Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Ensemble singular vectors and their use as additive inflation in EnKF

By SHU-CHIH YANG[1]*, EUGENIA KALNAY[2] and TAKESHI ENOMOTO[3], [1]*National Central University, Taoyuan, Taiwan*; [2]*University of Maryland, College Park, MD, USA*; [3]*Kyoto University, Sakyō-ku, Kyoto, Japan*

(Manuscript received 2 November 2014; in final form 23 June 2015)

ABSTRACT

Given an ensemble of forecasts, it is possible to determine the leading ensemble singular vector (ESV), that is, the linear combination of the forecasts that, given the choice of the perturbation norm and forecast interval, will maximise the growth of the perturbations. Because the ESV indicates the directions of the fastest growing forecast errors, we explore the potential of applying the leading ESVs in ensemble Kalman filter (EnKF) for correcting fast-growing errors. The ESVs are derived based on a quasi-geostrophic multi-level channel model, and data assimilation experiments are carried out under framework of the local ensemble transform Kalman filter. We confirm that even during the early spin-up starting with random initial conditions, the final ESVs of the first analysis with a 12-h window are strongly related to the background errors. Since initial ensemble singular vectors (IESVs) grow much faster than Lyapunov Vectors (LVs), and the final ensemble singular vectors (FESVs) are close to convergence to leading LVs, perturbations based on leading IESVs grow faster than those based on FESVs, and are therefore preferable as additive inflation. The IESVs are applied in the EnKF framework for constructing flow-dependent additive perturbations to inflate the analysis ensemble. Compared with using random perturbations as additive inflation, a positive impact from using ESVs is found especially in areas with large growing errors. When an EnKF is 'cold-started' from random perturbations and poor initial condition, results indicate that using the ESVs as additive inflation has the advantage of correcting large errors so that the spin-up of the EnKF can be accelerated.

Keywords: singular vector, dynamic sensitivity, ensemble Kalman filter, data assimilation

1. Background

In ensemble Kalman filter (EnKF), covariance inflation is a common strategy to compensate for the under-estimation of ensemble-based background error covariance due to the use of a limited ensemble size and imperfect models. Covariance inflation can be classified into three types. Multiplicative covariance inflation (Anderson and Anderson, 1999) increases the amplitude of the error covariance without modifying the structure, so that the weight given to the overconfident model state is reduced. Additive covariance inflation (Whitaker et al., 2008) aims to perturb the subspace spanned by the ensemble vectors and better capture the sub-growing directions that may be missed in the original ensemble. Finally, Zhang et al. (2004) relaxed the analysis ensemble vectors to the background ensemble so that the subspace of the ensemble space is not over-shrunk. Considering that the issue of rank deficiency becomes even more harmful when EnKF is performed with a small ensemble size, additive inflation is expected to be particularly beneficial for improving the EnKF performance. Whitaker et al. (2008) found that additive inflation was more effective than multiplicative inflation. The concept of additive inflation was demonstrated in Corazza et al. (2003) with a quasi-geostrophic (QG) channel model. Adding a small amount of the random perturbations on the bred vectors (BV) helps BVs capture sub-growing directions so that BVs better project on the background errors. In Yang et al. (2006, 2009), Corazza et al. (2007), Kalnay et al. (2007), random perturbations are used as the additive covariance inflation in the EnKF assimilation framework. However, in realistic models, using random perturbations to perturb the ensemble may introduce directions that are irrelevant with respect to the underlying background flow. This noise can even be accumulated in areas with sparse observations. To generate

*Corresponding author.
email: shuchih.yang@atm.ncu.edu.tw

Citation: Tellus A 2015, **67**, 26536, http://dx.doi.org/10.3402/tellusa.v67.26536

proper additive inflation for realistic models, Houtekamer et al. (2005) generated additive errors according to the 3D-Var errors covariance structures. Instead of using the large-scale and barotropic structures based on the 3D-Var errors covariance, Whitaker et al. (2008) generated additive noise by selecting random differences between adjacent 6-hourly analyses the NCEP/NCAR reanalysis, aiming to emphasise growing baroclinic synoptic-scale structures in middle latitudes. Although both methods should lead to more realistic model error statistics, it is not clear whether the additive perturbations are added in the 'desired' areas where errors are flow-dependently growing and dynamically active. To make the additive inflation effectively improve the capture of growing errors, we propose that additive perturbations should also be flow-dependent, instead of inflating the full space spanned by the ensemble vectors. To achieve this, perturbations that aim to represent the fast-growing errors should be applied to enlarge the error covariance. In this study, we propose that forecast sensitivities associated with fast-growing errors can be applied as additive inflation, so that the ensemble vectors can efficiently project on the subspace associated with the dynamically growing errors.

Perturbations associated with fast-growing errors have been extensively studied for applications such as improving the performance of ensemble prediction and targeting observations. Two methods that aim to capture the subspace of growing errors in the ensemble prediction systems are singular vectors (SV, Buizza et al., 1993; Buizza and Palmer, 1995), the computation of which is based on the tangent linear and adjoint models, and BVs (Toth and Kalnay, 1993, 1997) that use a fully non-linear model. SVs are a set of perturbation vectors that will maximise the growth of the perturbations given the choice of initial/final norms (the metric to measure the size of the perturbation) and optimisation period. The structures of SVs are very sensitive to the choice of initial and final norms and optimisation period (Errico and Vukicevic, 1992). For the purpose of operational forecasting, ECMWF uses the total energy norm at the initial time so that the energy spectrum of the initial SV best matches the spectrum of analysis errors estimated from analysis differences (Palmer et al., 1998). For tropical cyclone (TC) track prediction, moist and dry SVs derived in Japan Meteorology Agency (JMA) target on the TC-associated and mid-latitude regions (Yamaguchi et al., 2009) and moist SVs are computed with the moist total energy norm (Barkmeijer et al., 2001). The breeding cycle for deriving BV is a non-linear, finite-time, finite-amplitude generalisation of the method used to obtain the leading Lyapunov Vector (LV) (Kalnay and Toth, 1994). Independent from the choice of norms, BV captures the growing errors of interest with physically meaningful breeding parameters, including the breeding rescaling amplitude and interval (Peña and Kalnay, 2004; Yang et al., 2006). BVs with a regional rescaling were introduced in the NCEP ensemble forecasting system (Toth and Kalnay, 1997). A study by Yang (2005, Appendix B) has shown that BV and final SV are similar in shape and both strongly project on the background errors, which are the errors from a short-range forecast used as the initial guess in assimilation cycles. This confirms that both BVs and final SVs are strongly related to the dynamically growing errors. In addition to SV and BV, the ETKF rescaling scheme, which is regarded as the generalised breeding approach (Bojarova et al., 2011), also seeks to represent the dynamical growing perturbations.

For targeting observations, perturbations indicating fast-growing errors within an optimisation period can also be used to determine the sensitivity area that contributes to the forecast errors in the target region. Deploying additional observations in these areas can improve the forecast in the target region (Palmer et al., 1998; Bishop and Toth, 1999; Gelaro et al., 1999; Bishop et al., 2001). Adjoint forecast sensitivity has been applied to answer questions such as 'what would be the optimal initial analysis perturbation that leads to the best 3-d forecast?' (Rabier et al., 1996). For regional weather prediction, adjoint-based forecast sensitivity has also been well explored for understanding typhoon development (Chen et al., 2009, 2011; Wu et al., 2009; Doyle et al., 2012; Ito and Wu, 2013). Without the need of an adjoint, Bishop and Toth (1999) derived the ensemble-based SV and discussed the relationship between the ensemble-based prediction error covariance matrix and SV for measuring maximum growth. Enomoto et al. (2006, 2015) re-formulated ensemble singular vector (ESV) using a Lagrange multiplier in a simplified setting. Given an ensemble of forecasts, the ensemble sensitivities can be derived as a linear combination of the forecasts that, given the choice of the perturbation norm and forecast interval, will maximise the growth of the perturbations. Therefore, the leading ESV indicates the directions of the fastest growing forecast errors. With a dry total energy norm, ESV has been used to investigate the sensitivity of atmospheric predictability for a stratospheric sudden warming event (Nishii and Nakamura, 2010) and a blocking event (Matsueda et al., 2011).

In this study, we use the forecast sensitivity based on Enomoto et al. (2006, 2015). We refer to such ensemble sensitivity as the ESV sensitivity. We derive the ESVs from the ensemble in the EnKF assimilation implemented in a QG model (Rotunno and Bao, 1996), and explore whether these ESVs are related to the forecast errors, and whether they can be used in the EnKF system for further improving the analysis accuracy and help to capture the growing errors while the ensemble perturbations are building up the flow-dependent structures during the EnKF's spin-up period.

The paper is organised as follows. Section 2 presents the derivation of ESV and the local ensemble transform Kalman filter (LETKF) method implemented in the QG model. The QG model used in this study is introduced in Section 3 and the characteristics of the ESVs derived with this QG model are discussed in Section 4. Section 5 shows results of applying ESV as additive inflation in the QG-LETKF framework. Finally, a summary and conclusions are presented in Section 6.

2. Methodology

2.1. Ensemble singular vector

Assume we have an ensemble of m members with an initial state $\{y_{1,..}\, y_{i,..}\, y_m\}$ at time t_0, and a final state $\{z_{1,..}\, z_{i,..}\, z_m\}$ at time $t_0+\Delta t$, where $z_i = \underset{t_0 \to t_0+\Delta t}{M}(y_i)$ is the integration of the initial state y_i with the non-linear model M. The initial and final perturbations are defined with respect to the ensemble mean: $\delta y_i = y_i - \bar{y}$ are the initial state perturbations and $\delta z_i = z_i - \bar{z}$ are the final state perturbations. With $\mathbf{M}$ representing the tangent linear model of the non-linear model M, the final perturbation can be approximated as

$$\delta z_i = \underset{t_0 \to t_0+\Delta t}{\mathbf{M}} \delta y_i \text{ for all } i \tag{1}$$

Equation (1) assumes that the forecast integration (Δt) is not too long, so that the evolution of the perturbations is approximately linear. We note that the same constraint is needed in the adjoint-based sensitivity analysis method. To measure the size of perturbations, matrices $\mathbf{C}_I$ and $\mathbf{C}_F$ are used to respectively define the initial and final norms of model states or perturbations, so that $\|\delta y\| = \delta y^T \mathbf{C}_I \delta y$ and $\|\delta z\| = \delta z^T \mathbf{C}_F \delta z$.

We want to find an initial state perturbation δy such that the norm of the final state perturbation δz will be maximised among all perturbations of the same initial size 1. Using Lagrange multipliers [e.g. eq. (6.3.30) in Kalnay, 2003], a cost-function for optimisation can be defined as:

$$F(\delta y, \lambda) = \delta z^T \mathbf{C}_F \delta z + \lambda(1 - \delta y^T \mathbf{C}_I \delta y) \tag{2}$$

Given the relationship between δz and δy, from eq. (1), eq. (2) can be solved as an eigenvector problem:

$$\mathbf{M}^T \mathbf{C}_F \mathbf{M} \delta y = \lambda \mathbf{C}_I \delta y \tag{3}$$

Enomoto et al. (2006, 2015) re-formulated this for an ensemble of forecasts: Given the initial and final perturbations, the corresponding matrices of perturbations are defined as:

$$\mathbf{Y} = [\delta y_1, \delta y_2, \cdots, \delta y_m];\ \mathbf{Z} = [\delta z_1, \delta z_2, \cdots, \delta z_m] \tag{4}$$

We want to find the optimal linear combination of the perturbations such that

$$\delta y = \mathbf{Y}\mathbf{p},\ \ \delta z = \mathbf{Z}\mathbf{p} \tag{5}$$

where $\mathbf{p} = [p_1, p_2, \cdots, p_m]$ is the coefficients vector. Note that the coefficient vector, $\mathbf{p}$, is the same for the initial and final perturbations because of eq. (1), which requires the forecast interval to be enough short to assure that the forecast evolution stays close to linear.

