

Flow-dependent versus flow-independent initial perturbations for ensemble prediction

By LINUS MAGNUSSON*, JONAS NYCANDER and ERLAND KÄLLÉN, *Department of Meteorology, Stockholm University, S-106 91 Stockholm, Sweden*

(Manuscript received 12 June 2008; in final form 27 November 2008)

ABSTRACT

Ensemble prediction relies on a faithful representation of initial uncertainties in a forecasting system. Early research on initial perturbation methods tested random perturbations by adding ‘white noise’ to the analysis. Here, an alternative kind of random perturbations is introduced by using the difference between two randomly chosen atmospheric states (i.e. analyses). It yields perturbations (random field, RF, perturbations) in approximate flow balance.

The RF method is compared with the operational singular vector based ensemble at European Centre for Medium Range Weather Forecasts (ECMWF) and the ensemble transform (ET) method. All three methods have been implemented on the ECMWF IFS-model with resolution $T_L255L40$. The properties of the different perturbation methods have been investigated both by comparing the dynamical properties and the quality of the ensembles in terms of different skill scores. The results show that the RF perturbations initially have the same dynamical properties as the natural variability of the atmosphere. After a day of integration, the perturbations from all three methods converge. The skill scores indicate a statistically significant advantage for the RF method for the first 2–3 d for the most of the evaluated parameters. For the medium range (3–8 d), the differences are very small.

1. Introduction

Due to the chaotic behaviour of the atmosphere, small initial errors will grow with time and eventually make the weather forecast uncorrelated with the true state (Lorenz, 1963). The predictability is first lost for short scales and affects longer scales with increasing forecast time. After two weeks almost all predictability is lost (Lorenz, 1969). The growth rate of the forecast uncertainties varies from day to day, and to estimate the impact of initial errors and to provide probabilistic forecasts, forecast centres produce ensemble forecasts. In this study, we introduce a new method for ensemble forecasting that randomly samples the initial uncertainties by taking the difference between two arbitrary atmospheric states. The method will be compared with two perturbation methods used operationally at forecast centres.

The basis for ensemble forecasting is the idea of Monte Carlo simulations. By adding small perturbations to the analysis of the same magnitude as the analysis errors, and rerunning the model from the new initial condition, the effect of the initial errors can be estimated (Leith, 1974). During the 1970s, experiments with pure random perturbations scaled with the estimated analysis error were made, but it was soon found that random perturba-

tions were inappropriate (Hollingsworth, 1979). This is because purely random perturbations yield unbalanced flow structures and lead to the perturbation energy being dissipated as gravity waves during the initial time steps.

Instead of using pure random perturbations, forecast centres use constrained, flow-dependent perturbations. The European Centre for Medium Range Weather Forecasts (ECMWF) uses singular vector perturbations, aimed at sampling the fastest-growing perturbations in the atmosphere for a specified optimization time (Lorenz, 1965; Palmer, 1993). The National Center of Environmental Prediction (NCEP) uses the ensemble transform (ET) technique (Bishop and Toth, 1999; Wei et al., 2008), which is an improved version of the breeding-vector (BV) technique (Toth and Kalnay, 1993, 1997), to perturb the initial conditions. The ET technique transforms the forecast perturbations to being orthogonal in the inverse analysis error variance norm. The ET method has a strong connection to the perturbation method ET Kalman filter (ETKF) described in, for example, Wang and Bishop (2003). An ensemble system using ETKF is used operationally at MetOffice (Bowler et al., 2008) and a study by Szunyogh et al. (2008) shows promising results using local ETKF. The ETKF is an application of ensemble Kalman filter and yields perturbations orthonormal in the observational space. Both the SV and the ET techniques use information about the current flow situation to calculate the perturbations and thus to obtain perturbations consistent with the flow. The main

*Corresponding author.

e-mail: linusm@misu.su.se

DOI: 10.1111/j.1600-0870.2008.00385.x

motivation for using flow-dependent initial perturbations is to obtain an ensemble with sufficient dispersion in the medium range, without using excessively large initial perturbation amplitudes.

Using low-order models, pure random perturbations have shown advantages compared to singular vectors and BVs (Anderson, 1997; Bowler, 2006). Many attempts have also been made to use some kind of random perturbations in more realistic settings. During the 1980s, simulations were made using perturbations obtained by adding random numbers to the variables and then balancing the field using normal mode initialization (Tribbia and Baumhefner, 1988). Toth and Kalnay (1993) mention that using random but balanced perturbations is not the best way of making ensemble forecasts because the perturbations require hours or days before they organize into dynamically unstable modes. Some recent studies have used pure random perturbations for ensemble prediction on limited area models. Zhang (2005) used it for forecast error studies of a winter storm over North America. The study showed that the perturbation energy decreases in the mean during the first 12 h of integration, but that the random perturbations quickly evolve into coherent, multivariate, structures.

A different attempt to use random variables when creating initial perturbations for an NWP system is the use of stochastic Ensemble Kalman filter schemes (opposite to the deterministic schemes such as the ones used by Bowler et al. (2008) and Szunyogh et al. (2008)). One example is the Ensemble Kalman filter technique using perturbed observations (Houtekamer et al., 1996). By adding random perturbations to the observations and rerunning the assimilation system, a new analysis is obtained that is in balance with the atmospheric dynamics. Studies have shown advantages for the method compared to singular vectors and BVs (Talagrand et al., 2007). Disadvantages are that it is computer time consuming to run the data assimilation system several times and also that the perturbations of the observations need an inflation factor to yield a proper amplitude of the initial perturbation (Buizza et al., 2005).

Another stochastic approach for the ensemble Kalman filter technique is to add random perturbations to the ensemble member background forecast. Several methods were tested in Hamill and Whitaker (2005). Another possibility of constructing perturbations not coupled with the background flow is to use historical perturbations. In Wei and Toth (2003) breeding perturbations valid 8 d earlier are used as perturbations. The study showed that the method is not as good as the breeding method. Mureau et al. (1993) use a linear combination of the most recent differences between 6-h forecasts and the analysis valid at the same time (first guess error) as initial perturbations. They conclude that the singular vector method is more appropriate for ensemble forecasting, because of a faster initial perturbation growth.

For a predictability study in a two-level Global Climate Model (GCM), Schubert and Suarez (1989) use perturbations

determined by two randomly selected historical model states. The same strategy is used by Hamill and Whitaker (2005) to add 'noise' in ensemble data assimilation. We use this idea of constructing perturbations for ensemble forecasting by simply taking differences between randomly chosen analyses of atmospheric states (random field, RF, perturbations). Using this technique, we obtain atmospheric flow structures that include some balances but are unconnected with the actual flow situation.

Such a technique can be seen as a cheap, lowest order baseline method for initial perturbations for ensemble prediction. Flow-dependent perturbation techniques should give results that are superior to this baseline method. If that is not the case we may have to rethink the way in which flow-dependent perturbations are constructed.

A cheap alternative to create an ensemble by rerunning the same model several times with different initial conditions, are so-called poor man's ensembles. The ensemble could either be constructed using forecasts from different forecast centres (e.g. Arribas et al., 2005) or a lagged-forecast ensemble based on recent forecasts (recently discussed in Buizza, 2008). The disadvantage with both methods is that the number of ensemble members is limited (by the number of forecast centres or the number of recent forecast realizations). An open question is also how to weight the different members together in the ensemble. Another possibility is to select events from the past showing similarities with the current atmospheric situation to form an ensemble of historical weather developments (Talagrand et al., 1999; Candille and Talagrand, 2005). In this study, we will focus on an ensemble system that uses the same forecast model and starts at a given time instant with perturbed initial conditions.

In this study, we introduce the ET method into the ECMWF Integrated Forecast System (IFS) and compare with the operational singular vector system as in Magnusson et al. (2008b). We also introduce the RF perturbation method into the same forecast system. The three initial perturbation methods are described in Section 2, and the model and data in Section 3. In Section 4 we compare the horizontal distribution and several dynamical properties of the ensemble perturbations. The quality of the ensembles in terms of skill scores is investigated in Section 5 and finally we summarize the results in Section 6.

2. Initial perturbation methodologies

2.1. ECMWF singular vector approach

In the ECMWF, ensemble prediction system (EPS), singular vectors are used to find the initial perturbations (Barkmeijer et al., 1999; Leutbecher and Palmer, 2008). This approach is based on the idea that it is most important to sample directions in phase space that are characterized by maximal amplification rates. These directions can be found by the singular vector approach.

Using a tangent linear forward and adjoint models during a prescribed time period (the optimization time), the singular vectors are computed by solving an eigenvalue problem. The singular vectors used by ECMWF are computed to maximize total energy growth over a 48-h time interval. The total energy norm is defined as

$$|x|_{\text{Etot}}^2 = \iiint \frac{1}{2} \left[u^2 + v^2 + \frac{C_p}{T_r} T^2 \right] dS \frac{dp}{g} + \iint \frac{1}{2} \frac{R_d T_r P_r}{g} (\ln P_s)^2 dS, \quad (1)$$

where u and v are the wind perturbations, T is the temperature perturbation and P_s is the surface pressure perturbation. T_r (300 K) and P_r (800 hPa) are a reference temperature and pressure, respectively.

