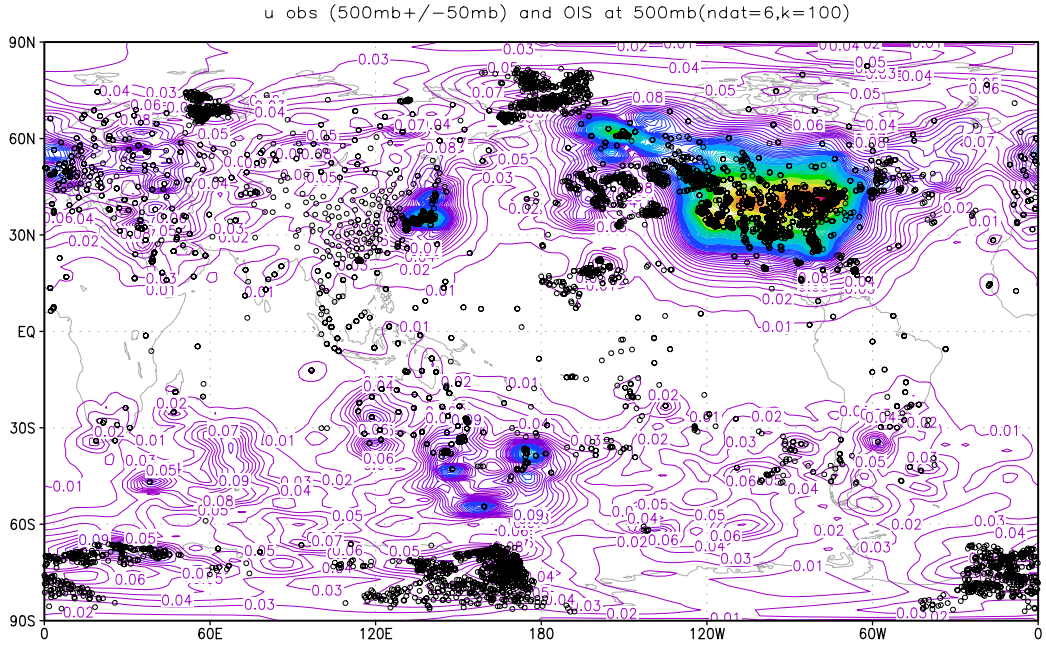

Fig. 10. The same as Fig. 9, but for u.

As described in previous sections, OIS is able to capture some of the observation impacts, especially over the regions with dense conventional observations. Results in this section indicate that 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 **B**) around the tropics are lower than in the extra-tropics. Since the observation errors do not depend on latitude in the GSI, the smaller background errors around