From eqs. (1) to (4) we can write the function to be maximised with a fixed perturbation norm:

$$F(\mathbf{p}, \lambda) = \mathbf{p}^T \mathbf{Z}^T \mathbf{C}_F \mathbf{Z}\mathbf{p} + \lambda(1 - \mathbf{p}^T \mathbf{Y}^T \mathbf{C}_I \mathbf{Y}\mathbf{p}) \tag{6}$$

Making the differential with respect to $\mathbf{p}$ equal to zero, we obtain

$$\left(\mathbf{Y}^T \mathbf{C}_I \mathbf{Y}\right)^{-1} \left(\mathbf{Z}^T \mathbf{C}_F \mathbf{Z}\right)\mathbf{p} = \lambda \mathbf{p} \tag{7}$$

Equation (7) is the same as eq. (9) in Enomoto et al. (2006), and eq. (12) in Matsueda et al. (2011) (except for the absence of the norms $\mathbf{C}$). By definition, the eigenvector associated with the largest eigenvalue λ_1 and the resulting initial ensemble singular vector (IESV) leads to the maximum growth of the corresponding final ensemble singular vector (FESV). With eq. (7), we can find m pairs of initial and final ensemble SVs (IESV_i and FESV_i, for $i = 1$ to m) and IESV_1 indicates the fastest growing perturbation.

Also, note that the initial and final norms, $\mathbf{C}_I$ and $\mathbf{C}_F$ do not need to be the same, and that they can include, for example, a mask to project only on the region of interest. If the interest of the forecast is certain types of scales, one may consider applying any kind of low/high pass filters to the ensemble and construct the ESV based on these filtered perturbations.

Using eq. (7) [as with the adjoint sensitivity in eq. (3), Rabier et al., 1996] one can formulate the response to general questions such as: 'What change in the initial conditions in a certain region will have maximum impact on the vorticity of the final forecast in another region?' Given that these ensemble SVs will provide the directions of fast-growing errors, we propose to test the use of ESVs as additive inflation for EnKF assimilation.

2.2. LETKF and running in place

The LETKF (Hunt et al., 2007) belongs to the square root type of EnKFs and it updates the ensemble mean and perturbation according to the local information of the background (a short-range forecast) and observations. In the LETKF, optimal weights for the background ensemble perturbations are derived, so that this linear combination of the ensemble perturbations minimises the analysis error variance (in the local domain). With K background

ensemble members at time t_n, the analysis ensemble perturbation (deviations from the ensemble mean) at the analysis time t_n are computed as follows:

$$\mathbf{X}_n^a = \mathbf{X}_n^b \mathbf{W}_n^a \tag{8}$$

Here, $\mathbf{X}_n^b = [\delta \mathbf{x}_n^{b,1}] \ldots [\delta \mathbf{x}_n^{b,k}]$ is the matrix of the background perturbations whose columns are the vectors of ensemble perturbations with respect to the ensemble mean: that is, $\delta \mathbf{x}_n^{b,k} = \mathbf{x}_n^{b,k} - \overline{\mathbf{x}}_n^b$, where $\mathbf{x}_n^{b,k}$ is the kth background ensemble member and $\overline{\mathbf{x}}_n^b$ is the background ensemble mean. Similar definitions are applied to the analysis ensemble mean ($\overline{\mathbf{x}}_n^a$) and perturbations ($\mathbf{X}_n^a$). The analysis perturbation weight matrix $\mathbf{W}_n^a$ is computed by:

$$\mathbf{W}_n^a = \left[\mathbf{I} + \frac{1}{(K-1)} \mathbf{Y}_n^{bT} \mathbf{R}^{-1} \mathbf{Y}_n^b \right]^{-\frac{1}{2}} \tag{9}$$

Here, $\mathbf{Y}_n^b = \left[\delta \mathbf{y}_n^{b,1}\right] \ldots \left[\delta \mathbf{y}_n^{b,k}\right]$ is the matrix of the background ensemble perturbations in observation space where $\delta \mathbf{y}_n^{b,k} = h(\mathbf{x}_n^{b,k}) - \overline{h(\mathbf{x}_n^b)}$, superscript T stands for matrix transpose, $\mathbf{R}$ is the observation error covariance matrix and $h(\bullet)$ is the observation operator that converts a variable from model to observation space.

The analysis ensemble mean at time t_{n} is obtained from

$$\mathbf{x}_n^a = \mathbf{X}_n^b \overline{\mathbf{w}}_n^a + \overline{\mathbf{x}}_n^b, \tag{10}$$

where

$$\overline{\mathbf{w}}_n^a = \hat{\mathbf{P}}_n^a \mathbf{Y}_n^{bT} \mathbf{R}^{-1} \left(\mathbf{y}_n^o - \overline{\mathbf{y}}_n^b \right) \tag{11}$$

In eq. (11), $\mathbf{y}_n^o$ and $\overline{\mathbf{y}}_n^b = \overline{h(\mathbf{x}_h^b)}$ are the column vectors for the observations and the background ensemble mean in observation space, respectively. Equations (8)–(11) provide the basic formulas of the standard LETKF. The R-localisation method (Hunt et al., 2007; Greybush et al., 2011) is adopted, which increases observation errors as the distance between the observation and analysis grid increases. Multiplicative covariance inflation is applied on the background ensemble perturbations, as $\mathbf{X'}_n^b = (1 + \Delta)\mathbf{X}_n^b$ and Δ is the inflation factor. Additive inflation is applied by adding new perturbations (Q_n) onto the analysis ensemble perturbations. Through non-linear integration, the additive inflation can have impact on the background ensemble perturbations.

$$\mathbf{X'}_n^a = \mathbf{X}_n^a + \mathbf{Q}_n \tag{12}$$

Based on the framework of the standard LETKF, Kalnay and Yang (2010) proposed the running in place (RIP) method to accelerate the spin-up period of an EnKF when initialising the assimilation from a state far from the true dynamics (e.g. a cold start), or when the background error statistics suddenly change (e.g. a rapid regime change in the dynamics). Below, RIP is described briefly within the LETKF framework. Further details are presented in Kalnay and Yang (2010) and Yang and Kalnay (2012).

The RIP scheme has two steps: (1) the use of the no-cost smoother derived with the latest observational and dynamical information to adjust the ensemble at a time earlier than the current analysis time, and (2) a forward integration of these smoothed (improved) ensemble states to the current analysis time and assimilation of the same set of observations. In Kalnay and Yang (2010), these two steps are repeated iteratively for a window between the previous and current analysis times, (t_{n-1}, t_n). We note that with a dynamically complex model such as a regional weather model, the window for applying the no-cost smoother can be shorter than the analysis interval, given the non-linearity of the mesoscale dynamics (Yang et al., 2012).

In step (1), the no-cost smoother applies the LETKF weights derived at t_n to the analysis ensemble at the previous analysis time t_{n-1}, as indicated in eqs. (13) and (14) below. These weights carry the information about the later observations and dynamical uncertainties so that the mean and ensemble anomalies can be smoothed:

$$\overline{\mathbf{x}}_{n-1}^{a,i+1} = \overline{\mathbf{x}}_{n-1}^{b,i} + \mathbf{X}_{n-1}^{b,i} \overline{\mathbf{w}}_n^{a,i} \tag{13}$$

$$\mathbf{X}_{n-1}^{a,i+1} = \mathbf{X}_{n-1}^{b,i} \mathbf{W}_n^{a,i} \tag{14}$$

At the ith iteration, the weights ($\overline{\mathbf{w}}_n^{a,i}$ and $\mathbf{W}_n^{a,i}$) obtained during the LETKF analysis computation at t_n are applied to the mean and the perturbations of the model ensemble at t_{n-1} ($\overline{\mathbf{x}}_{n-1}^{b,i}$ and $\mathbf{X}_{n-1}^{b,i}$). Equations (13) and (14) start at $i = 0$, where $\overline{\mathbf{w}}_n^{a,0}$ and $\mathbf{W}_n^{a,0}$ are the weight coefficients from the standard LETKF (with the observation assimilated once) and $\overline{\mathbf{x}}_{n-1}^{b,0}$ and $\mathbf{X}_{n-1}^{b,0}$ are the mean and the perturbations of the final analysis ensemble derived at the previous analysis cycle at t_{n-1}.

In step (2), a forward integration of the ensemble states from t_{n-1} to t_n provides the new background ensemble ($\mathbf{x}_n^{b,i+1}$) for the next iteration. The LETKF computation is repeated to obtain the new analysis ensemble ($\mathbf{x}_n^{a,i+1}$) and weight coefficients ($\overline{\mathbf{w}}_n^{a,i+1}$ and $\mathbf{W}_n^{a,i+1}$). Following Kalnay and Yang (2010), the iteration procedure is repeated until the improvement of the observational increment in the new iteration is less than 5%:

With RIP, the accuracy of the mean state and ensemble-based background error covariance are both improved simultaneously to capture the underlying true dynamics, as represented by the observations. Also, the RIP method has been proposed to serve as a generalised outer-loop to deal with the non-linearity and non-Gaussianity issues that may lead to filter divergence (Yang and Kalnay, 2012; Yang et al., 2013).

3. QG channel model and experimental setup

ESVs are computed under the framework of LETKF implemented on a QG model (Rotunno and Bao, 1996). The dimensional values for the model parameters are the same as in Morss (1998). There are 64 grid points in the zonal direction, 33 grid points in the meridional direction, and a total of seven vertical levels, including the bottom, top and five internal levels. The non-dimensional time step is 0.1, about 30 min. The prognostic variables are potential vorticity (q) at the interior levels, and potential temperature (θ) at the bottom and top levels. The governing equations are:

$$\frac{\partial q}{\partial t}+\mathbf{v}\bullet\nabla q=(-\tau^{-1}+D)(q-q_{ref})$$
$$\frac{\partial \theta}{\partial t}+\mathbf{v}\bullet\nabla\theta=\begin{cases}-\Gamma\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\phi+(-\tau^{-1}+D)(\theta-\theta_{ref}) & \textit{at the bottom}\\ (-\tau^{-1}+D)(\theta-\theta_{ref}) & \textit{at the top levels}\end{cases} \tag{15}$$

q is defined through the stream-function ϕ, as $q=\beta y+\partial^2\phi/\partial z^2+\nabla^2\phi$. The velocity and temperature are obtained through the stream-function (SF): $(u,v,\theta)=(-\partial\phi/\partial y,\partial\phi/\partial x,\partial\phi/\partial z)$. In eq. (15), q_{ref} and θ_{ref} are the zonal mean reference state for potential vorticity (PV) and temperature, τ is the relaxation time, D denotes the fourth-order horizontal diffusion, and Γ controls the Ekman pumping at the bottom level. Further details about the mathematical formulation and the numerical schemes are described in Rotunno and Bao (1996). The forcing and dissipation included in the model are specified as in Snyder et al. (2003).

With a perfect model configuration and observation system simulation experiments (OSSEs), Yang et al. (2009) showed that the analysis derived from the QG-LETKF was more accurate than the analyses from the QG-3D-Var or from the QG-4D-Var with a short assimilation window (12 h), but was comparable with the QG-4D-Var analysis with a 24-h assimilation window, which requires more computational time.

When the model is initialised from a very inaccurate state (e.g. climatology), the QG-LETKF system needs a very long spin-up period to converge to its asymptotic level of performance. This is because the flow-dependent error covariance used in the LETKF system can effectively estimate the observation corrections only after the ensemble perturbations capture the structures of the growing errors. During the spin-up, the accuracy of analysis mean state is far from the asymptotic level of performance. Ensemble perturbations evolved upon such poor mean state cannot capture well the structures of the dynamically growing errors and thus the background error covariance based on these ensemble perturbations is very suboptimal in representing the flow-dependent uncertainties. As a result, the observations are less effectively assimilated with the poor background error covariance and thus the system takes time to correct the analysis accuracy. As shown in Kalnay and Yang (2010) and Section 5.3, such long EnKF's spin-up can be significantly accelerated with the RIP method so that the spin-up period becomes even shorter than those required for the variational-based methods.

We propose to apply additive inflation with ESVs within the LETKF and LETKF-RIP frameworks with 20 ensemble members. Following the OSSE setup in Yang et al. (2009), the data assimilation experiments in this study are performed for 150 d using a 12-h analysis cycle (i.e. 300 analysis cycles). There are 64 'rawinsonde' observations, randomly distributed in the model domain. The observations are vertical profiles of zonal and meridional wind components and temperature, generated by adding random Gaussian errors on the truth. The observation error covariance is constructed following Dey and Morone (1985), where the observation error is assumed to be uncorrelated between observations and between different variables. Only vertical correlations for the same variable are considered. The observation error is 0.8 m s^{-1} for the zonal wind, 0.5 m s^{-1} for the meridional wind and 0.8°C for temperature. Details of the QG assimilation setup can be found in Yang et al. (2009). The performance of this assimilation system is measured by the RMS analysis error, defined as the domain-averaged RMS difference of the model variables (PV and temperature) between the analysis and truth.