Fifty singular vectors are computed separately for each hemisphere, aimed at finding perturbations localized to the extratropics for each hemisphere (30°–90° latitude). A global computation of singular vectors would yield too few singular vectors in the summer hemisphere; therefore two separate sets of singular vectors are computed. In addition to initial singular vectors, the initial perturbations are also based on 48-h linearly evolved SV, computed 48 h prior to the ensemble starting time. This is done to represent slower-growing large-scale structures. For the tropics, the singular vector computation has been limited to the vicinity of tropical depressions and tropical cyclones (Barkmeijer et al., 2001).

In our study, the different sets of singular vectors are combined to 10 perturbations, using Gaussian sampling, which is used operationally since 2004 (Leutbecher and Palmer, 2008). The standard deviation of the Gaussian distribution depends on analysis error estimates provided by the ECMWF four-dimensional data assimilation (4D-Var) scheme (Fisher and Courtier, 1995).

2.2. Ensemble transform perturbation

The operational ensemble system at NCEP uses ET perturbations (Wei et al., 2008), which have also been implemented at Naval Research Laboratory (McLay et al., 2008). The ET method was first introduced in Bishop and Toth (1999) aiming for adaptive observations. It is an extension of the breeding method (Toth and Kalnay, 1993, 1997), which dynamically recycles the perturbations and is aimed at sampling the subspace of the fastest-growing error modes (Buizza et al., 2005). The ET method transforms the forecast perturbations so that they are orthonormal in the inverse analysis error variance norm.

The ET system at NCEP uses the simplex transformation (Wang et al., 2004) to ensure that the ensemble is centred around the analysis. In this study, we instead use a plus–minus symmetry to obtain an ensemble with the same kind of symmetry as the operational singular vector setup. To use plus–minus symmetry, the ET method described in Wei et al. (2008) has been slightly modified.

Define a forecast and analysis perturbation matrix for an ensemble with k members as

$$X^f = [x_1^f, x_2^f, \dots, x_m^f, \dots, x_k^f], \quad (2)$$

$$X^a = [x_1^a, x_2^a, \dots, x_m^a, \dots, x_k^a], \quad (3)$$

where x_m^f is the difference between the model state vector of an ensemble member m and the control (unperturbed) forecast after 6 h and x_m^a is the difference between ensemble member m and the analysis initially. Every 6 h the forecast perturbations are transformed into analysis perturbation by the transformation

$$X^a = X^f C \Gamma^{-1/2}, \quad (4)$$

where the columns of C are orthonormal eigenvectors of $\frac{X^{fT} P_{op}^{a-1} X^f}{n}$ and Γ is a diagonal matrix containing the associated eigenvalues. The diagonal matrix P_{op}^a containing the analysis error variances is derived from the 4D-Var data assimilation system and n is the length of the model state vector. The resulting matrix X^a contains orthogonal columns, that is, the analysis perturbations are orthogonal in the analysis error variance norm.

It was found by Wei et al. (2008) that ET perturbations, although using the analysis variance norm, cannot form an initial perturbation distribution similar to the analysis error. Therefore, the operational version of ET used at NCEP also includes a regionally varying rescaling (Toth and Kalnay, 1997; Wang and Bishop, 2003), to ensure that the perturbation magnitude varies in accordance with the uncertainties in the analysis. The mask ($c_m(i, j)$), used in our study, is only a function of latitude and longitude and is defined as

$$c_m(i, j) = \begin{cases} 1 & y_m^a(i, j) \leq e(i, j) \\ \frac{e(i, j)}{y_m^a(i, j)} & y_m^a(i, j) > e(i, j) \end{cases}, \quad (5)$$

where $e(i, j)$ is the vertically integrated analysis error variance (P_{op}^a) measured in total energy and $y_m^a(i, j)$ is the total energy of the analysis perturbation ($x_m^a(i, j, k)$) vertically integrated. Both the analysis uncertainty and the forecast perturbation are horizontally smoothed (truncated to T7) before the vertical integration and the total energy calculation.

The mask is applied as

$$x_m^{aR}(i, j, k) = \alpha \frac{x_m^a(i, j, k) c_m(i, j)}{|x_m^a(i, j, k) c_m(i, j)|_{AE}}, \quad (6)$$

where the global analysis error variance norm in the denominator is defined as

$$|x|_{AE}^2 = \frac{x^T P_{op}^{a-1} x}{n}. \quad (7)$$

The constant parameter α is used to tune the initial amplitude of the perturbation. The use of a mask can have a minor effect on the orthogonality of the analysis perturbations. In addition, the mask could give rise to undesirable perturbation imbalances as it does not have any dynamical constraints.

2.3. Random field perturbation

The simplest way to perturb an initial state is to add white noise. However, in a dynamical system the solutions should be in line with the underlying dynamics (Judd et al., 2008). This can be illustrated with a low-order model such as Lorenz 63 (see e.g. Magnusson et al., 2008a). If the system is perturbed away from the Lorenz attractor it rapidly returns to it, and only a small fraction of the initial perturbation remains. In an atmospheric model, an unbalanced initial state generally generates gravity waves. Although it is uncertain whether the atmosphere has a low dimensional attractor, we know that there exist constraints for the solutions, for example, geostrophy in the mid-latitudes and equatorial waves in the tropics.

It is illustrated in Magnusson et al. (2008a) that an efficient way of generating growing perturbations is to construct random perturbations in vicinity of the attractor (normal mode perturbations). To create random perturbations under realistic constraints in an NWP model, we introduce a method called RF perturbations. A similar method was previously used by Schubert and Suarez (1989) for predictability studies in a two-layer GCM. The perturbations are constructed by randomly choosing two independent analysed atmospheric states and calculating the difference between them. This difference is then rescaled to an amplitude convenient for ensemble perturbations. During these linear operations linear balances in the analyses remain. However, the balances in the analyses could be different from those in the NWP model due to non-linearities, for example, in gradient wind balance (Holton, 2004). Because of this, the perturbations are not perfectly balanced initially.

The perturbations are calculated as

$$x_m^a = \alpha \frac{a_{d1} - a_{d2}}{|a_{d1} - a_{d2}|_{\text{Etot}}}, \quad (8)$$

where $a_{d1,d2}$ is the state vector of the analysis from the date $d1$ or $d2$, respectively. The dates are randomly chosen from year 2000 to 2005 with the constraint that they should be from the same season as the actual date of the forecast and that $d1$ and $d2$ should be from different years. The former constraint is to preserve the seasonal characteristics of the variability as well as possible. The latter is to ensure that the analyses are uncorrelated. The difference field is then normalized so that all initial perturbations have the same amplitude in the total energy norm. As for the ET method a ‘tuning constant’ α is used to obtain a sufficient ensemble dispersion. The value of α is the same for all RF ensembles.

3. Model and data

For the experiments, the ECMWF IFS (IFS-model) was used with spectral resolution T_L255L40 and model cycle 31r1 (for details see <http://www.ecmwf.int/research/ifsdocs/CY31r1>). For each method, 20 ensemble members are used. The ensemble systems were run once per day (00Z) between 1 December 2005

and 28 February 2006 resulting in 90 forecast cases. For all three experiments, the temperature, surface pressure and wind fields are perturbed.

To make it possible to use 6-h cycle length for the ET perturbation, three additional 6-h forecasts (06Z, 12Z and 18Z) were run for the ET ensemble for each day.

For all three experiments, 10 different perturbation vectors are calculated. These are both added to and subtracted from the analysis to obtain a 20-member ensemble centred on the analysis. This strategy to obtain an ensemble with the same mean as the analysis yields an ensemble with variance in only half as many directions as there are perturbations. This applies to the forecasts as long as the NWP model operator is linear. As mentioned in Section 2.2, it is possible to use a simplex transformation to overcome this problem. Using a simplex transformation (for a k member ensemble), $k - 1$ orthogonal modes are spanned initially compared to $\frac{k}{2}$ for a plus-minus symmetric ensemble. To test the impact of the simplex transformation for the ET ensemble, an additional experiment was run for the period 1 December 2005–15 January 2006 using ET with simplex transformation. The method was implemented as described in Wei et al. (2008).

The initial perturbation amplitude, for both ET and RF, has been chosen to obtain a mean ensemble spread of the same magnitude for 500 hPa geopotential height as SV-EPS for the Northern Hemisphere at day 3 (see Section 5, Fig. 9a). The period 1–15 December 2005 is used to find the value of the tuning factor α that satisfies this criterion. For ET this implies that the initial amplitude is set to 90% of the estimated analysis error standard deviation, calculated from the analysis error variance norm (eq. 7; further discussed in Section 6). For RF, the initial amplitude is fixed for a value of the total energy, yielding on average 70% of the analysis error standard deviation. This tuning strategy has been chosen to avoid differences in scores depending on differences in ensemble spread for the Northern Hemisphere and also to make it possible comparing the ensemble spread for different variables between the experiments.