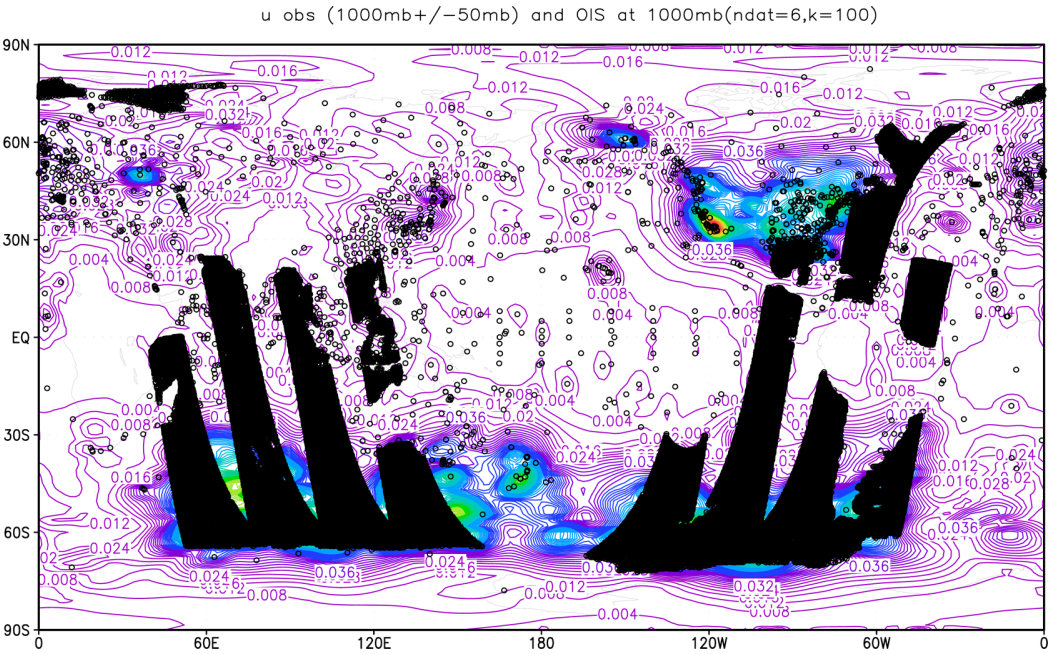

Fig. 11. The same as Fig. 10, but for surface level (around 1000 mb).

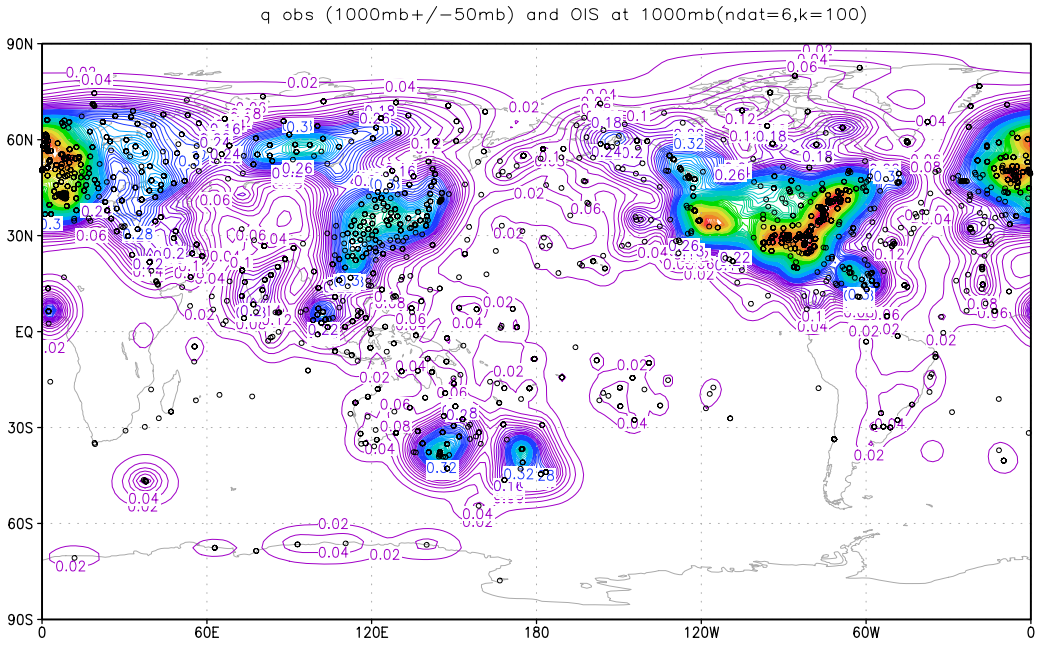

Fig. 12. The same as Fig. 11, but for relative humidity q.

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 Eq. (2).

Corresponding to the observation locations and the OIS in Figs. 9–12, Fig. 13a–d shows the analysis increments for t_{500}, u_{500}, u_{1000}, and q_{1000}. If we compare Figs. 9 and 10 with Fig. 13a and b, 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. 11 and 12 are compared with Fig. 13a and d, 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.