4. The ESVs in the QG model

4.1. Characteristics of the ESVs and their relationship with the background errors

The initial and final ESVs are derived within the QG-LETKF system. To initialise the experiment, the mean state of the ensemble is initialised from a 3D-Var solution (Morss, 1998) and the initial perturbations are sampled from the 3D-Var error covariance. To illustrate how quickly the leading FESV can project on the fast-growing error, Fig. 1 compares the leading ESV, ensemble perturbations from the first ensemble member and errors of the mean state. In the following, the background and analysis ensemble perturbations are referred to as BP and AP respectively. In Fig. 1, the left panel shows the leading IESV, the analysis errors and AP at the initial time ($t=1$), and the right panel shows the leading FESV, the background errors and BP 12-h later ($t=2$). Since the system is initialised from a 3D-Var analysis and the errors from the mean state are mixed with the growing (with dynamically stretching) and non-growing errors (isotropic-like), as shown in Fig. 1a. At such early time of assimilation, the

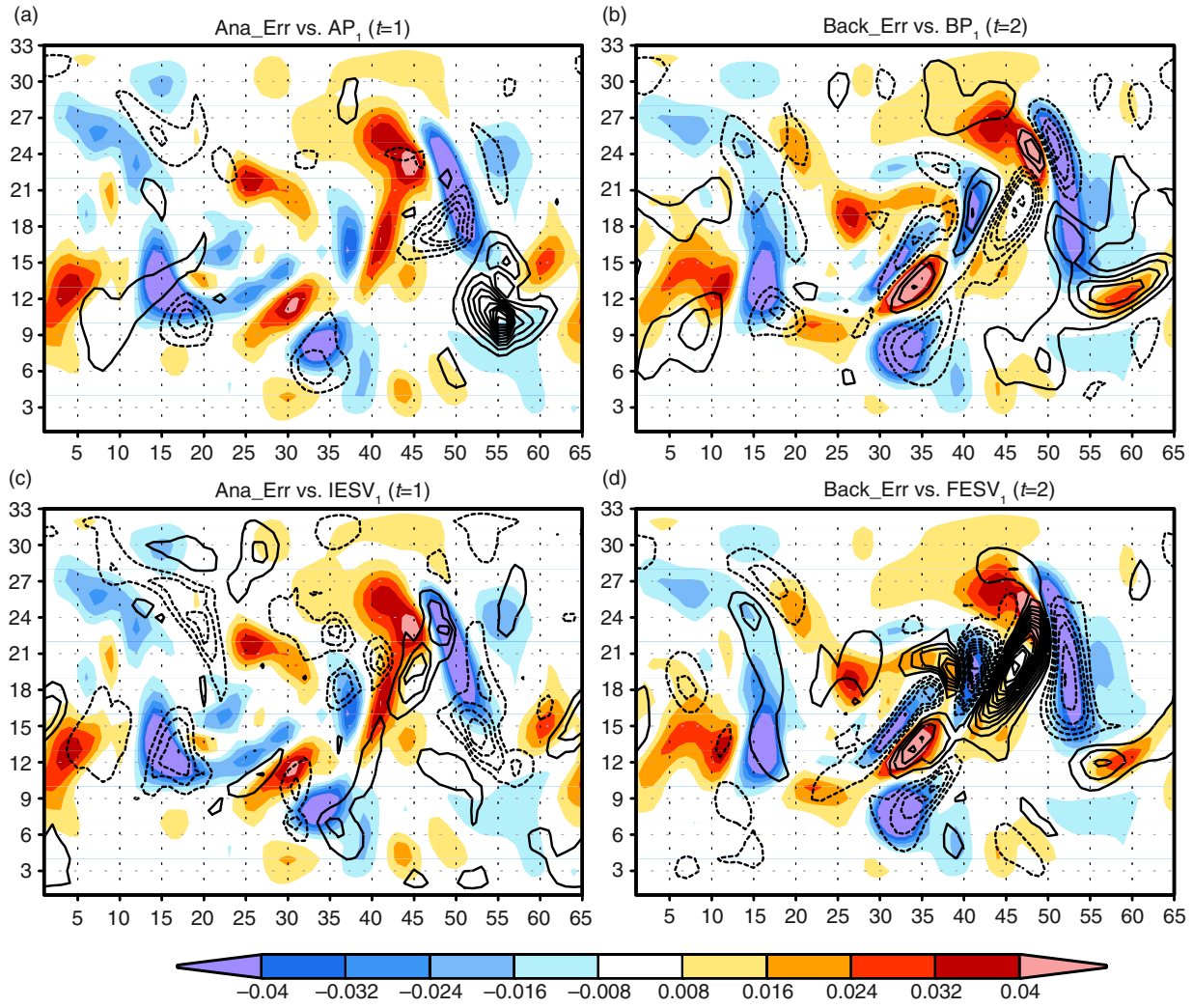

Fig. 1. (a) Analysis ensemble perturbation (contour) and analysis errors (colour shading) at $t=1$ and (b) and background ensemble perturbation (contour) and background errors (colour shading) at $t=2$. (c) and (d) are the same as (a) and (b), except that the contours are, $IESV_1$ and $FESV_1$, respectively. The contour interval is 0.004.

ensemble perturbations are still strongly dominated by the 3D-Var Gaussian structures (contours in Fig. 1a) and they are not yet able to represent the locations and structure of analysis error. In Fig. 1c, the analysis errors at $t=1$ are compared with the first initial SV ($IESV_1$), which being an IESVs is not similar to the analysis errors either. The right top and bottom figures show the effect of 12 h of dynamic evolution, with the colours being the background (forecast) error after 12 h in Fig. 1b and d. In Fig. 1b, the perturbation from the first background ensemble member has done a fair job in representing the true background errors. In contrast, in Fig. 1d, the fastest growing $FESV_1$ is representing exceedingly well the true forecast errors.

Although the ESVs are derived from the same set of ensemble perturbations, $FESV_1$ at this early time already projects very strongly on the background errors, identifying the locations and stretching directions of fast-growing errors (e.g. area near $x=52$, $y=18$ in Fig. 1d). This suggests that even at the early time of model forecast and assimilation, when ensemble is starting to develop the structures related to growing errors, the computation of the leading ESV can help to efficiently estimate the shapes of the fast-growing errors. The fastest growing modes eventually correspond to the uncertainty pattern that determines the background errors. In addition to the leading ESV, $FESV_2$ to $FESV_5$, identified as growing ESVs, also project strongly on different parts of the dominant errors, as highlighted by the green boxes in Fig. 2. Therefore, these large errors with growing property can be derived by constructing the ESVs. In principle, perturbations containing very little energy, perturbations projecting on poorly observable directions or perturbations corresponding to the phenomena

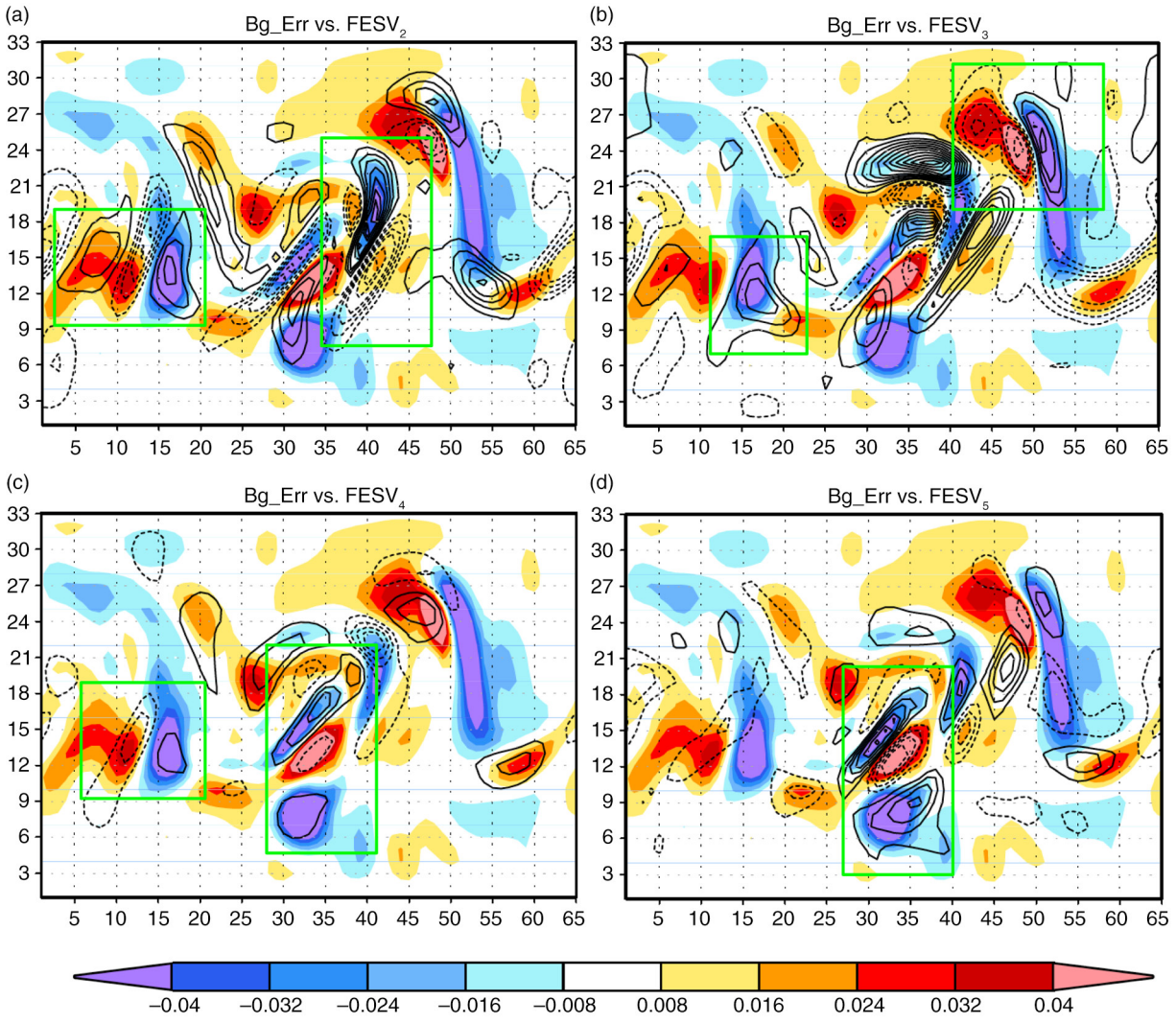

Fig. 2. Background errors (colour shading) and (a) $FESV_2$, (b) $FESV_3$, (c) $FESV_4$ and (d) $FESV_5$ at $t=2$. Green boxes indicate areas where FESV projects well on the background errors. The contour interval is 0.004.

'not-of-interest' may also project on these ESVs. As a consequence, several ESVs with growing modes may be needed to project well on different parts of the background error.

Even well after the LETKF has spun up, the leading FESVs can still better estimate the fast-growing errors based on the well-developed ensemble, as will be discussed later. According to the temporal mean singular values, there are on the average 10 growing modes from all the available ESVs, as shown in Fig. 3, in which the growth rates are computed based on PV (in red) and SF (in blue) norms. In comparison, the ensemble perturbations exhibit modes growing moderately. Figure 3 also indicates that when using 10 vectors for perturbations, the first 10 FESVs will be more effective to capture the growing errors than using 10 BPs.