The analysis error variances are obtained from the 4D-Var data assimilation system as an operationally archived field (Fisher and Courtier, 1995; Barkmeijer et al., 1998). The analysis error variances are calculated as the difference between the background errors and a correction derived from the eigenvectors of the Hessian of the cost function (ECMWF IFS Documentation CY28r1). The estimate of the analysis error variances depends on the observation density (affecting the cost function) and the flow situation within the data assimilation window (Gustafsson, 2007).

To represent model error, ECMWF uses the stochastic diabatic tendency perturbations (stochastic physics; Buizza et al., 1999). By scaling the tendencies from the physical schemes with a random number, the uncertainties in the parametrizations are simulated. The random values are between 0.5 and 1.5 drawn from a uniform distribution. The same random number is used in a box of $10^\circ \times 10^\circ$ extent and over a period of six time steps

(4.5 h). This stochastic scheme is used in all ensemble forecast experiments.

4. Perturbation properties

4.1. Geographical perturbation distribution

Studying the temperature at 700 hPa of one arbitrary perturbation for each method (not shown), we see very distinct features. The ET perturbations are dominated by a few (2–5) local maxima with very high amplitude. The SV perturbations are also localized but contain more local maxima in each perturbation compared to ET. This is mainly due to the use of several singular vectors combined into one perturbation field. The amplitude of each local maximum is lower than for the ET perturbations. The SV perturbations have, especially in the Southern Hemisphere, a wave pattern, indicating the connection to baroclinic zones (see below). Regarding the RF perturbations, the perturbation structure appears to be more large scale and with several maxima around the globe, especially around the mid-latitudes. The amplitude of each maximum is about half of the maximum amplitude for the SV perturbations and one-third of the ET perturbation maxima.

During the first days of integration, the perturbation patterns in the different experiments converge. It indicates that the perturbations grow by extracting energy from the same atmospheric structures in all experiments.

To illustrate the mean perturbation distribution, Fig. 1 shows the initial perturbation energy (see eq. 1) on the 500 hPa level for the three different experiments as a mean for the period 1 December 2005–28 February 2006. For the SV-EPS we see that, in the Northern Hemisphere, the energy is large in the western Pacific but also in the North Atlantic. The initial perturbation location coincides with cyclogenetic areas (Hoskins and Hodges, 2002). In the Southern Hemisphere, the energy is more equally spread in meridional bands with a maximum around 50°S. For ET-EPS, the initial perturbation energy is more equally spread over the hemispheres. This is mainly due to the regional rescaling mask applied for this experiment. But we can see that the small, dominant structures in each perturbation imply that the mean for 90 forecasts (900 independent perturbation fields) is not so smooth. For RF-EPS, by definition, the perturbations appear in the regions where the variability of the atmosphere is highest. We see that for the Northern Hemisphere, this is in the storm tracks over the Pacific and the Atlantic, somewhat east of the maxima for the SV-EPS.

Figure 2 shows the same as Fig. 1 for the estimated analysis error from the 4D-Var. Comparing the analysis error with the perturbations, we see that ET-EPS and RF-EPS show the best resemblance. For ET-EPS, this is because of the regional rescaling mask that is constrained by the estimated analysis error. For RF-EPS, this means that the highest uncertainties in the atmosphere are present in regions with the high-

est variability. For SV-EPS, the patterns of the analysis error and the distribution of perturbation energy are less similar over the Northern Hemisphere than compared to the other two experiments.

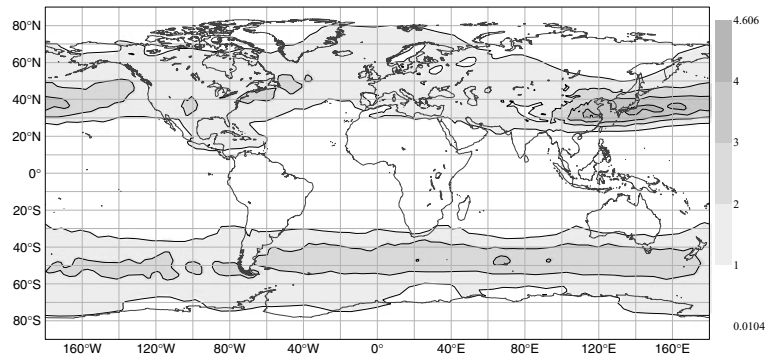
The clearest difference between the experiments is the initial amplitude of the perturbations. The SV-EPS experiments start with the lowest (global average of 0.92 J kg^{-1} for 500 hPa level) initial amplitude in a total energy sense compared to the other two experiments. The RF experiment start with a lower initial amplitude (1.73 J kg^{-1}) compared to ET (3.93 J kg^{-1}). This is mainly because the ET perturbations are more equally spread over the globe while the RF perturbations are dominated by the Northern Hemisphere (the area the initial amplitude is tuned for). But the need for a lower initial amplitude for RF-EPS also indicates that the RF perturbations are close to balance initially and do not need a high initial amplitude to compensate for generation of gravity waves.

Figure 3 shows the perturbation energy after 48 h. Here, we see large similarities between all three experiments. In the Northern Hemisphere, maxima of perturbation energy in all three experiments are located in the Atlantic and the Pacific and minima over Central Asia and Canada. Regarding the mean amplitude, RF-EPS has a somewhat lower amplitude of the perturbations for the 500 hPa level (11.3 J kg^{-1}) than the other two experiments (SV— 12.6 J kg^{-1} , ET— 13.1 J kg^{-1}), mainly because of a lower spread on the Southern Hemisphere. The reason for the fast growth of the SV perturbations is that the ECMWF singular vectors are optimized to yield the fastest perturbation growth for the total energy during the first 48 h. Comparing the perturbation energy initially and after 48 h ET-EPS shows the slowest perturbation growth. One explanation is that the perturbations in the tropics generally grow slowly (and in some cases even decrease). This could be due to imbalances introduced by the regional rescaling mask. But also for the mid-latitudes, we see a slower perturbation growth compared to the RF perturbations.

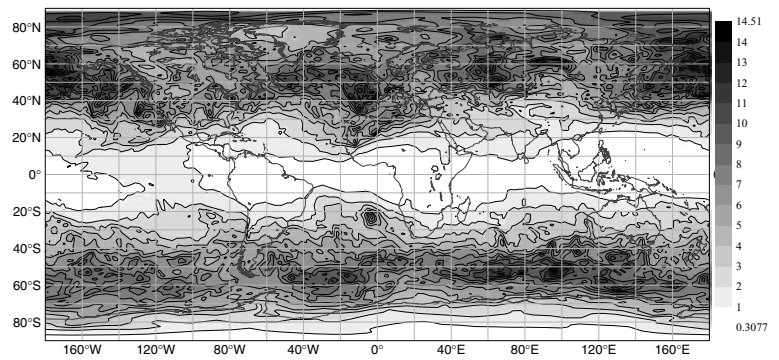
To compare the perturbation distribution after 48 h with the forecast error, the forecast error for the control forecast is plotted in Fig. 4 (Because the error of the control forecast is used, the field is only an average of 90 realizations, compared to 900 for the perturbation fields. Therefore, the field of the forecast error is not as smooth as the perturbation field). The forecast error in the Northern Hemisphere is largest in the storm tracks. This is the same pattern as seen for the perturbations after 48 h for all experiments. Therefore, we can conclude that the perturbation distribution for all three experiments resembles the distribution of the forecast error in the mid-latitudes after 48 h. For the tropics, we see that all three experiments have too little perturbation energy compared to the forecast error.

To further investigate how the different perturbation strategies are linked to cyclogenetic areas and baroclinic instability zones, we have calculated the spatial correlation coefficient between the standard deviation of the ensemble (ensemble spread) and the

(a) SV-EPS



(b) ET-EPS



(c) RF-EPS

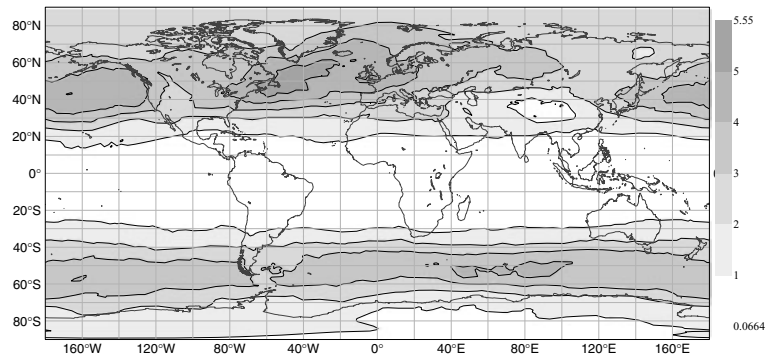


Fig. 1. Initial perturbations in total energy for 500 hPa level. Mean of 1 December 2005–28 February 2006. Contour interval 1 J kg^{-1} .