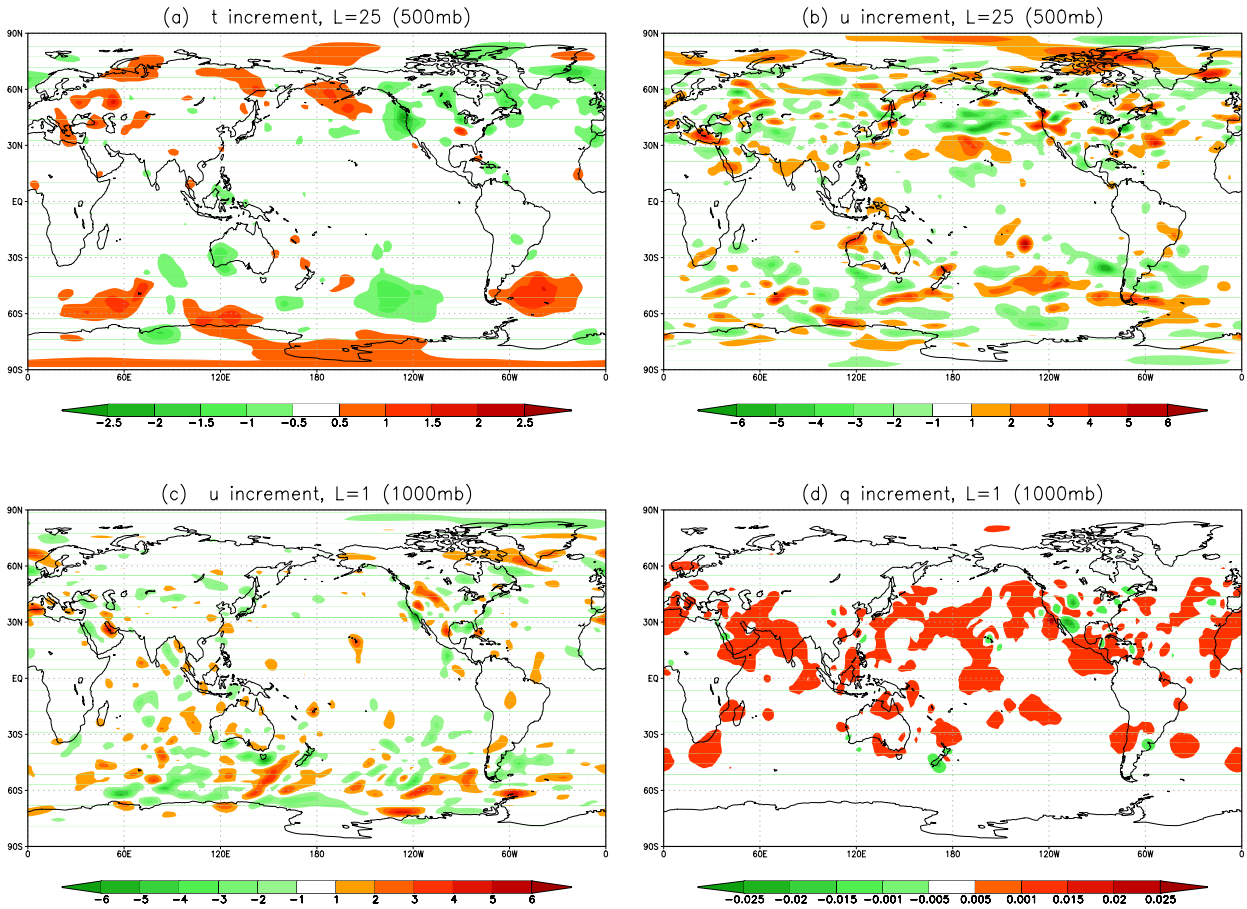

Fig. 13. The analysis increments of **(a)** t_{500}, **(b)** u_{500}, **(c)** u_{1000}, and **(d)** q_{1000}.

6 Discussion and conclusions

This paper represents another effort in estimating the analysis error variance, using the Lanczos method proposed by Fisher and Courtier (1995), in the NCEP global 3-D-Var GSI DA system. We have applied this method to the global 3-D-Var GSI and studied other different aspects of this method that were not exploited in Fisher and Courtier (1995) and Pondeca and Manikin (2009). The properties of convergence in different calibration schemes discovered in this paper have greatly improved our understanding of the method and its implications in practical applications in an operational environment. 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 direct measure of the error reduction due to the observations assimilated.