To investigate the relationship between FESVs and background errors, we compute the local angle between the background error and the subspace of FESVs or BPs. This also indicates the extent to which the background error lies in the subspace spanned by these vectors (Corazza et al., 2003). As shown in eq. (16), the computation is done by sequentially removing the projection from each vector on background errors. In eq. (16), $\langle \mathbf{v}_i, \mathbf{e}_0 \rangle$ denotes the inner product between the vector of background error, $\mathbf{e}_0$, and the ith ensemble vector, $\mathbf{v}_i$. $\mathbf{e}'$ is the remaining part unexplained by the vectors and the calculation is initialised from $\mathbf{e}' = \mathbf{e}_0$. $\cos^2(\theta)$ denotes the percentage of the explained variance.

$$\begin{aligned} \mathbf{e}' &= \mathbf{e}' - \langle \mathbf{e}_0, \mathbf{v}_i \rangle \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|^2} \, i = 1, \cdots, K \\ \cos^2(\theta) &= 1 - \|\mathbf{e}'\|^2 \Big/ \|\mathbf{e}_0\|^2 \end{aligned} \tag{16}$$

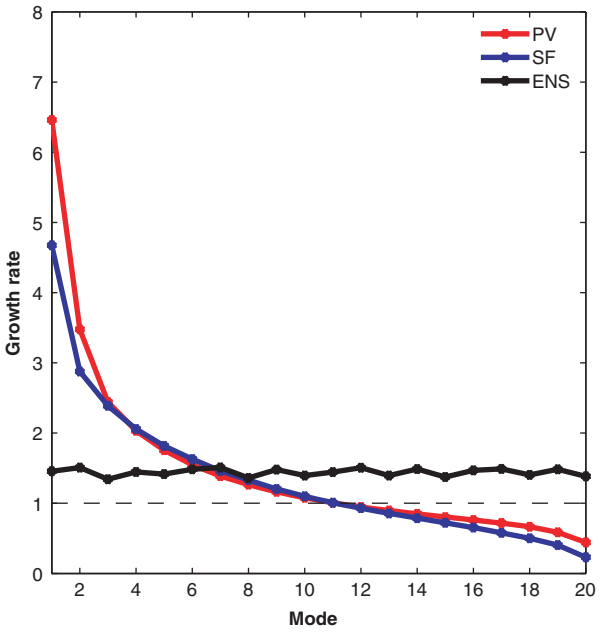

Fig. 3. Time mean growth rates of the 20 ESVs based on the potential vorticity and stream-function norms and ensemble perturbations from 300 analysis cycles. An ESV with a growth rate larger than one indicates a growing mode.

The computation of eq. (16) is done sequentially with a 9×9 local box centred at each analysis grid point and K is the number of the total vectors. For $K > 1$, the local vectors $(\mathbf{V}_1, \mathbf{V}_2, \ldots \mathbf{V}_k)$ are orthogonalised before the computation to avoid double counting. Figure 4 shows the time series of the mean of the local explained variance from using the leading and first 10 FESVs, in comparison with the one derived from using the same number of BP vectors. To avoid over-representation of certain ensemble members, the BP vectors are randomly chosen from the 20 ensemble perturbations. As shown in Fig. 4a, the FESV_1 has larger explained variance most of the times than using one BP. Even though, FESV_1 occasionally, has less explained variance than BP (e.g. at the 10th analysis cycle), other FESVs help capture the growing errors. For example, the explained variance with the first 10 FESVs at the 10th analysis cycle becomes larger than the one derived with 10 BPs (Fig. 4b). With 10 vectors, it is clear that the background error is better confined with the subspace spanned by the 10 FESVs than the space by 10 BPs. In other words, the subspace spanned by the FESVs with growing modes always represents better the background errors than the BPs. It would be more efficient to make corrections using a subspace that better encompasses the background errors. In Section 5, these growing ESVs will be used to enhance the structures of ensemble instead of uniformly enlarging the amplitude of BPs like the multiplicative inflation.

The amount of explained variance is strongly related to the amplitude of the errors, which is also associated with the strength of dynamical instability. The background error is more strongly confined in the subspace of FESVs when the dynamical instability becomes stronger. Therefore, it is expected that using ESVs as the additive inflation may have advantage in regions with large growing errors.

4.2. *ESVs with different norms*

In Section 4.1, the IESVs and FESVs were derived with the L_2 norm, which can also be defined as the PV norm. The ESVs can be derived with different choice of initial and final norms, as indicated in eq. (7). In the following, we demonstrate that the structures of the ESVs are sensitive to the choice of norms by comparing the ESVs derived from the PV and SF norms. We saw in Fig. 3 that the growth rate of the fastest growing SVs is larger for the PV norm than for the SF norm.

It is well known that, in general, initial SVs grow very fast and are very sensitive to the choice of norm. Final SVs, by contrast, evolve to become close to the local leading LV therefore grow more slowly (like LV) and are much less sensitive to the choice of norm (e.g. Norwood et al., 2013). This is also observed among ESVs. Figure 5 shows the ESV_1 for potential temperature at the bottom level derived from the PV and SF norms. FESV_1 from two norms are generally similarities over the areas of dominant structures, while the PV- and SF-based IESV_1 are, as expected for initial SVs, much less similar. Also, the IESV_1 with the SF norm has some structures that are less stretched than the IESV_1 with the PV norm. The similarity between the PV- and SF-based FESV_1 suggests that they are dominated by the same instabilities associated with the local leading LVs, and thus indicate the growing errors at the same areas.

Results from Fig. 5 also reflect the fact that SVs are sensitive to the choice of norms and that it is straightforward to change the choice of norms with the ensemble-based method. The PV-based FESVs show larger projections on the background errors (figure not shown), indicating a closer relationship to the background errors. This is expected since the PV-based norm uses the perturbations of the prognostic variables of the model. On the average, the PV-based ESV_1 has a larger growth rate than the SF-based ESV_1, as shown in their singular values (Fig. 3). Therefore, we expect that the PV-based ESV_1 grows faster and better represent the fast growing errors. The efficiency of the data assimilation scheme depends on how well the observations can correct the growing errors

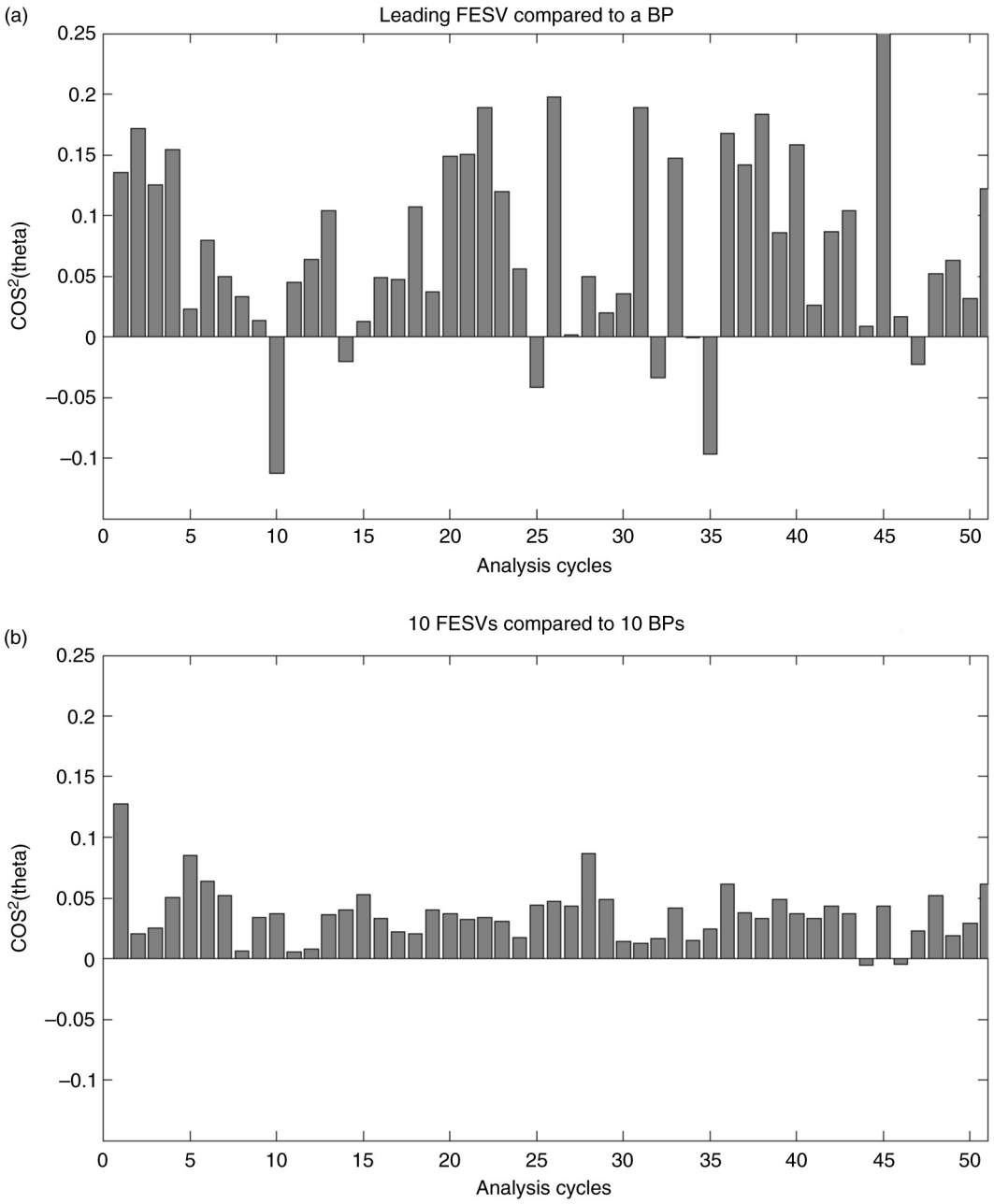

Fig. 4. Time series of the improvement of the projection of the background errors on the FESVs compared with the BPs. (a) leading FESV compared with a BP and (b) 10 FESVs compared with 10 BPs.

during the forecast-analysis cycles. Although the observing network used in these experiments is dense enough to constrain the large-scale error growth, the uncorrected errors (analysis errors) are dominated with flow-dependent structures, giving the basis for the fast growing errors during the forecast step. Given our purpose of capturing the fast growing errors in the prognostic variables of the QG model, it is reasonable to apply the PV-based ESVs to enhance the ensemble perturbations during the EnKF assimilation.

In the following, the ESVs are incorporated in the background error covariance used in the assimilation to better capture the structures of flow-dependent fast growing errors. For this purpose, the set of ESVs are used as the flow-dependent additive covariance inflation for the QG-LETKF and QG-LETKF-RIP frameworks.

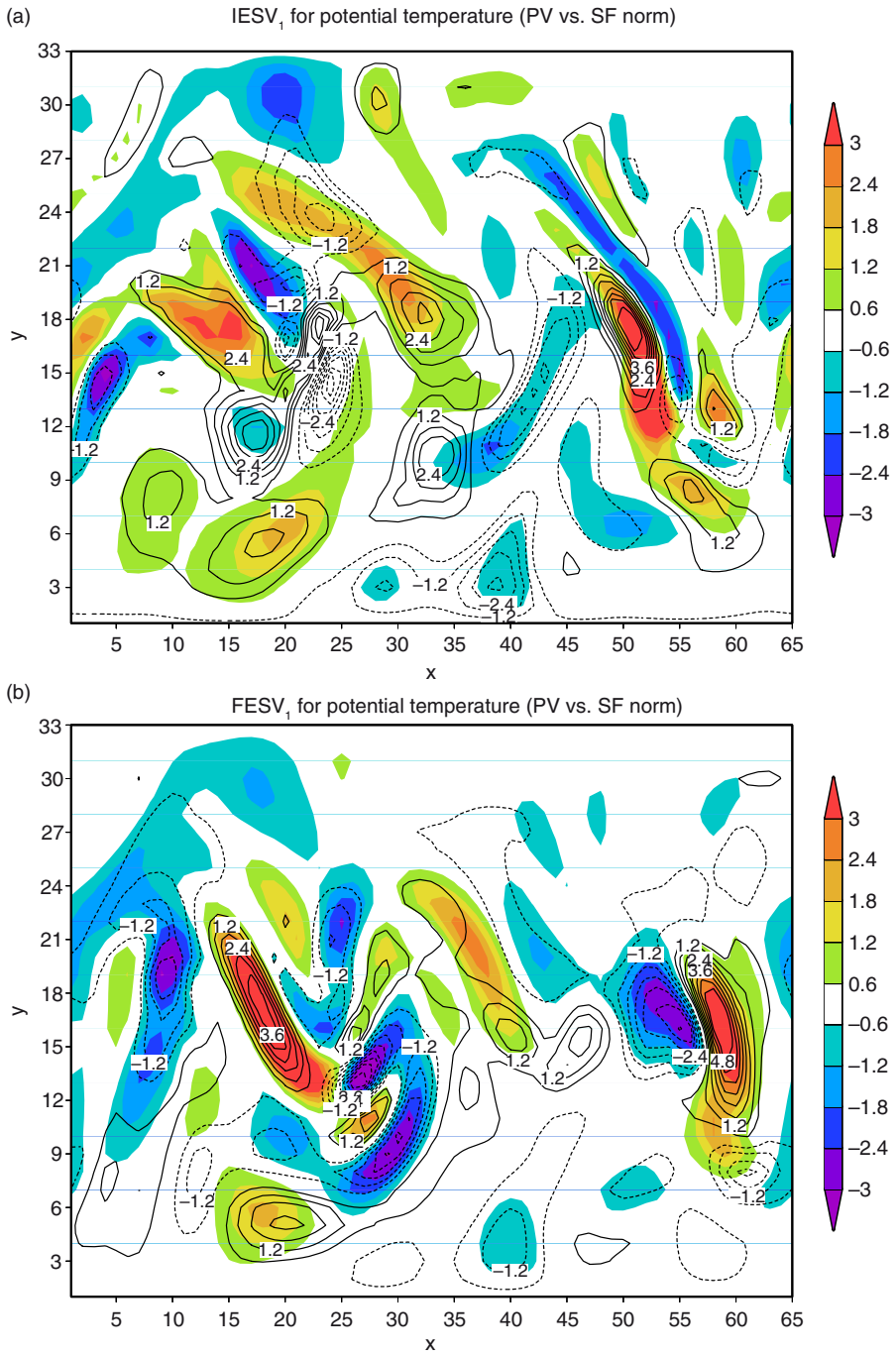

Fig. 5. (a) $IESV_1$ and (b) $FESV_1$ at $t = 50$, derived with the potential vorticity norm (colour shading) and stream-function (contours) norms. For plotting purpose, variables are rescaled and the contour interval is 0.6.