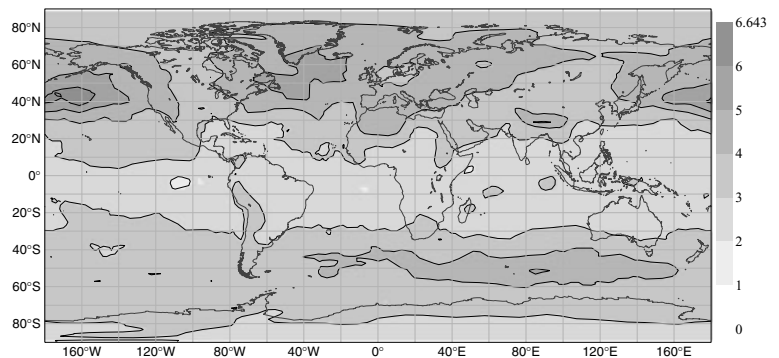
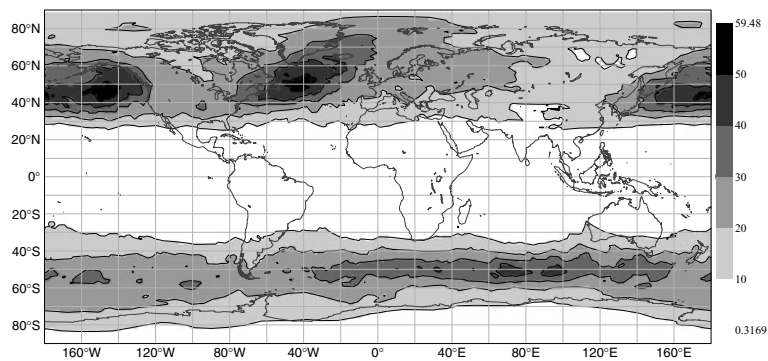
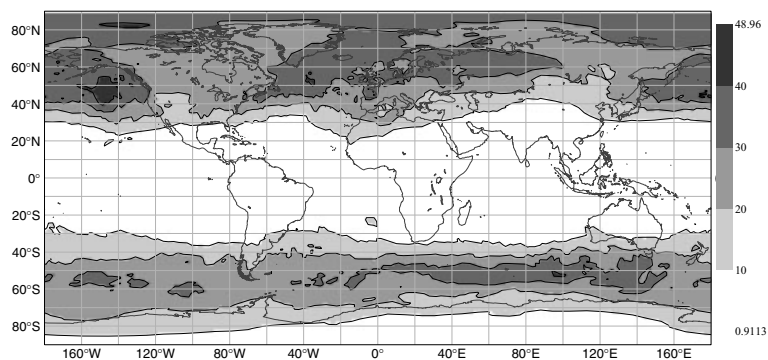


Fig. 2. Analysis error variance estimate provided by 4D-Var in total (dry) energy for 500 hPa level. Mean of 1 December 2005–28 February 2006. Contour interval 1 J kg^{-1} .

(a) SV-EPS



(b) ET-EPS



(c) RF-EPS

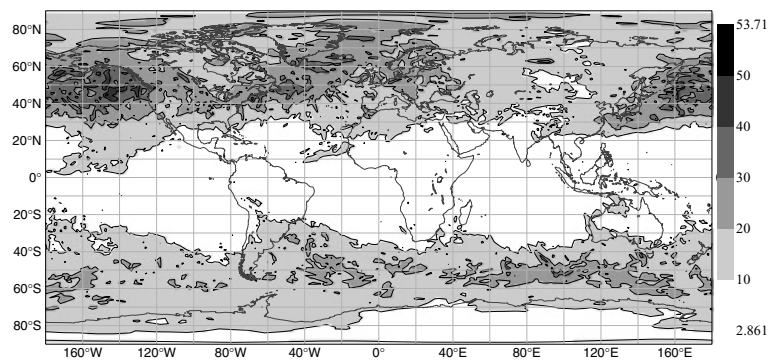
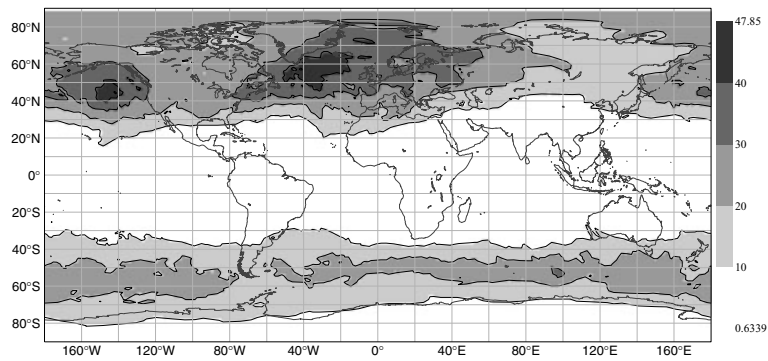


Fig. 3. Total energy of ensemble perturbations after 48 h for 500 hPa level. Mean of 1 December 2005–28 February 2006. Contour interval 10 J kg⁻¹.

Fig. 4. Error of control forecast for 48-h forecast time in total (dry) energy for 500 hPa level. Mean of 1 December 2005–28 February 2006. Contour interval 10 J kg⁻¹.

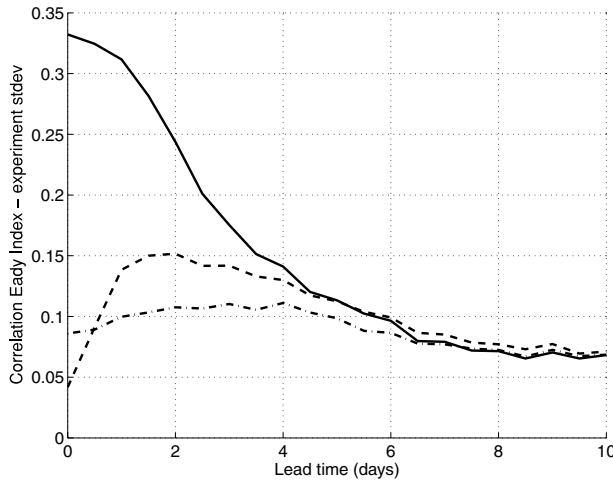


Fig. 5. Correlation coefficient between Eady index and the ensemble spread for z500 (1 December 2005–28 February 2006). The ET-EPS (dash-dotted), RF-EPS (dashed) and SV-EPS (solid).

Eady index between 20°N and 70°N. The Eady index is defined as (Lindzen and Farrell, 1980)

$$E = 0.3125 \frac{f}{N} \frac{dV}{dz}, \quad (9)$$

where V is the horizontal wind and the static stability is defined as $N = (\frac{1}{g} \frac{d\ln\theta}{dz})^{1/2}$. The vertical derivatives are calculated between 700 and 300 hPa from the control forecast.

The correlation between the Eady index and the ensemble spread for z500 as a function of forecast time is plotted in Fig. 5. Initially, SV has a higher correlation with the Eady index, compared to the other two experiments. It indicates that the SV-EPS perturbations are connected to baroclinic zones (Hoskins et al., 2000). For RF-EPS and ET-EPS, the initial perturbations are almost uncorrelated with the Eady index, but we see for RF-EPS that the correlation increases during the first day. This could be related to RF perturbations in the vicinity of baroclinic zones being ‘activated’ and starting to grow. For the ET experiment, the initial perturbations are not concentrated around storm tracks and that could be a reason why the correlation with the Eady index is low compared to the RF-EPS experiment. For long-forecast times, the results for the different experiments converge to a low correlation value.

To test the significance of the results of the correlation between the Eady index and the ensemble spread the Student’s t -test (Wilks, 2006) with confidence limit 95% was applied to the differences between the experiments. The difference between SV-EPS and RF-EPS is statistically significant until day 3.5, RF-EPS and ET-EPS until day 4 (except day 0.5) and ET-EPS and SV-EPS until day 4.

It is worth mentioning that the results for long-forecast length are affected by the fact that the control forecast (used to calculate the Eady index) and the ensemble mean are uncorrelated. Therefore, the correlation between the ensemble spread

and Eady index for the control forecast is not relevant to study for long-forecast ranges.

4.2. Energetics

To evaluate the dynamical properties of the perturbations from the different methods, we have calculated the ratio between the baroclinic and barotropic modes in the perturbations (Fig. 6), the ratio between potential and kinetic energy (Fig. 7) for the 500 hPa level and the perturbation spectra for z500 (Fig. 8).