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 convergence 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 largest eigenvectors when K is small show larger impacts in the regions where conventional data are dominant. If K is increased, the OIS can represent other areas with fewer 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 a 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 fewer 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, and 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 $K = 60, 100$, and 116 are about 1.5, 2.0, and 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 estimates of observation impacts are not accurate. 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 a 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. A lot more eigenvectors are required to recover the whole satellite observation impacts. The area in which the method performs worst is 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 major 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 3-D/4-D-Var system to estimate the observation impacts, and it can be used as part of a routine verification package.

Appendix A

Basic formulation of Lanczos algorithm in 3-D-Var GSI

The basic 3-D-Var equations in GSI are described in Eqs. (1)–(3). 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

$$\boldsymbol{z} = \mathbf{B}^{-1}\boldsymbol{x}. \tag{A1}$$

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

$$\begin{aligned}\boldsymbol{g} = \nabla_{\boldsymbol{x}} J(\boldsymbol{x}) &= (\mathbf{B}^{-1} + \boldsymbol{H}^T\mathbf{R}^{-1}\boldsymbol{H})\boldsymbol{x} - \boldsymbol{H}^T\mathbf{R}^{-1}\boldsymbol{y}_0 \\ &= \mathbf{M}x - \boldsymbol{H}^T\mathbf{R}^{-1}\boldsymbol{y}_0,\end{aligned} \tag{A2}$$

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

$$\mathbf{M} = \frac{\partial^2 J(\boldsymbol{x})}{\partial \boldsymbol{x}^2} = \mathbf{B}^{-1} + \boldsymbol{H}^T\mathbf{R}^{-1}\boldsymbol{H} = \boldsymbol{A}^{-1}. \tag{A3}$$

Equation (A2) is equivalent to

$$\mathbf{M}x = \boldsymbol{g} + \boldsymbol{H}^T\mathbf{R}^{-1}\boldsymbol{y}_0. \tag{A4}$$

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

$$\boldsymbol{h} = \nabla_{\boldsymbol{z}} J(\boldsymbol{x}) = \mathbf{B}\nabla_{\boldsymbol{x}} J(\boldsymbol{x}) = \boldsymbol{x} + \mathbf{B}H^T\mathbf{R}^{-1}(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}_0). \tag{A5}$$

The preconditioned conjugate gradient method used to minimize the cost function defined in Eq. (1) in GSI can be expressed as:

$$\begin{aligned}\boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \\ \boldsymbol{g}_{k+1} &= \nabla_{\boldsymbol{x}} J(\boldsymbol{x}_{k+1}) \\ \boldsymbol{h}_k &= \mathbf{B}g_{k+1} \\ \boldsymbol{d}_{k+1} &= -\boldsymbol{h}_{k+1} + \beta_k \boldsymbol{d}_k \\ \boldsymbol{e}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{e}_k,\end{aligned} \tag{A6}$$

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

$$\beta_k = \frac{(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)^T \boldsymbol{h}_{k+1}}{(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)^T \boldsymbol{d}_k}. \tag{A7}$$

$\boldsymbol{g}_k$ and $\boldsymbol{h}_k$ are two independent variables and can be normalized such that

$$\boldsymbol{g}_k^n = c_k \boldsymbol{g}_k \text{ and } \boldsymbol{h}_k^n = c_k \boldsymbol{h}_k, \tag{A8}$$

where $c_k = \frac{1}{\sqrt{\boldsymbol{g}_k^T \boldsymbol{h}_k}}$.

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

$$\left\langle \boldsymbol{g}_i^n,\ \boldsymbol{h}_{\mathrm{j}}^n \right\rangle = \delta_{ij}, \tag{A9}$$

where δ_{ij} is a Kronecker delta function.