5. Using ESVs as additive inflation in the EnKF framework

5.1. *Results with the regular LETKF framework*

In the original LETKF framework, the additive covariance inflation is performed by adding a set of random perturbations onto the analysis ensemble. The random perturbations are white Gaussian noise. Through the non-linear forecast, the additive perturbations are expected to help the ensemble better capture the growing directions that were originally missed in subspace spanned by the ensemble perturbations (Corazza et al., 2003). To show the impact of

using additive covariance inflation, four experiments are conducted within the LETKF framework. The control experiment (CNTL) uses only the multiplicative covariance inflation and the amount of inflation is a function of the vertical levels to consider the errors distribution (Yang et al., 2009). The multiplication inflation ranges from 8.4% at the bottom level to 13.8% at the top levels. In addition to the multiplicative inflation used in CNTL, the other experiments also use additive inflation, including random perturbations, all IESVs, 10 IESVs and 10 FESVs. In the following, these experiments are denoted as RDM, IESVall, IESV10 and FESV10. The amplitude of the additive perturbations is 2% of the analysis ensemble perturbations. In IESV10, the positive–negative pairs of the first 10 IESVs, recognised as the growing modes in Fig. 3, are used to form the $\mathbf{Q}_n$ in eq. (12). Although ESVs have the structures associated with the errors, we observe that the signs of the local ESVs are not always consistent with the sign of the errors. In FESV10, the additive inflation is then composed of the first 10 FESVs and applied on the background ensemble. Besides, in all the ESV-associated experiments, the order of ESVs is randomised before they are applied as the additive inflation so that each ensemble member has the same chance to use the leading IESV, instead of having only the first ensemble member with the privilege of boosting its perturbation into the growing errors.

Figure 6 is a schematic plot of how IESVs are applied as the additive inflation for the background error covariance used in the LETKF framework. In the IESVall and IESV10 experiments, IESVs are added on the analysis ensemble. The reason that we apply IESVs on the analysis ensemble instead of adding FESVs on the background ensemble is because that by adding perturbations on the analysis ensemble, the IESVs (which grow faster than the FESVs) can be integrated with the fully non-linear model, which allows adding a non-linear component in the evolution of IESVs. However, we should note that these IESVs are derived based on the previous APs at the previous analysis time and BPs at the current analysis time (eq. 7) because at t_i the IESVs are not yet available. Therefore if we want to use perturbations that will grow fast and become non-linear, we need to perturb with the IESVs derived from the previous window. As shown in Fig. 6, the analysis ensemble at time t_i is perturbed by the IESVs derived based on the ensemble evolutions between time t_{i-12} and t_i. At the next analysis time (t_{i+12}), the new background ensemble, denoted as $\mathbf{X}'_b$ in Fig. 6, will be perturbed through the non-linear evolution of IESV. Given that it is the 'current' analysis accuracy that we want to improve, it is not feasible to apply the current IESVs as the additive inflation since the acquisition of the current IESV requires the current analysis ensemble and the following 12-h ensemble forecasts. Despite the shift of the reference state 12-h back, we find no significant difference in its application as additive inflation because the structure of IESV does not change rapidly in 1 d for a QG model. We found that the structures of FESVs are similar to the structures of IESVs after a 12-h integration starting with either the correct reference state or the reference state shifted 12 h later, especially in the areas with large amplitudes where fast growing instabilities take place.

Figure 7 shows the time series of RMS of the total analysis errors and Table 1 lists out the mean RMS analysis errors from all experiments. We note that, all LETKF experiments with 20 members converge to an error level much lower than the 3D-Var analysis error (grey line in Fig. 7a). However, the RMS errors occasionally spike in the CNTL analysis, which uses just multiplicative inflation. With additive inflation, RDM improves the CNTL performance and successfully reduces the amplitude of the large errors. This confirms that by applying random perturbations as the additive inflation, we gain better chance of capturing

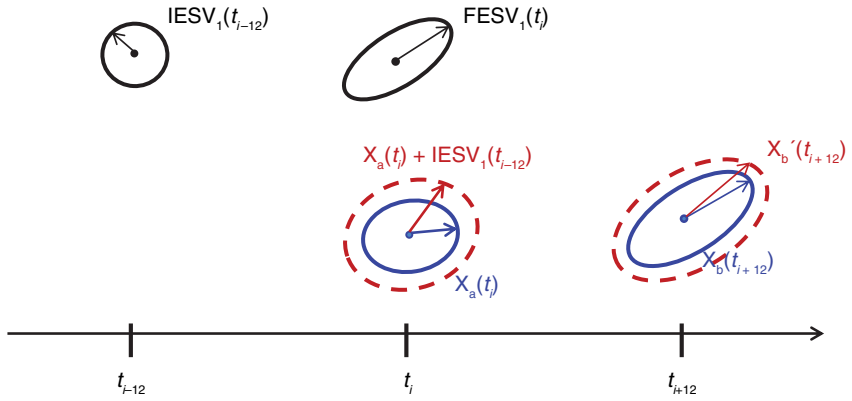

Fig. 6. A schematic plot for applying IESVs as the additive inflation in the LETKF assimilation framework.

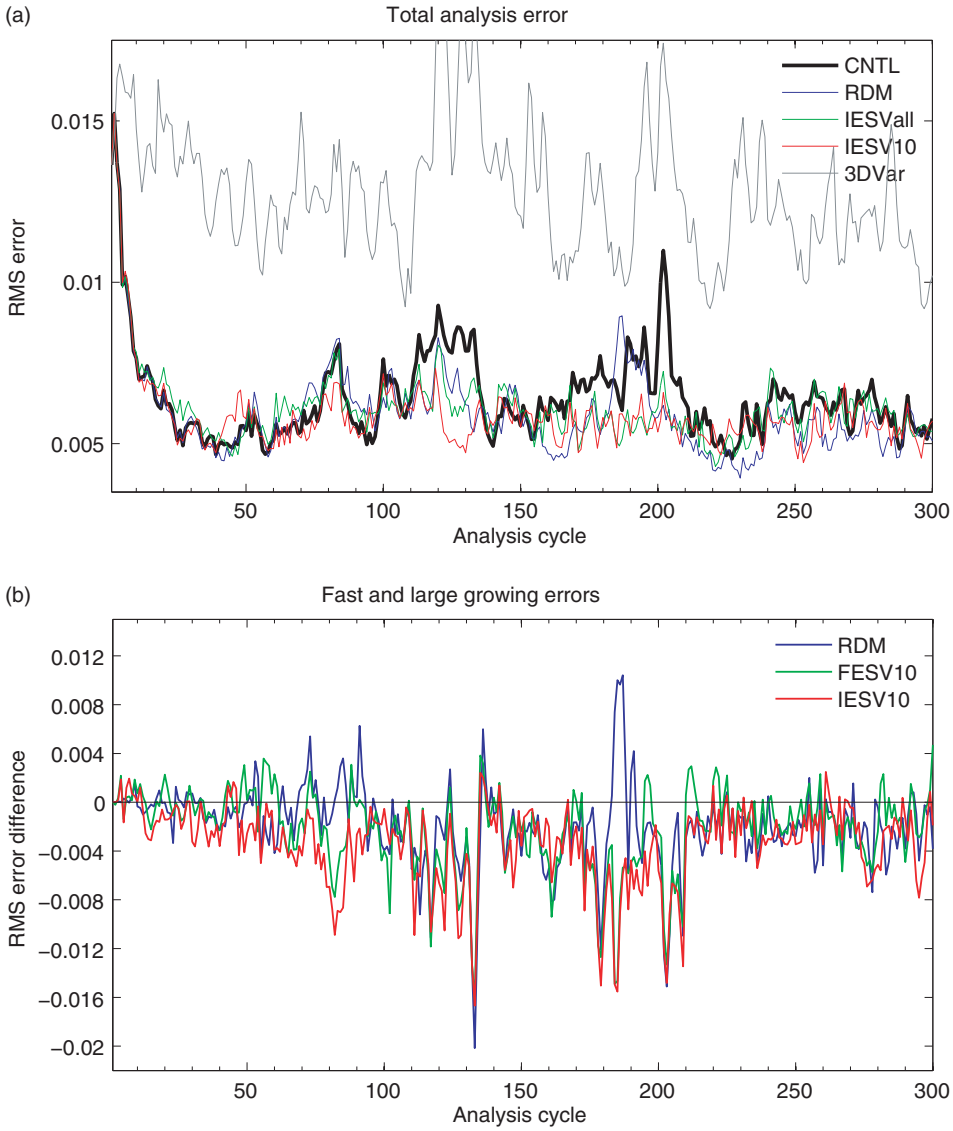

Fig. 7. Time series of the RMS of (a) the total analysis error and (b) analysis error with the mask of large errors. A mask of large errors is defined as the grid points whose analysis error from the CNTL experiment is two standard deviation larger than the domain-average analysis error.

the subspace of the growing errors and can further improve the analysis accuracy. Particularly, the RMS errors of IESV10 and IESVall are both even smaller than the RDM when the CNTL RMS errors are large, for example. the 110th to 140th and 180th to 202th analysis cycles. However, using the first 10 IESVs as additive inflation has a better performance than the one using all IESVs. This suggests that more effective corrections are obtained when using structures associated with dominant errors to enhance the background error covariance.

To emphasise the catastrophic error growth characterised by rapid error growth and large error amplitudes, we define a mask according to the CNTL analysis and forecast errors. The mask is defined as the model grids that have errors grow rapidly during the first 12-h forecast ($\ln(\|e_1\|/\|e_0\|) > 0.5$, where e_0 and e_1 are the analysis error and 12-h forecast errors, respectively) and have 48-h error amplitude larger than two standard deviations of the domain mean 48-h forecast error. Based on this mask, Fig. 7b shows the RMS analysis error difference between the experiments with additive inflation and the CNTL. Negative and positive differences indicate improvement and degradation, respectively. As shown in Fig. 7b, experiments with additive inflation are all very successful for reducing such large errors most of the time and IESV10 generally has smaller errors than RDM does, as also shown in Table 1.

Table 1. Time mean analysis and background errors of all experiments

$\times 10^{-2}$	CNTL	RDM	IESVall	IESV10	NGIESV10	FESV10	IESV10_NRO
RMSE (analysis/background)	6.4 (7.5)	5.8 (6.7)	5.9 (6.8)	5.6 (6.4)	6.3 (7.2)	5.9 (6.9)	6.5 (7.4)
RMSE_Large (analysis/background)	13.0 (18.3)	10.8 (14.8)	9.7 (13.4)	9.1 (12.2)	9.9 (13.8)	10.3 (14.4)	10.7 (14.8)

With IESV10, the improvement is almost 30%, compared with a 20% improvement obtained with RDM. We should note that, RDM, occasionally, has errors larger than CNTL (e.g. at the 187th analysis cycles), indicating that random perturbations fails to represent these large errors at these times. In contrast, IESV10 successfully reduces large growing errors at these analysis times. All the ESV-incorporated experiments show improvement over RDM for the large growing errors but using the first 10 IESVs as the additive inflation is most effective (see Table 1). However, when 10 non-growing IESVs (NGIESV10) are used as additive inflation, the performance is worse than IESVall. This again confirms that the structures used for additive inflation can have a large impact on improving the LETKF performance and the effect will be limited when augmenting the background error covariance with structures less relevant to errors. Furthermore, the advantage of IESV10 can be attributed to improvement in the background ensemble for representing the errors as well as to the orthogonality in the ESVs (Annan, 2003). The improvement in the analysis also affects the subsequent accuracy of the 12-h forecast, that is, the background state. IESV10, which constrains better the growing modes, provides the most improvement for the background state and shows the best performance among all analysis systems shown in Table 1.