To evaluate the baroclinicity in the perturbations, we have calculated the baroclinic and barotropic modes defined as

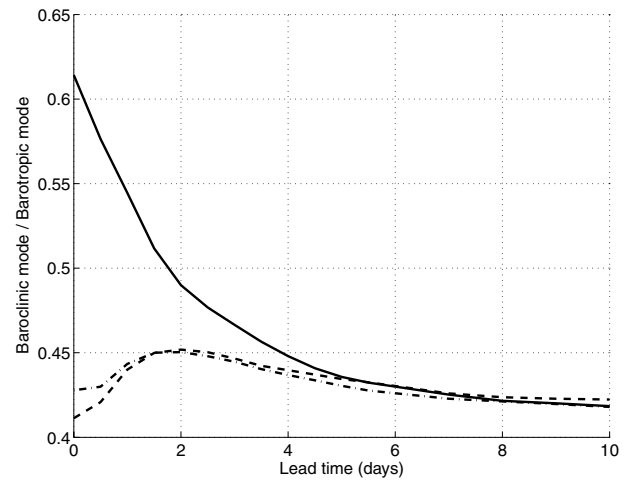


Fig. 6. The ratio $\frac{z_T}{z_M}$ between the baroclinic and barotropic mode amplitudes as defined in eqs. (10)–(11) (1 December 2005–28 February 2006). The ET-EPS (dash-dotted), RF-EPS (dashed) and SV-EPS (solid).

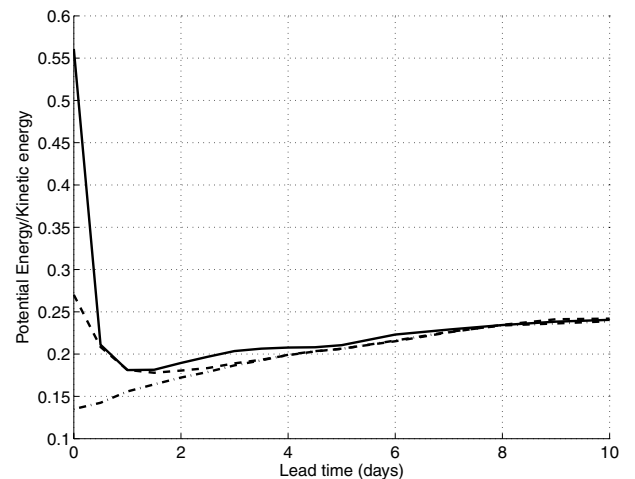


Fig. 7. The ratio between the potential and kinetic perturbation energy of the 500 hPa level (1 December 2005–28 February 2006). The ET-EPS (dash-dotted), RF-EPS (dashed) and SV-EPS (solid).

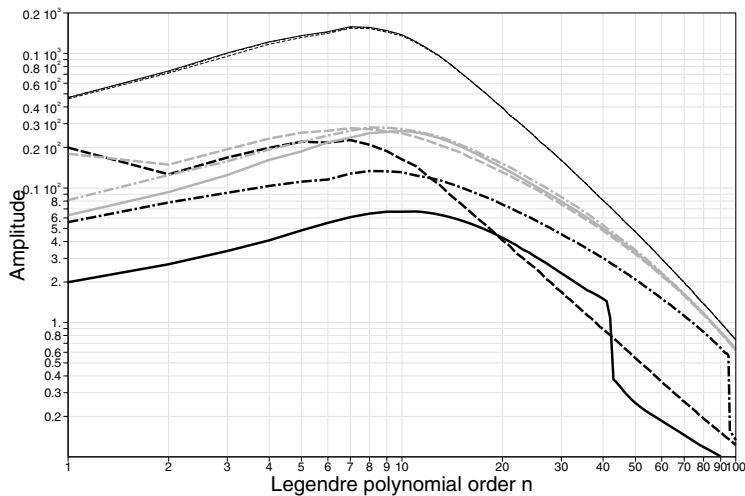


Fig. 8. Amplitude of the spectral components of the perturbations for z500 (1 December 2005–28 February 2006). Initial perturbation (black, thick), +48 h (grey, thick) and +240 h (black, thin). The ET-EPS (dash-dotted), RF-EPS (dashed) and SV-EPS (solid).

(Holton, 2004)

$$z_T = \frac{(z'_{300} - z'_{700})}{2} \text{ baroclinic,} \quad (10)$$

$$z_M = \frac{(z'_{300} + z'_{700})}{2} \text{ barotropic,} \quad (11)$$

where z' is the geopotential height perturbation from the ensemble mean. The baroclinic mode (eq. 10) represents the thermal part of the perturbation and is given by the average temperature perturbation between the 300 and 700 hPa levels. The barotropic mode is defined as the vertically averaged height perturbation between the 300 and 700 hPa levels, given by eq. (11). We have calculated the mean of the absolute value of the modes over the globe. This measure of the baroclinicity is not able to distinguish between an equivalent barotropic perturbation, with no heat transport, and a growing, vertically sloping perturbation with a non-zero heat transport. This diagnostic quantity therefore only gives information about the relative contribution to the barotropic and baroclinic parts of the streamfunction as discussed in (Vallis, 2006); it does not detect the presence of baroclinic instability. To complement this measure, we have also computed the energy conversions involved in the perturbation growth.

The singular vector perturbations are initially dominated by baroclinic structures. This is clearly seen in the ratio between the baroclinic and barotropic modes (Fig. 6, solid line). The baroclinic structures initially contain potential energy. The potential energy is quickly transformed into kinetic energy during the first day of integration (Fig. 7, solid line). Note that these clear baroclinic properties appear even though the singular vector perturbations used here are a combination between initial and evolved singular vectors (see Section 2).

For forecast times longer than the optimization time for the singular vectors, the perturbations are more barotropic in their structure.

The properties of the ET perturbations initially are similar to the short-forecast perturbations (lead time 12–48 h), both when studying the shape of the spectrum (Fig. 8, dash-dotted line) and the energy ratio (Fig. 7, dash-dotted line). This is related to the recycling of perturbations when generating the ET perturbations, yielding initial perturbations with the same properties as the short-forecast perturbations. For the baroclinicity in the perturbations, the ratio (Fig. 6, dash-dotted line) increases during the first days when the perturbations organize into baroclinic structures and reaches a maximum after one and a half day.

For the RF perturbations, both when studying the shape of the spectrum (Fig. 8, dashed line) and the ratio between the baroclinic and barotropic modes (Fig. 6, dashed line), we see similarities between the initial perturbations and properties for 10-d forecast perturbations. This is because the RF perturbations have the same properties initially as the atmospheric variability (also found in Schubert and Suarez, 1989). Because long-forecast perturbations approach the structure of the atmospheric variability, the initial and 10-d forecast perturbations should show similarities.

The RF perturbations have low amplitude for short waves in the initial spectra compared to ET (and also SV). This is because the perturbations obtained by the recycling procedure (ET) are more close to the scale of short-range forecast errors than the atmospheric variability. The difference between short and long forecasts is illustrated in Boer (2003), who studied the spectra of the forecast error for different forecast lengths. During the first 48 h of integration, disturbances of small scales grow rapidly with the RF perturbations (grey dashed) and obtain the same amplitude as with ET (grey dash-dotted) and SV (grey solid).

Another property seen in the spectra for the RF perturbations is that wavenumber 1 has a larger amplitude than in the other two experiments. This is because the two analyses from random dates could be in somewhat different seasonal phases, which could yield differences in mass-balance and temperature between the hemispheres.

For the energy ratio (Fig. 7, dashed line), the RF perturbations initially contain a larger portion of potential energy than ET-EPS, but both experiments converge during the first 2 d of integration. The difference in the energy ratio could be related to the differences in the spectra. Generally, structures with large spatial scales favour potential energy, due to the fact that (geostrophically) the wind is the derivative of the geopotential (yielding greater value for high wavenumbers in quadratic sense). When small-scale structures grow in RF-EPS the energy ratio also decreases. This initial transformation, together with the higher perturbation growth for RF-EPS compared to ET-EPS, indicates a period of transient perturbation growth (Trevisan, 1993) for RF-EPS. It was recently discussed in Magnusson et al. (2008a) that perturbations randomly distributed in vicinity of the attractor of the Lorenz 63 system shows a perturbation growth faster than the breeding modes during the first steps of integration. If model errors in a more realistic atmospheric model have an initial error growth different to the exponential approximation (Lorenz, 1982) is recently discussed in Judd et al. (2008).

Regarding the ratio between the baroclinic and barotropic modes for RF perturbations (Fig. 6, dashed line), the initial perturbations are barotropic, due to the large-scale of the perturbations. During the first days, the perturbations become more baroclinic and approach the ratio of ET when small-scale structures develop (see above). This development of baroclinic structures is in line with results from an earlier study by Källén and Huang (1988). They found that a barotropic perturbation, obtained by assimilating isolated surface pressure observations, could yield growing perturbations using the ECMWF forecasting system. Especially in the vicinity of strong temperature gradients, a fast perturbation growth is obtained even if the perturbation structures are not explicitly connected to growing modes.