By using the relations in Eqs. (A1)–(A9), it can be shown that the final analysis error covariance matrix can be expressed as

$$\boldsymbol{A} = \mathbf{B} - \sum_{k=1}^{m} (1 - \lambda_k^{-1})(\boldsymbol{Q}_k^n \boldsymbol{v}_k)(\boldsymbol{Q}_k^n \boldsymbol{v}_k)^T, \tag{A10}$$

where m is the total number of inner iterations and also the number of gradients in $\boldsymbol{x}$ and $\boldsymbol{z}$, λ_k and $\boldsymbol{v}_k$ are the dominant eigenvalues and eigenvectors of a $m \times m$ tri-diagonal matrix $\mathbf{T}_m$ consisting of different coefficients as

$$\mathbf{T}_m = \begin{bmatrix} \delta_1 & \theta_1 & 0 & 0 \\ \eta_1 & \delta_2 & \theta_2 & 0 \\ 0 & \eta_2 & \delta_3 & \bullet \\ 0 & 0 & \bullet & \bullet \end{bmatrix}, \tag{A11}$$

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

$$\boldsymbol{Q}_m^n = [\boldsymbol{h}_1^n, \boldsymbol{h}_2^n, \boldsymbol{h}_3^n \ \ldots\ldots \boldsymbol{h}_{\mathrm{m}}^n]. \tag{A12}$$

It can be shown that λ_k and $\boldsymbol{e}_k = \boldsymbol{Q}_k^n \boldsymbol{v}_k$ are the dominant eigenvalues and eigenvectors of the transformed Hessian matrix.

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 Steve Lord for his constant support and encouragement during this work, Mary Hart for improving the presentation, and Carolyn Reynolds and Craig Bishop at NRL for helpful discussion and sharing the FNMOC analysis error estimates with us. M. Wei was funded through SEMESTER 6.2 project at NRL and supported by the Office of Naval Research (Program Element 0602435N).

Edited by: S. Vannitsem
Reviewed by: P. M. Jermey and one anonymous referee

References

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

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

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

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

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

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

Daley, R. and Barker, E.: NAVDAS: Formulation and Diagnostics, Mon. Weather Rev., 129, 869–883, 2001.

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

Derber, J. C. and Rosati, A.: A global oceanic data assimilation system, J. Phys. Oceanogr., 19, 1333–1347, 1989.

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

Derber, J. C., Purser, R. J., Wu, W.-S., Tredon, R., Pondeca, M., Parish,D., and Kleist, D.: 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, 2003.

Errico, R. M., Yang, R., Masutani, M., and Woollen, J.: The estimation of analysis error characteristics using an observation system simulation experiment, Meteorol. Z., 16, 695–708, 2007.

Fisher, M.: 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, 1998.

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

Fowler, A. and van Leeuwen, P. J.: Measures of observation impact in non-Gaussian data assimilation, Tellus, 64A, 17192, doi:10.3402/tellusa.v64i0.17192, 2012.

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

Houtekamer, P. L., Lefaivrem, L., Derome, J., Ritchie, J., and Mitchell, H. L.: A system simulation approach to ensemble prediction, Mon. Weather Rev., 124, 1225–1242, 1996.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J.: 4-D-Var or ensemble Kalman filter?, Tellus, 59A, 758–773, 2007.

Kleist, D. T., Parrish, D., Derber, J. C., and Treadon, R.: Improving incremental balance in the GSI 3D Var analysis system, Mon.

Weather Rev., 137, 1046–1060, 2009.

Langlang, R. H., Maue, R. N., and Bishop, S. H.: Uncertainty in atmospheric temperature analysis, Tellus, 60A, 598–603, 2008.

Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, J. Comput. Phys., 227, 3515–3539, 2008.

Manikin, G. S. and Pondeca, M.: 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, 2009.

McLay, J. and Reynolds, C. A.: Two alternative implementations of the ensemble-transform (ET) analysis-perturbation scheme: The ET with extended cycling intervals, and the ET without cycling, Q. J. Roy. Meteor. Soc., 135, 1200–1213, 2009.