To illustrate the ability of IESV10 in removing growing errors, Fig. 8 compares the background error and analysis increment at four consecutive analysis times derived from CNTL, RDM and IESV10. Because of the use of a flow-dependent background error covariance, the analysis increments derived from the LETKF assimilation are expected to exhibit structures related to the background errors. Nevertheless, there are still some areas that have almost no analysis increments (corrections) on the background errors (e.g. the box area in Fig. 8a). Without removing the background errors completely, these errors in the box area reside, grow and propagate with the underlying flow (Fig. 8a–d). The LETKF system with multiplicative inflation is able to provide the analysis corrections related to these growing errors later at $t=81$, but the amplitude is not large enough to remove these large errors. Comparing Fig. 8a and e, the analysis increments are generally similar, and the one derived from RDM did not particularly show an advantage of correcting the background errors at this time, indicating that there is no significant difference in the ensemble-based background error covariance between CNTL and RDM. With IESV10, the analysis increment derived at $t=79$ shows corrections in the box area and thus have much smaller background errors at $t=80$ than the other two experiments. With the proper corrections at $t=81$, the background errors are further removed at $t=82$, while the box area of CNTL and RDM is still dominated by large growing errors. The example in Fig. 8 suggests that during an event of catastrophic error growth, using growing IESVs as the additive inflation can better capture the subspace of growing errors than the random perturbations and thus the observations can provide effective correction, resulting in a more accurate analysis.

Removing the growing errors with large amplitudes is important to reduce the forecast errors at long forecast times. Figure 9 shows the time mean RMS errors for the forecasts initialised from the analysis means of CNTL, RDM and IESV10. Figure 9 is computed based on the mask of large errors with fast growth rate used in Fig. 7b to emphasise the catastrophic errors. As shown in Fig. 9, the result with the CNTL forecast not only has the largest errors but also the slope of the error growth increases after the 12-h forecast time, indicating that these incompletely removed growing errors keep amplifying through nonlinear dynamics. With additive inflation, not only the performance is improved with a smaller RMS error, the error growth rate is also decreased. This again suggests that the additive inflation we applied helps to project more efficiently on the subspace of the dominating errors so that the analysis increments can effectively correct them and alleviate growth of large errors. Although the random perturbations have already done a generally good job in reducing the forecast errors, the IESVs provides better corrections for growing errors, especially during the first 12-h forecast and thus continuously improve both the accuracy of the initial and later forecast state. As indicated by the red dashed lines on Fig. 9, the error growth (the slope of the RMS error) during the first 12 forecast hour is the smallest with IESV10. We note that the advantage of using additive inflation can be accumulated through the analysis-forecast cycles and therefore it is difficult to quantify how long the effect of additive inflation persists.

Compared with IESV10 that perturbs the analysis ensemble, FESV10 directly perturbs the background ensemble with perturbations that grow slower because they are close to converging to the LVs. Result from FESV10 shows comparable mean analysis accuracy with the RDM

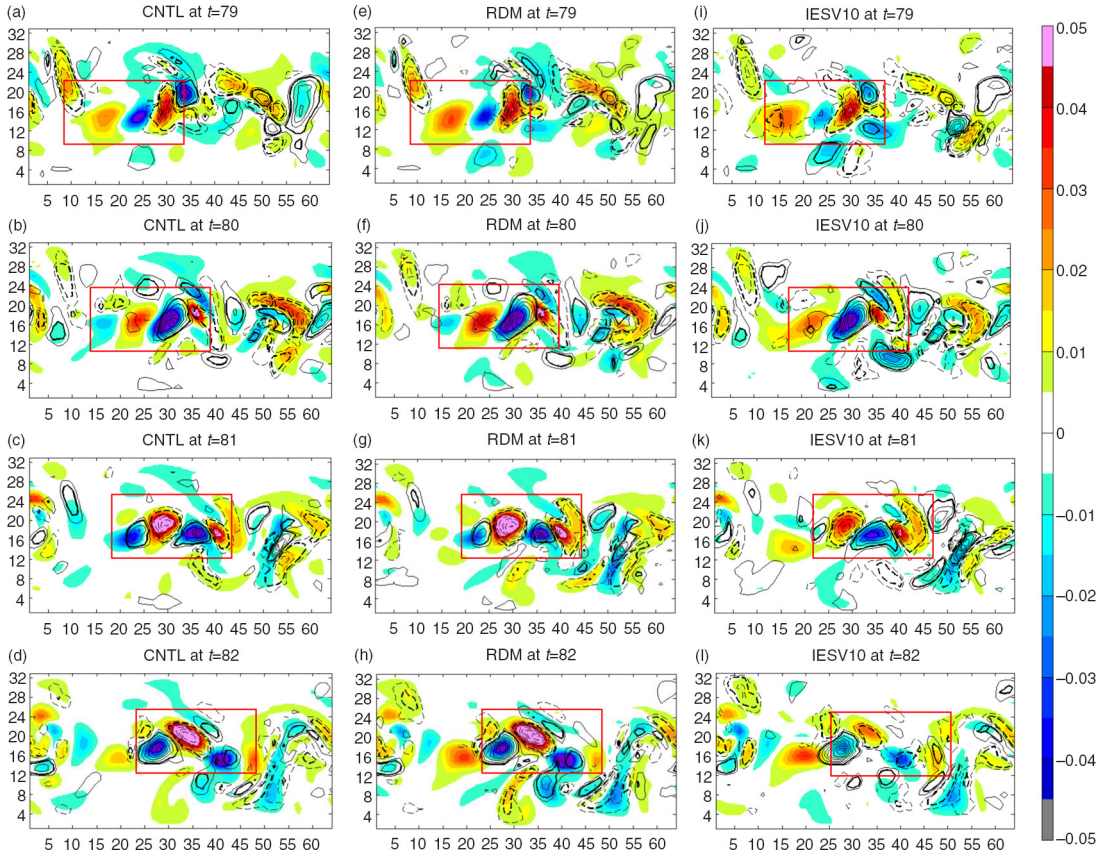

Fig. 8. (a)–(d) Background error (shading) and analysis increment (contour, black: positive, grey: negative) of potential temperature at the bottom level from the CNTL experiment from $t = 79$ to $t = 82$. Figure 8(e)–(h) are the same as Fig. 8(a)–(d) except that they are derived from the RDM experiment. Figure 8(i)–(k) are the same as Fig. 8(a)–(d) except that they are derived from the IESV10 experiment. The contour interval is 0.005, and the solid and dashed thick lines denote iso-lines of 0.005 and −0.005, respectively.

analysis and not as good as IESV10 (Table 1). However, we should note that using the first 10 FESVs is still very effective in correcting errors during the periods with large errors, as shown in Fig. 7b. The difference between IESV10 and FESV10 suggests that the space spanned by the background ensemble can be further perturbed by the growth of the non-linear component, giving more chances to capture the growing errors.

We also note that when applying ESVs as the additive inflation, it is essential to randomise the order of ESVs so that the chance for all the ensemble members to be perturbed by the fast growing modes is the same. Without such procedure, there is limited improvement from using ESVs, as indicated by the IESV10_NRO experiment in Table 1.

5.2. *Spectral analysis*

To further understand what have been corrected with the ESV-based additive inflation, a spectral analysis is performed for the analysis increments, forecast and analysis errors. As mentioned in Yang et al. (2009) and shown in Fig. 10a, forecast and analysis errors in this model are characterised by large- to mid-scale structures (global wavenumber smaller than 50). With the observing network used in this study, the corrections are mainly characterised by structures with wavenumber smaller than 15, corresponding to the dominant structures of the errors, and are effective enough to constrain the growing errors. By halving the number of observations, the error growth for wavenumber smaller than 10 almost doubles. Also, in Fig. 10a, additional noises with small amplitude in the analysis errors are introduced during the assimilation step, but they are damped after the model integration.

Compared with CNTL and RDM, forecast and analysis errors of IESV10 show smallest amplitudes at all scales (Fig. 10c and d), and such advantage is particularly evident at large- to mid-scales. As shown in Fig. 10b, the analysis increment derived from IESV10 is characterised by large- to mid-scale structure (wavenumber smaller than 20), corresponding to the scales of the growing part of the errors shown in Fig. 10a. This is another evidence that the

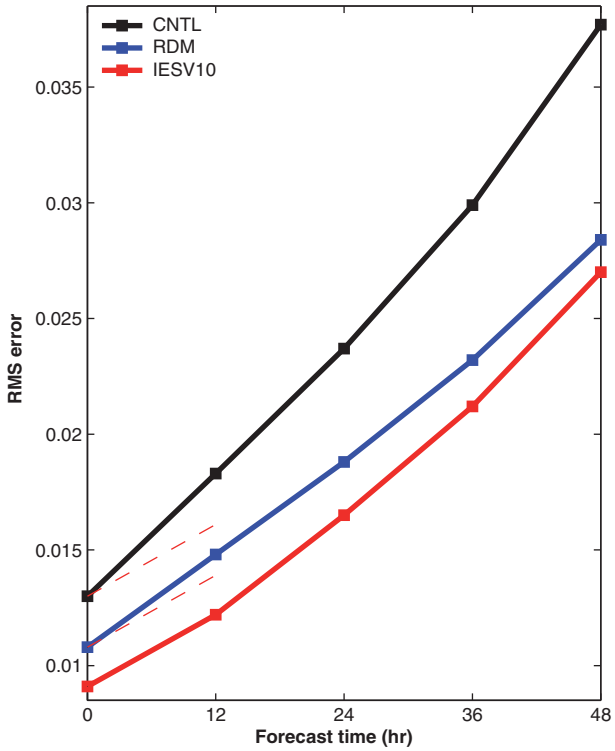

Fig. 9. Time mean RMS analysis/forecast errors at different forecast lead times with the mask of large errors. The red dashed line is the error of IESV10.

ESV-based additive inflation is more effective in correcting the growing errors than the random perturbations.

5.3. *Results with the LETKF-RIP framework*

In realistic application of the EnKF, the ensemble is often initialised from a condition far from the nature, for example, a model state without TCs or strong convection for regional assimilation and prediction. Therefore, a spin-up period is required for building up the flow-dependent characteristic of the ensemble perturbations before the EnKF system can achieve its asymptotic level of performance. The LETKF-RIP method (Kalnay and Yang, 2010, Section 2.2) is proposed to catch up the true dynamics by repeatedly using observations and iteratively improving the non-linear evolution of the ensemble. During RIP iterations, the smoothed analysis ensemble at the previous analysis time step is obtained with a 'no-cost smoother' and the accuracy of the smoothed mean state is improved due to observation information from a later time (Kalnay et al., 2007). It is expected that an improved mean state at the previous analysis time can lead to a dynamical evolution closer to the truth. Therefore, the ensemble perturbations, re-evolving upon this improved mean state, can better represent the flow-dependent dynamic uncertainties, that is, an improved background error covariance. In the LETKF-RIP system presented in Kalnay and Yang (2010), the smoothed analysis ensemble is perturbed by a small amount of random perturbations to avoid having the smoothed analysis evolve into the same current analysis ensemble (see Appendix A in Kalnay and Yang, 2010). Through perturbing the smoothed analysis ensemble and non-linear integration, the ensemble space can be perturbed to better capture the growing errors. Instead of using random Gaussian perturbations as in Kalnay and Yang (2010), we propose to further improve the flow-dependent characteristics of the ensemble perturbations with the first 10 ESVs so that the observations can be more effectively used and the spin-up of the LETKF system can be further shortened.

For the QG experiments using the LETKF-RIP method, the mean of the ensemble is initialised from the climatology state to highlight the issue of spin-up problem and the advantage of using a proper background error covariance. In contrast with the standard LETKF method, the additive inflation will be used to perturb the smoothed analysis during the RIP iterations, as the step (1) in Section 2.2. In the original LETKF-RIP method, random Gaussian perturbations with amplitude 20% of the variance of the analysis ensemble perturbations are added on the smoothed analysis ensemble.

In the following, results from three experiments are discussed: (1) the standard LETKF, (2) the LETKF-RIP with the random perturbations as the additive inflation for the smoothed analysis and (3) same as (2) except the additive perturbations are based on the first 10 IESVs. These experiments are referred to as CNTL-noRIP, RIP-RDM and RIP-IESV, respectively.