During the first days of integration, the properties of all three experiments converge. This could be related to the ‘preferred structures’ discussed in Reinhold (1986). By using a two-layer model and studying the phase difference and amplitude ratio between the streamfunctions in each layer he found, if dissipation is included in the model, that an arbitrary initial perturbation evolves asymptotically to an externally prescribed phase difference between the streamfunction for the two layers. This phase difference is related to the ratio between the modes we have studied. Therefore, a baroclinic structure in medium-range forecasts should be independent of the initial structure.

For longer forecast times, the perturbations for all experiments develop towards larger scale structures, yielding a larger portion of potential energy and a more barotropic structure. The spectrum obtains the same slope and amplitude as +240 h for longer and longer waves as the forecast time increases as seen in Schubert and Suarez (1989). This indicates that the predictability is lost on these scales in the same manner as discussed in Lorenz (1969). One remaining difference between the experiments in the medium range is the high amplitude in wavenumber 1 for

the RF perturbations. This difference remains until days 6–7 (not shown in Fig. 8).

To test the significance of the differences between the experiments for the energy ratio and for ratio of the baroclinic and barotropic modes, the same test as for the Eady index (see above) is applied. The difference in the energy ratio between ET-EPS and RF-EPS is statistically significant until day 2, between SV-EPS and RF-EPS for the initial time step and around day 3 and between SV-EPS and ET-EPS until day 3.5. For ratio of the baroclinic and barotropic modes, the difference between RF-EPS and ET-EPS shows no statistical significance. Both RF-EPS and ET-EPS are significantly different from SV-EPS until day 3.5.

5. Ensemble skill

To evaluate the skill of the ensemble forecast, we have calculated skill scores for several parameters for 90 forecast cases (1 December 2005–28 February 2006). The skill scores we used are root mean square error (RMSE) and anomaly correlation coefficient (ACC) of the ensemble mean. For the probabilistic forecasts, we calculated the ranked probability skill score (RPSS), area under relative operating characteristics (ROCA) and ignorance skill score (ISS). The parameters evaluated are z500, z1000, t850 and v200. The area Northern Hemisphere refers to the area north of 20°N and Southern Hemisphere south of 20°S.

To test whether the difference in scores between the experiments is statistically significant, we have used a moving-block bootstrap method for the paired data. The sample of N cases is resampled with replacement to obtain M different samples of N cases. For our calculation, we have used $M = 5000$. The resampling uses blocks of L consecutive dates to account for the temporal correlation of the score differences. The block size L depends on forecast lead-time. The block length is set using an AR-2 model following Wilks (1997). The empirical distribution of the difference of the average statistic (e.g. RPSS or RMSE) is computed. Then, the confidence interval is determined as the x th and $(100 - x)$ th percentile of the empirical distribution. To determine if the difference between the experiments significantly differs from zero we have used the 99% confidence limit. The significance test has been applied for all scores except the ROC area

5.1. RMSE of ensemble mean and ensemble spread

Figure 9a shows the RMSE of the ensemble mean and the standard deviation of the ensemble (often referred to as ensemble spread) for z500 in the Northern Hemisphere. We see that RF-EPS and ET-EPS start with a much larger initial spread than SV-EPS. For RF-EPS, the spread grows very slowly during the first 12 h. This is mainly because the potential energy in the perturbations grows very slowly for the first time steps while

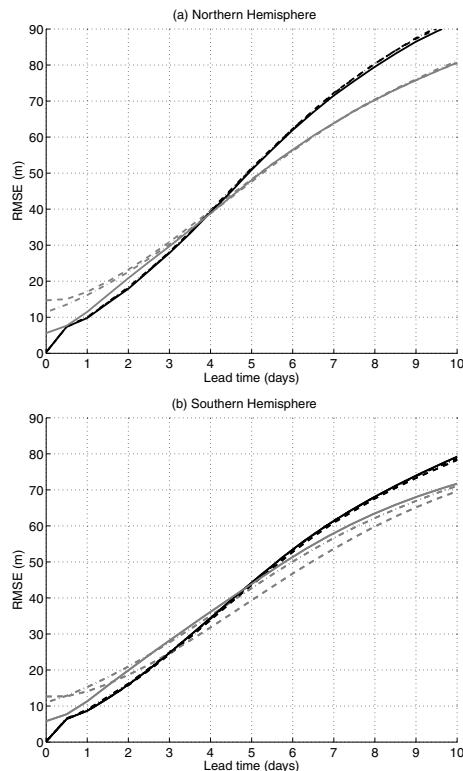


Fig. 9. Ensemble mean rms error (black) and ensemble standard deviation (grey) for z500. The ET-EPS (dash-dotted), RF-EPS (dashed) and SV-EPS (solid). 1 December 2005–28 February 2006.

the kinetic energy grows fast as seen in Fig. 7 and discussed in Section 4.2. As discussed in Section 4.1, the RF perturbations grow faster than ET perturbations in total perturbation energy sense. But after 12 h, the perturbations in z500 start to grow with the same rate as ET perturbations. The spread of SV-EPS grows much faster than the other two during the first 3 d. This is due to the baroclinic structures, as discussed in Section 4.2. After 3 d all experiments have the same spread (due to the objective of the tuning strategy, see Section 3) and grow with the same rate. Regarding the RMSE of the ensemble mean, it shows somewhat better results for SV-EPS for all forecast lengths. But the difference to RF-EPS is only significant forecast days 1–2 and to ET-EPS forecast days 1–4. The RF-EPS is significantly better than ET-EPS for forecast days 1–3.

Figure 9b shows the same as Fig. 9a for the Southern Hemisphere. The ensemble spread growth for the three experiments shows the same properties as for the Northern Hemisphere. But for the medium-range (days 3–10) forecast times, there are differences between the experiments in terms of spread. The SV-EPS has the highest ensemble spread and RF-EPS the lowest. One reason for the differences in spread in the Southern Hemisphere between the experiments is that the initial amplitude for the ET-EPS and RF-EPS has been tuned to yield the same spread as SV in the Northern Hemisphere for medium-range forecast

times. Comparing RMSE in the Southern Hemisphere, RF-EPS is significantly better than SV-EPS for days 2–4. The SV-EPS is significantly better than ET-EPS for days 1–2 and RF-EPS is significantly better than ET-EPS for days 1–4. For longer forecast lengths, RF-EPS shows lower RMSE than the other two experiments, but the differences are not significant.

For all experiments we see that the spread is lower than the RMSE for ensemble mean for medium-range forecasts. For a ‘perfect’ ensemble, the ratio should be one-to-one (Leutbecher and Palmer, 2008). This indicates that the atmospheric model is under-dispersive and has a negative effect on the ensemble spread. For further discussion see Bengtsson et al. (2008).

5.2. Ranked probability skill score (RPSS)

The RPSS (Candille and Talagrand, 2005) is based on the Brier skill score for (in this study) 10 climatologically equally likely categories. The overall conclusion when evaluating RPSS for the three experiments is that the results are similar. This is illustrated in Fig. 10 where the RPSS for t850 is shown. For the Northern Hemisphere, there is a slight but statistically significant advantage for RF-EPS and ET-EPS for the first 2 d compared to SV-EPS. For a longer forecast length, SV-EPS has slightly better performance but the difference is only statistically significant for

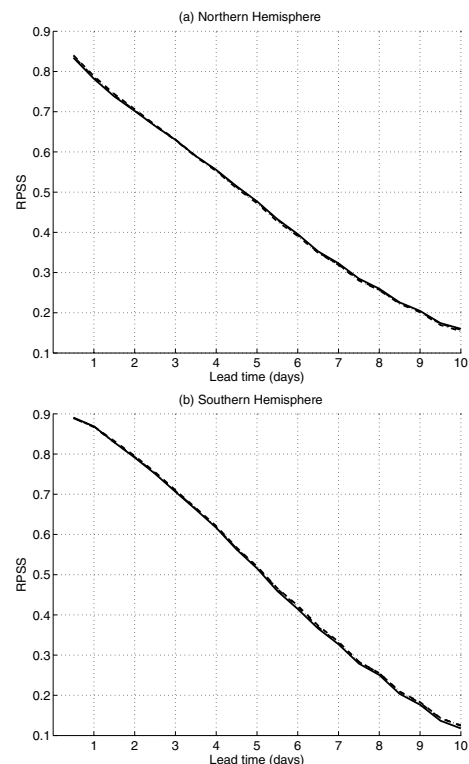


Fig. 10. Ranked probability skill score (based on 10 climatologically equally likely categories) for t850 for ET-EPS (dash-dotted), RF-EPS (dashed) and SV-EPS (solid). 1 December 2005–28 February 2006.

one time step between SV-EPS and ET-EPS and never between SV-EPS and RF-EPS. For the Southern Hemisphere, a slight advantage for ET-EPS and RF-EPS over SV-EPS is present for all forecast lengths (statistically significant days 1–3).

These results for RPSS can be compared with the results from the study by Magnusson et al. (2008b), where SV-EPS had a slight but statistically significant advantage compared to breeding ensembles for both hemispheres. Comparing ET-EPS with masked breeding (not shown) we found an improved performance using the ET method, as also earlier found in Wei et al. (2008).