McLay, J., Bishop, C. H., and Reynolds, C. A.: The ensemble-transform scheme adapted for the generation of stochastic forecast perturbations, Q. J. Roy. Meteor. Soc., 133, 1257–1266, 2007.

McLay, J., Bishop, C. H., and Reynolds, C.A.: Evaluation of the ensemble transform analysis perturbation scheme at NRL, Mon. Weather Rev., 136, 1093–1108, 2008.

Molteni, F., Buizza, R., Palmer, T., and Petroliagis, T.: The ECMWF ensemble prediction system: Methodology and validation, Q. J. Roy. Meteor. Soc., 122, 73–119, 1996.

Park, Y.-Y., Buizza, R., and Leutbecher, M.: TIGGE: Preliminary results on comparing and combining ensembles, Q. J. Roy. Met. Soc., 134, 2029–2050, 2008.

Parrish, D. F. and Derber, J.: The National Meteorological Center's spectral statistical-interpolation analysis system, Mon. Weather Rev., 120, 1747–1763, 1992.

Pondeca, M. and Manikin, G. S.: 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, 2009.

Pondeca, M., Manikin, G. S., DiMego, G., Benjamin, S., Parrish, D., Purser, J., Wu, W.-S., Horel, J., Myrick, D., Ling, Y., Aune, R., Keyser, D., Colman, B., Mann, G., and Vavra, J.: The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: current status and development, Weather Forecast., 26, 593–612, 2011.

Rabier, F., Jarvine, H.,.Klinker, E., Mahfouf, J., and Simmons, A.: The ECMWF operational implementation of four-dimensional variational assimilation, I: experimental results with simplified physics, Q. J. Roy. Meteor. Soc., 126, 1143–1170, 2000.

Reynolds, C. A., Teixeira, J., and McLay, J. G.: Impact of stochastic convection on the ensemble transform, Mon. Weather Rev., 136, 4517–4526, 2008.

Rodgers, C. D. : Inverse methods for atmospheric sounding: Theory and practice, World Scientific Publishing, Singapore, 2000.

Swanson, K. L. and Roebber, P.: The impact of analysis error on medium-range weather forecasts, Mon. Weather Rev., 136, 3425–3431, 2008.

Szunyogh, I., Kostelich, E. J., Gyarmati, G., Kalnay, E., Hunt, B., Ott, E., Satterfield, E., and York, J. A.: A local ensemble transform Kalman filter data assimilation system for the NCEP global model, Tellus, 60A, 113–130, 2008.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T., and Whitaker, J. S.: Ensemble square root filters, Mon. Weather Rev., 131, 1485–1490, 2003.

Toth, Z. and Kalnay, E.: Ensemble forecasting at NMC: the generation of perturbations, B. Am. Meteorol. Soc., 174, 2317–2330, 1993.

Toth, Z. and Kalnay, E.: Ensemble forecasting at NCEP and the breeding method, Mon. Weather Rev., 125, 3297–3319, 1997.

Wei, M. and Toth, Z.: A new measure of ensemble performance: Perturbations versus Error Correlation Analysis (PECA), Mon. Weather Rev., 131, 1549–1565, 2003.

Wei, M., Z. Toth, Z., Wobus, R., Zhu, Y., and Bishop, C.: Initial Perturbations for NCEP Ensemble Forecast System. Thorpex Symposium Proceedings for the 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, 227–230, 2005.

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

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

Wei, M., Toth, Z., and Zhu, Y. : Analysis differences and error variance estimates from multi-center analysis data, Austr. Meteorol. Oceanogr. J., 59, 25–34, 2010.

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

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

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

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

Xu, Q.: Measuring information content from observations for data assimilation: relative entropy versus Shannon entropy difference, Tellus, 59A, 198–209, 2007.

Zupanski, D., Hou, A. Y., Zhang, S. Q., Zupanski, M., Kummerow, C. D., and Cheung, S. H.: Application of information theory in ensemble data assimilation, Q. J. Roy. Meteor. Soc., 133, 1533–1545, 2007.

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