In CNTL-noRIP, the amount of multiplicative inflation is much larger than the ones used in RIP-RDM and RIP-IESV to avoid filter divergence. Nevertheless, the standard LETKF system still requires a long period to converge to its asymptotic level of performance, as indicated by the black line in Fig. 10. This is because that the initial ensemble is too poor, resulting in a poor background error covariance for assimilating observations and thus the observations cannot effectively correct the model state. As demonstrated in Kalnay and Yang (2010), RIP-RDM (the blue line in Fig. 11) successfully accelerates the spin-up period by simultaneously improving the accuracy of the mean state and the flow-dependent properties carried by the ensemble perturbations. RIP-IESV (the red line in Fig. 11) further improves the analysis accuracy during the spin-up period. In other words, using IESVs as the additive perturbations allows the LETKF-RIP scheme uses observations even more effectively and further accelerates the spin-up of the LETKF system. We note that after spin-up, both the RIP-related experiments converge to an accuracy

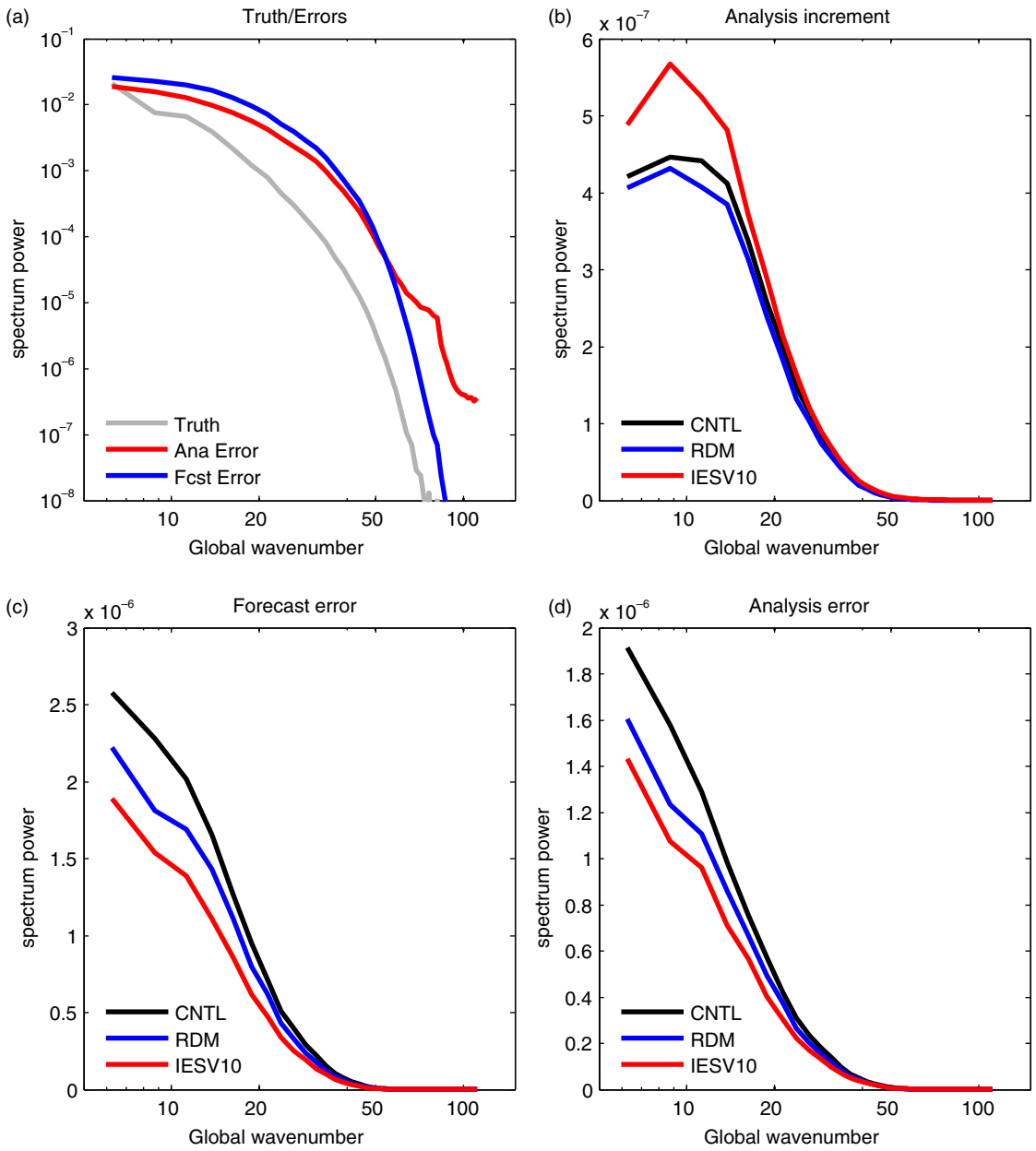

Fig. 10. Spectrum power of (a) truth, forecast and analysis errors of CNTL, (b) analysis increments from CNTL, RDM and IESV10 (c) same as (b) except for the forecast error and (d) same as (b) except for the analysis error. The amplitude of the spectrum power of analysis and forecast errors is multiplied in Fig. 9a by a constant factor, 10^4, for plotting purposes.

level of RMSE = 0.0044, which is more accurate than the one obtained with CNTL-noRIP (RMSE = 0.0053). The CNTL_noRIP has a larger RMSE is mainly because the use of a larger inflation so that the LETKF can converge; such larger inflation should be adaptively tuned after the system has spun up.

Figure 12 shows an example of the background error (colour) and analysis increment (contour) of the potential temperature at the bottom level at $t = 26$. In Fig. 12a, we changed the scale of the CNTL-noRIP error and increment by a constant factor of 0.5 to have comparable ranges as those of the RIP-related experiments. At this early time of the assimilation experiment, the corrections with large amplitudes are not collocated well with large background errors (e.g. area near $x = 47$, $y = 24$) and thus the fast growing errors are not corrected effectively yet with the available observations. At this time, large errors appear at the areas with strong gradient of the potential temperature

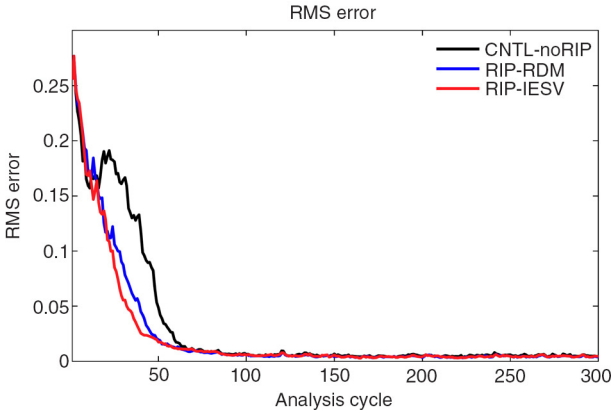

Fig. 11. Time series of the RMS analysis error of the potential temperature at the bottom level.

(figure not shown). Compared with Fig. 12a, the analysis corrections from the RIP-related experiments are smaller in the interior region where the observations locate and correspond well to the shapes and locations of the large background errors. However, at some areas (e.g. $x=23$, $y=20$), the analysis corrections do not correspond well with the background errors and thus the errors cannot be effectively removed with available observations. Compared with RIP-RDM, RIP-IESV is able to better correct the large interior errors, giving better correspondence between corrections and errors. This again confirms that by applying the fast growing errors as the additive perturbations, the background ensemble can more effectively capture the subspace of the growing errors so that the observations can be better used for analysis correction. We also note that, in RIP-IESV, the background errors near the northern and southern boundaries are smaller than those shown in RIP-RDM. This indicates that the error correlations away from observations are now more reliable and thus the corrections at far distance are more effective.

6. Summary and conclusion

In this study, we propose to use ESVs as additive covariance inflation. The derivation of ESVs was originally proposed by Enomoto et al. (2006, 2015) to use a set of ensemble forecasts, involving initial and final perturbations, to find the SV-like vectors that will have the maximum perturbation growth, given a chosen norm and optimisation period. Here, the ESVs are used to help the ensemble perturbations better capture the subspace of the growing errors, compared with the random perturbations that aim to capture the missing or sub-growing directions randomly.

The derivations of ESV and assimilation experiments are carried out within the QG-LETKF framework. The derivations of ESVs involve the analysis perturbations

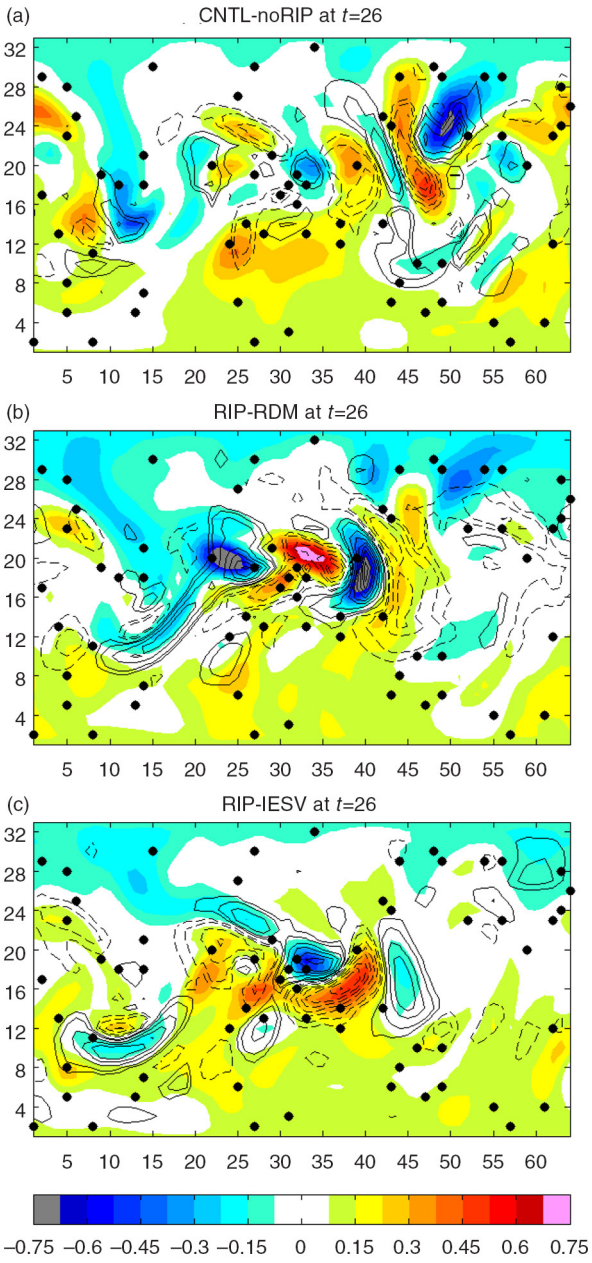

Fig. 12. Analysis increment (contours) and background error (colour shading) of the potential temperature at the bottom level at $t=26$ from (a) CNTL-noRIP, (b) RIP-RDM and (c) RIP-IESV. For Fig. 12, the scale is twice as shown. The contour interval is 0.075.

at the previous analysis times (initial) and background perturbations at the current analysis time (final). Since ESVs are derived from the LETKF ensemble, it is almost cost-free to identify the growing modes.

During the first analysis cycle of the LETKF assimilation, the leading FESV is already able to effectively capture the structure of the growing error related to background instabilities while the ensemble perturbations are still

developing these structures. Averagely, there are 10 growing modes in the pairs of ESVs derived from the QG model with 20 ensemble members. In terms of the relationship between the background errors, FESVs and ensemble perturbations, results suggest that the background error is better confined in the subspace spanned by the FESVs of growing modes than the subspace of the background ensemble perturbations with the same number of vectors. In addition, results also show that the structures of ESVs are sensitive to the choice of norm. The SF-based ESV_1 has less-stretched structures and a slower growth rate smaller than the PV-based ESV_1. Therefore, the PV-based ESVs are the natural choice for augmenting the background error covariance and for the purpose of capturing the fast growing errors during analysis corrections. Since IESVs grow much faster than LVs, and FESVs are close to convergence to LVs, perturbations based on leading IESVs grow faster than those based on FESVs, and are therefore preferable as additive inflation, even though they are not available at the beginning of the analysis window. For this reason, and since the leading ESVs do not change much over a window length of 12-h, we use the IESVs corresponding to the previous window.