Evaluating RPSS for the other parameters, we found an advantage for SV-EPS in the Northern Hemisphere for $z1000$ (not shown). The difference to RF-EPS is statistically significant for days 4 and 7 and to ET-EPS for day 4. The difference in the Northern Hemisphere could be related to the high perturbation amplitude in the planetary waves for RF perturbations affecting the ensemble unfavourably in the medium range. For the Southern Hemisphere, the differences are opposite, RF-EPS shows a better performance compared to SV-EPS (significant for days 0–2). For $z500$, the results are in line with $t850$ except for day 1 where we can see a clear advantage for SV-EPS. This is because ET-EPS and in particular RF-EPS is over-dispersive in $z500$ initially, and this affects the RPSS negatively.

The Brier score can be decomposed into a reliability and a resolution term (Murphy, 1973; Wilks, 2006). The reliability term summarizes the statistical consistency between the predicted and observed probabilities and the resolution term indicates how well the forecast resolves the variations in the observations. Using the climatological probability of an event will result in ‘perfect’ reliability but very bad resolution. Studying the reliability and resolution terms of the Brier scores for the different experiments (not shown), the reliability and the spread-skill ratio agree well. For $t850$ and a short-forecast length (up to day 2), RF-EPS and ET-EPS show a better reliability than SV-EPS. This is because SV-EPS is under-dispersive for $t850$ and short lead times. The opposite holds true for $z500$, when especially RF-EPS is over-dispersive for the first day of integration. Regarding the resolution, the differences are small for the Northern Hemisphere. For the Southern Hemisphere and the medium range, SV-EPS has a better reliability (due to better spread-skill ratio) while the resolution term shows worse results than the other two experiments. It results in a slight advantage in RPSS for ET-EPS and RF-EPS in the Southern Hemisphere as discussed above.

5.3. Ignorance skill score (ISS) and ROC area

The clearest differences between the three experiments are found when evaluating the ISS (Roulston and Smith, 2002). The ISS is calculated for six different probabilistic events. Figure 11a shows ISS for $t850$ anomaly < -0.5 standard deviation for the Northern Hemisphere. For this anomaly, we see that RF-EPS has the best performance for the first 2–3 d followed by ET-EPS. This

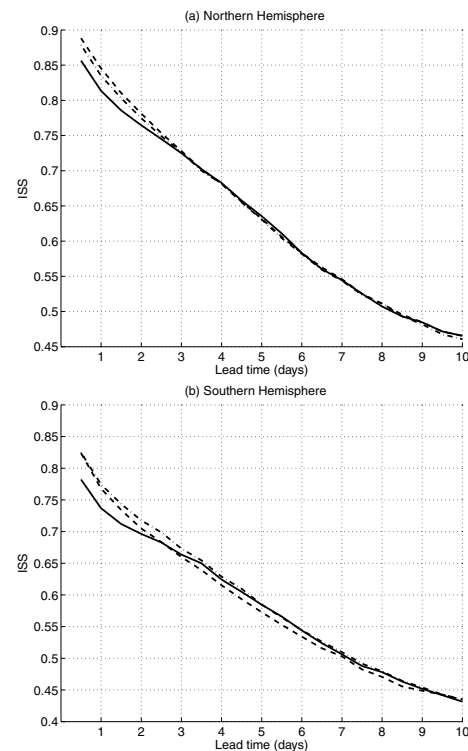


Fig. 11. Ignorance skill score for $t850$ anomaly < -0.5 standard deviation on the Northern Hemisphere and Southern Hemisphere for ET-EPS (dash-dotted), RF-EPS (dashed) and SV-EPS (solid). 1 December 2005–28 February 2006.

difference between the experiments is statistically significant and appears in all evaluated events. One explanation of the results could be that the ISS is sensitive to under-dispersive ensembles. This is the case for SV-EPS ($t850$) during the first forecast days, which could partly explain the lower scores. For medium-range forecasts, ET-EPS shows the lowest ISS for positive anomalies and for negative anomalies all experiments are equal.

Evaluating ISS for the Southern Hemisphere (Fig. 11b), the ET-EPS shows the best performance for days 1–3 followed by RF-EPS. In contrast to the Northern Hemisphere, RF-EPS shows somewhat worse results than the other two experiments for forecast days 3–6. This could be explained by the fact that RF-EPS is under-dispersive in the Southern Hemisphere for the medium range.

For the ROC area (not shown), the differences show the same properties on the Northern Hemisphere as seen for ISS. Evaluating the ROC area for the Southern Hemisphere, ET-EPS and RF-EPS show an advantage over SV-EPS for almost all evaluated events.

5.4. Tropics

Due to the absence of geostrophy in the equatorial region, the mass field (temperature and geopotential) has low variability.

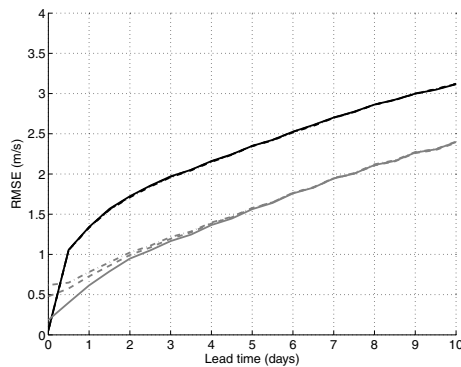


Fig. 12. Ensemble mean rms error (black) and ensemble standard deviation (grey) for u850 and the tropics. The ET-EPS (dash-dotted), RF-EPS (dashed) and SV-EPS (solid). 1 December 2005–28 February 2006.

Therefore, the ensemble systems have been compared for the zonal wind component (u) at 850 hPa for the tropical region. For SV-EPS, tropical singular vectors are only computed in the vicinity of tropical cyclones, therefore low initial spread is expected for the tropics (see Section 2.1 and Magnusson et al., 2008b). Figure 12 shows the RMSE and spread for u850 in the tropical region (20°S – 20°N). Here, all three experiments have too low ensemble spread. The spread for SV-EPS is initially lowest (as expected) but approaches the spread for other two experiments and becomes equal after day 6 when perturbations from the mid-latitudes have propagated into the tropics. The highest initial spread is obtained for ET-EPS. This is mainly because of the use of a regional rescaling mask. The relatively low initial amplitude for RF-EPS initially is because the amplitude of the atmospheric variability is largest in the mid-latitudes even if the short-range forecast errors have a large amplitude in the tropics.

The differences in RMSE are small. The lowest RMSE for short-forecasts length is yielded by ET-EPS. The difference to SV-EPS and RF-EPS is statistically significant until day 4. The RF-EPS has an advantage over SV-EPS until day 3.

Evaluating the RPSS for u850 in the tropics (not shown) a small but statistically significant advantage is present for ET-EPS compared to RF-EPS for the first 3 d. Compared to SV-EPS both ET-EPS and RF-EPS shows an advantage for the first 6 d. But as mentioned earlier the performance of the SV-EPS is very dependent on the configuration of the tropical singular vectors and the ET-EPS is dependent on the regional rescaling mask.

5.5. ET with simplex transformation

In Wang et al. (2004), the superiority of using simplex transformation instead of using plus-minus symmetric perturbations is shown for ETKF. To quantify the difference between ET with plus-minus symmetry and simplex transformation in the present study an experiment using ET with simplex transformation has

been performed for the period 1 December 2005–15 January 2006. Studying the variance spectra (as in Magnusson et al., 2008b) a clear difference between the experiments is present initially (not shown). The ET experiment using simplex transformation has variance in 19 orthogonal modes of 20 compared to 10 for the plus-minus experiment. After 48 h, the difference between the experiments is much smaller but the variance spectrum for the simplex transformation experiment is still flatter and it remains so also after 72 h, but after 120 h almost no difference remains. Evaluating the RPSS for t850 for the two experiments only small and statistically insignificant differences are present.

6. Conclusions

In this study, we have compared two existing flow-dependent perturbation methods, singular vectors and ET perturbations, with a method using random perturbations obtained by rescaling the difference between atmospheric analyses from two random dates (RF perturbations). The results for medium-range forecasts are in general very similar for the three different ensemble techniques, despite the simplicity of the RF technique.

To obtain a similar ensemble spread for medium-range forecasts, it is necessary to tune the initial amplitude of the perturbations differently for the different experiments. Singular vectors need a low initial amplitude because of the rapid initial perturbation growth characterizing the method. The RF perturbations need to start with a larger initial amplitude for z500 because of the slow perturbation growth for z500 during the first forecast steps. But in total energy sense, the RF perturbations grow faster than the ET perturbations during the first days.

For ET-EPS, the initial amplitude was set to 90% of the estimated analysis error standard deviation. One could expect that the initial amplitude should be equal to the estimated analysis error but because the initial spread for neither ET-EPS nor for any other method used here resembles the analysis error variance perfectly between regions and variables, we had to use a lower value to obtain a proper spread for z500 in the Northern Hemisphere. This implies that, for example, the wind perturbations in the tropics are too low compared to the analysis error variance.