Results show that all the experiments associated with additive covariance inflation can successfully improve the accuracy of the LETKF analysis compared with only using multiplicative inflation (CNTL). By better removing the growing errors in the analysis, the forecast errors at long forecast times are also reduced and the error growth rate becomes slower, compared with the CNTL forecast. Applying ESVs for additive inflation generally improves the LETKF analysis performance, whether perturbing the analysis ensemble with IESVs or the background ensemble with FESVs. Results also suggest that for making the growing modes effective as additive inflation, it is important to randomise the order of the ESVs so that each member has equal chance to be perturbed by the fast-growing modes. When using IESVs as the additive inflation, which grow faster, the subspace of the ensemble for capturing growing errors can be further expanded through non-linear evolution. Therefore, IESVs as the additive inflation (applied on the analysis ensemble) outperforms than the ones with FESVs (applied on the background ensemble). Among the experiments, significant improvement is further obtained when only the first 10 IESVs, identified as growing modes, are used, and the benefit is most evident at the area with large errors. Although FESV10 shows a comparable accuracy with the RDM, the overall performance for areas with large growing errors is still better than RDM. Finally, by improving the analysis accuracy, the advantage of using additive inflation can last till 2-d forecast and using ESV is even more advantageous, especially during the first 12-h forecast.

To show that using IESVs as additive perturbations can improve the flow-dependent structures associated with background instabilities, the IESVs are used in the LETKF-RIP system to perturb the smoothed analysis ensemble during the RIP iterations. The LETKF-RIP method (Kalnay and Yang, 2010) aims to accelerate EnKF's spin-up when the system is initialised from a state far from the truth (e.g. a cold start). Random perturbations, originally adopted in Kalnay and Yang (2010), are added on the smoothed analysis ensemble to avoid the smoothed analysis ensemble at previous analysis time evolving into the same analysis ensemble at the current analysis time. Through non-linear dynamics, these perturbations grow and stimulate the ensemble to better capture the dynamical growing errors and effectively use the observations. Results show that by using the first 10 IESVs, the LETKF's spin-up time is even shorter and more effectively corrects the growing errors. This again confirms that IESVs can be used to enhance the ensemble perturbations and help to better capture the subspace of growing errors.

In this work, the idea of using ESVs as the flow-dependent additive inflation is demonstrated with the QG model with simple dynamics. Such method can be easily implemented in any ensemble-based data assimilation framework, with negligible additional computational cost. In real applications with complex model, it may occur that some of the dynamical growing modes associated with model uncertainties are not captured if using the same model to generate the ensemble forecast. To consider the effect of the model errors in the ensemble, EVSs could be derived by performing ensemble forecasting with different convection, PBL and microphysics parameterisation schemes (Fujita et al., 2007) or with stochastic physics like the stochastic kinetic energy backscatter scheme (Lang et al., 2012). Further investigations such as the optimisation period for deriving ESVs or the possibility of capturing growing error in a target area such a hurricane will be carried out with realistic models with full physics and dynamics to understand the feasibility of this approach in realistic EnKF applications.

7. Acknowledgements

The authors are very grateful for valuable discussions with Prof. M. Zupanski from Colorado State University and Prof. Fuqing Zhang from Pennsylvania State University. The authors are also thankful to two anonymous reviewers and the subject editor, Dr. Nils Gustaffson, for their valuable comments and suggestions for improving this manuscript. S.-C. Yang is sponsored by National Science Council grant NSC-102-2111-M-008-202-MY3.

References

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

Annan, J. D. 2003. On the orthogonality of bred vectors. *Mon. Weather Rev.* **132**, 843–849.

Barkmeijer, J., Buizza, R., Palmer, T. N., Puri, K. and Mahfouf, J.-F. 2001. Tropical singular vectors computed with linearized diabatic physics. *Q. J. R. Meteorol. Soc.* **127**, 685–708.

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

Bishop, C. H. and Toth, Z. 1999. Ensemble transformation and adaptive observations. *J. Atmos. Sci.* **56**, 1748–1765.

Bojarova, J., Gustafsson, N., Johansson, Å. and Vignes, O. 2011. The ETKF rescaling scheme in HIRLAM. *Tellus A*. **63**, 385–401. DOI: 10.1111/j.1600-0870.2011.00513.x.

Buizza, R. and Palmer, T. N. 1995. The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.* **52**, 1434–1456.

Buizza, R., Tribba, J., Molteni, F. and Palmer, T. 1993. Computational of optimal unstable structures for a numerical weather prediction model. *Tellus A*. **45**, 388–407.

Chen, J.-H., Peng, M. S., Reynolds, C. A. and Wu, C.-C. 2009. Interpretation of tropical cyclone forecast sensitivity from the singular vector perspective. *J. Atmos. Sci.* **56**, 1748–1765.

Chen, S.-G., Wu, C.-C., Chen, J.-H. and Chou, K.-H. 2011. Validation and interpretation of adjoint-derived sensitivity steering vector as targeted observation guidance. *Mon. Weather Rev.* **139**, 1608–1625.

Corazza, M., Kalnay, E., Patil, D. J., Yang, S.-C., Morss, R. and co-authors. 2003. Use of the breeding technique to estimate the structure of the analysis "error of the day." *Nonlinear Process. Geophys.* **10**, 233–243.

Corazza, M., Kalnay, E. and Yang, S.-C. 2007. An implementation of the local ensemble Kalman filter in a quasigeostrophic model and comparison with 3D-Var. *Nonlinear Proc. Phys.* **14**, 89–101.

Dey, C. and Morone, L. L. 1985. Evolution of the national meteorological center global data assimilation system: January 1982–December 1983. *Mon. Weather Rev.* **113**, 304–318.

Doyle, J. D., Reynolds, C. A., Amerault, C. and Moskaitis, J. 2012. Adjoint sensitivity and predictability of tropical cyclogenesis. *J. Atmos. Sci.* **69**, 3535–3557.

Enomoto, T., Yamane, S. and Ohfuchi, W. 2006. Simple sensitivity analysis using ensemble forecast. In: *Proceedings of Third Workshop on Mechanisms of Climate Variation and its Predictability*, Disaster Prevention Research Institute, Kyoto University, Kyoto, pp. 40–43. (In Japanese.).

Enomoto, T., Yamane, S. and Ohfuchi, W. 2015. Simple sensitivity using ensemble forecasts. *J. Meteorol. Soc. Jpn.* **93**, 199–213. DOI: 10.2151/jmsj.2015-011.

Errico, R. M. and Vukićević, T. 1992. Sensitivity analysis using an adjoint of the PSU–NCAR mesoscale model. *Mon. Weather Rev.* **120**, 1644–1660.

Fujita, T., Stensrud, D. and Dowell, D. C. 2007. Surface data assimilation using an ensemble filter approach with initial condition and model physics uncertainties. *Mon. Weather Rev.* **135**, 1846–1868.

Gelaro, R., Langland, R., Rohaly, G. D. and Rosmond, T. E. 1999. An assessment of the singular vector approach to targeted observing using the FASTEX dataset. *Q. J. Roy. Meteorol. Soc.* **125**, 3299–3328.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K. and Hunt, B. R. 2011. Balance and ensemble Kalman filter localization techniques. *Mon. Weather Rev.* **139**, 511–522.

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M. and coauthors. 2005. Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. *Mon. Weather Rev.* **133**, 604–620.

Hunt, B., Kostelich, E. and Syzunogh, I. 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*. **230**, 112–126.

Ito, K. and Wu, C.-C. 2013. Typhoon-position-oriented sensitivity analysis. Part I: theory and verification. *J. Atmos. Sci.* **70**, 2525–2546.

Kalnay, E. 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, New York, NY, USA, 340 pp.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C. and Ballabrera-Poy, J. 2007. Response to the discussion on "4D-var or ensemble Kalman filter?". *Tellus A*. **59**, 778–780.

Kalnay, E. and Toth, Z. 1994. Removing growing errors in the analysis cycle, In: *Tenth Conference on Numerical Weather Prediction*, American Meteorological Society, Portland, pp. 212–215.

Kalnay, E. and Yang, S.-C. 2010. Accelerating the spin up of the LETKF. *Q. J. Roy. Meteorol. Soc.* **136**, 1644–1651.

Lang, S. T. K., Leutbecher, M. and Jones, S. C. 2012. Impact of perturbation methods in the ECMWF ensemble prediction system on tropical cyclone forecasts. *Q. J. Roy. Meteorol. Soc.* **138**, 2030–2046.

Matsueda, M., Kyouda, M., Toth, Z., Tanaka, H. L. and Tsuyuki, T. 2011. Predictability of an atmospheric blocking event that occurred on 15 December 2005. *Mon. Weather Rev.* **139**, 2455–2470.

Morss, R. E. 1998. *Adaptive observations: idealized sampling strategies for improving numerical weather prediction*. PhD Thesis. Massachusetts Institute of Technology, Cambridge, MA, 225 pp.

Nishii, K. and Nakamura, H. 2010. Three-dimensional evolution of ensemble forecast spread during the onset of a stratospheric sudden warming event in January 2006. *Q. J. Roy. Meteorol. Soc.* **136**, 894–905.

Norwood, A., Kalnay, E., Ide, K., Yang, S-C. and Wolfe, C. 2013. Lyapunov, singular and bred vectors in a multi-scale system: an empirical exploration of vectors related to instabilities. *J. Phys. A*, special issue of Lyapunov vectors **46**, 20 pp.

Palmer, T. N., Gelaro, R., Barkmeijer, J. and Buizza, R. 1998. Singular vectors, metrics and adaptive observations. *J. Atmos. Sci.* **55**, 633–653.

Peña, M. and Kalnay, E. 2004. Separating fast and slow modes in coupled chaotic systems. *Nonlinear Process. Geophys.* **11**, 319–327.

Rabier, F., Klinker, E., Courtier, P. and Hollingsworth, A. 1996. Sensitivity of forecast errors to initial conditions. *Q. J. Roy. Meteorol. Soc.* **122**, 121–150.

Rotunno, R. and Bao, J. W. 1996. A case study of cyclogenesis using a model hierarchy. *Mon. Weather Rev.* **124**, 1051–1066.

Snyder, C., Hamill, T. M. and Trier, S. B. 2003. Linear evolution of error covariances in a quasigeostrophic model. *Mon. Weather Rev.* **131**, 189–205.

Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.* **74**, 2317–2330.

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

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y. and Toth, Z. 2008. Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Weather Rev.* **136**, 463–481.

Wu, C.-C., Chen, J.-H., Majumdar, S. J., Peng, M. S., Reynolds, C. A. and co-authors. 2009. Inter-comparison of targeted observation guidance for tropical cyclones in the north western Pacific. *Mon. Weather Rev.* **137**, 2471–2492.

Yamaguchi, M., Iriguchi, T., Nakazawa, T. and Wu, C.-C. 2009. An observing system experiment for typhoon *Conson* (2004) using a singular vector method and DOTSTAR data. *Mon. Weather Rev.* **137**, 2801–2816.

Yang, S.-C. 2005. *Bred vectors in the NASA NSIPP Global Coupled Model and their application to coupled ensemble predictions and data assimilation, Appendix B.* PhD Thesis. University of Maryland, College Park, 174 pp. Online at: http://hdl.handle.net/1903/2477

Yang, S.-C., Baker, D., Li, H., Huff, M., Nagpal, G. and co-authors. 2006. Data assimilation as synchronization of truth and model: experiments with the 3-variable Lorenz system. *J. Atmos. Sci.* **63**, 2340–2354.

Yang, S.-C., Corazza, M., Carrassi, A., Kalnay, E. and Miyoshi, T. 2009. Comparison of ensemble-based and variational-based data assimilation schemes in a quasi-geostrophic model. *Mov. Weather Rev.* **137**, 693–709.

Yang, S.-C. and Kalnay, E. 2012. Handling nonlinearity and non-Gaussianity in ensemble Kalman filter. submitted to a special collection ''Intercomparisons of 4D-variational assimilation and the ensemble Kalman filter.'' *Mon. Weather Rev.* **140**, 2628–2646.

Yang, S-C., Kalnay, E. and Miyoshi, T. 2012. Improve EnKF spin-up for typhoon assimilation and prediction. *Weather Forecast.* **27**, 878–897.

Yang, S.-C., Lin, K.-J., Miyoshi, T. and Kalnay, E. 2013. Improving the spin-up of regional EnKF for typhoon assimilation and forecasting with Typhoon Sinlaku (2008). *Tellus A.* **65**, 20804, doi: http://dx.doi.org/10.3402/tellusa.v65i0.20804

Zhang, F., Snyder, C. and Sun, J. 2004. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Weather Rev.* **132**, 1238–1253.