To increase the amplitude in the tropics, a regional rescaling mask was used for the ET ensemble, following the setup at NCEP (Wei et al., 2008). The disadvantage of using a regional rescaling mask is that it introduces imbalances in the model (this effect is not investigated in the present study). Another way of increasing the amplitude in the tropics is to introduce perturbations that represent model errors. Using a representation of model errors due to unresolved scales, McLay et al. (2008) obtained a variance in the ET perturbation in better agreement with the analysis error variances in the tropics. In our study, we have used the stochastic physics scheme to represent model errors. We concluded that, for the ECMWF NWP model, a regional rescaling mask is also needed (This is discussed further in Magnusson et al. (2008b) for

a BV ensemble). The effects of different approaches to account for model errors are investigated in Houtekamer et al. (2008) and Reynolds et al. (2008).

Another uncertainty in the ET calculation is the dependence on the quality of the analysis error variance obtained from the data assimilation system. This issue needs to be studied further. Regarding all three perturbation strategies, it should be possible to find an initial spread yielding a better ensemble in terms of any verification score. All possible values of the initial spread have not been tested.

The RF perturbations initially show a different spectral signature for short waves than the other methods. This could be explained by the fact that the spectral signature for short-forecast perturbations (represented in the ET method) and the atmospheric variability (represented by the RF method) are different. The analysis uncertainty, which is desirable to sample, should be closer to the short-forecast perturbations than the atmospheric variability. Therefore, one could have expected that the initial spectral distribution for ET-EPS and SV-EPS should affect the quality of the short-range forecasts favourably. This seems not to be the case, at least after 12 h, which is our first evaluation step. One reason could be that the probability to sample the 'true' analysis uncertainties is very low for the shortest scales.

Initially all three experiments show differences in the geographical distribution of the perturbations and the connection to baroclinic structures. For example, the singular vectors yield very baroclinic perturbations and RF perturbations are more barotropic. But after the first forecast days, we see that the properties of the different experiments converge. Also in the localization, we see differences initially, the experiments converge during the first days. This indicates that the model itself 'decides' which perturbations that will grow and that the choice of perturbation method only has a minor influence after the first few days of a forecast.

Evaluating the quality of the experiments, we see almost no significant difference between them for medium-range forecasts for the Northern Hemisphere. The clearest difference, for t850, is found for the first 2 d when RF-EPS shows the best performance in the sense of probabilistic scores (RPSS, ISSs and ROC area), followed by ET-EPS, while the singular vector experiment shows the worst performance. The fact that only small differences appear in the medium range indicates that uncertainties caused by model errors dominate over the initial uncertainties, as also discussed in, for example, Descamps and Talagrand (2007).

Regarding the tropics, all three experiments are under-dispersive for medium-range forecasts. Initially, ET-EPS has the largest ensemble spread and therefore an advantage in RPSS. When the spread for the different experiments converges also the RPSS converges. For SV-EPS, the low spread is due to the configuration of the singular vectors. The low spread for the RF in the tropics could be seen as a potential area of development of the method.

Our general conclusion is that the RF method gives at least as good results as either SV or ET. One possible explanation is that the perturbations are not localized to a few local maxima around the globe. Singular vectors are originally localized but by combining several SV with the Gaussian sampling more global perturbations are obtained. For the ET method, the perturbations seem to be confined to only a few local maxima.

The main difference between our experiment with the ET method and the one used operationally at NCEP is their use of a simplex transformation to obtain an ensemble centred around the unperturbed analysis. To evaluate the impact of the centring method, we have made additional simulations using simplex transformation for the ET method. These show only small and not statistically significant differences in the scores for the ET method. We have not investigated the differences for SV and RF using simplex transformation and that is the reason why we used the same centring method for all experiments in this study.

We have not attempted to develop the RF method further. It should be possible to find a more selective sampling of analyses to improve the method. It should also be possible to find a method to avoid the unexpectedly high amplitude in wavenumber 1, or finding random perturbations that have a spectrum closer to the analysis error uncertainty.

The RF method is not primarily meant as a new way of generating ensembles in operational forecasting. It should rather be regarded as a baseline method that provides a benchmark when assessing the performance of more sophisticated methods. By taking the properties of the current flow situation into account, it should be possible to obtain better results than with RF ensembles. The main conclusion of the present work is that both ET and SV fail to meet this criterion. Thus, until better methods have been devised, RF is in fact a viable alternative, since it is simpler to use (and at least cheaper than singular vector computations) and performs as well (in some respects even better) than the other methods.

7. Acknowledgments

We express our gratitude to Martin Leutbecher, ECMWF, for providing the evaluation software and for valuable discussions about our project. We acknowledge the Swedish Meteorological and Hydrological Institute, SMHI, for allowing us to use their computer allocation at ECMWF. We are also grateful to anonymous reviewers for providing thoughtful and constructive criticism.

References

- Anderson, J. L. 1997. The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: low-order perfect model results. *Mon. Wea. Rev.* **125**, 2969–2983.
- Arribas, A., Robertson, K. B. and Mylne, K. R. 2005. Test of a poor man's ensemble prediction system for short-range probability forecasting. *Mon. Wea. Rev.* **133**, 1825–1839.

- Barkmeijer, J., van Gijzen, M. and Bouttier, F. 1998. Singular vectors and estimates of the analysis-error covariance metric. *Quart. J. R. Meteorol. Soc.* **124**, 1695–1713.
- Barkmeijer, J., Buizza, R. and Palmer, T. N. 1999. 3D-Var Hessian singular vectors and their potential use in ECMWF Ensemble Prediction System. *Quart. J. R. Meteorol. Soc.* **125**, 2333–2351.
- Barkmeijer, J., Buizza, R., Palmer, T. N., Puri, K. and Mahfouf, J.-F. 2001. Tropical singular vectors computed with linearized diabatic physics. *Quart. J. R. Meteorol. Soc.* **127**, 685–708.
- Bengtsson, L. K., Magnusson, L. and Källén, E. 2008. Independent estimations of the asymptotic variability in an ensemble forecast system. *Mon. Wea. Rev.* **136**, 4105–4112.
- Bishop, C. H. and Toth, Z. 1999. Ensemble transformation and adaptive observations. *J. Atmos. Sci.* **56**, 1748–1765.
- Boer, G. J. 2003. Predictability as a function of scale. *Atmos.-Ocean* **41**, 203–215.
- Bowler, N. 2006. Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus* **58A**, 538–548.
- Bowler, N., Arribas, A., Mylne, K. R., Robertson, K. B. and Beare, S. E. 2008. The MOGREPS short-range ensemble prediction system. *Quart. J. R. Meteorol. Soc.* **134**, 703–722.
- Buizza, R. 2008. Comparison of a 51-member low-resolution (TL399L62) ensemble with a 6-member high-resolution (TL799L91) lagged-forecast ensemble. *Mon. Wea. Rev.* **136**, 3343–3362.
- Buizza, R., Miller, M. and Palmer, T. 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. R. Meteorol. Soc.* **125**, 2887–2908.
- Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M. and co-authors. 2005. A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. *Mon. Wea. Rev.* **133**, 1067–1097.
- Candille, G. and Talagrand, O. 2005. Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. R. Meteorol. Soc.* **131**, 2131–2150.
- Descamps, L. and Talagrand, O. 2007. On some aspects of the definition of initial conditions for ensemble prediction. *Mon. Wea. Rev.* **135**, 3260–3272.
- Fisher, M. and Courtier, P. 1995. Estimating the covariance matrices of analysis and forecast error in variational data assimilation, *Technical Memorandum 220*, ECMWF.
- Gustafsson, N. 2007. Discussion on '4D-Var or EnKF?'. *Tellus* **59A**, 774–777.
- Hamill, T. H. and Whitaker, J. S. 2005. Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon. Wea. Rev.* **133**, 3132–3147.
- Hollingsworth, A. 1979. An experiment in Monte Carlo forecasting, In: *Proceedings of Workshop on Stochastic Dynamic Forecasting*, Reading, UK, ECMWF.
- Holton, J. R. 2004. *An Introduction to Dynamic Meteorology*. Academic Press, London, p. 535.
- Hoskins, B. J. and Hodges, K. I. 2002. New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.* **59**, 1041–1061.
- Hoskins, B. J., Buizza, R. and Badger, J. 2000. The nature of singular vector growth and structure. *Quart. J. R. Meteorol. Soc.* **126**, 1565–1581.
- Houtekamer, P. L., Lefaiivre, L., Derome, J., Ritchie, H. and Mitchell, H. 1996. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.* **124**, 1225–1242.
- Houtekamer, P. L., Mitchell, H. and Deng, X. 2008. Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.* in press.
- Judd, K., Reynolds, C. A., Rosmond, T. E. and Smith, L. A. 2008. The geometry of model error. *J. Atmos. Sci.* **65**, 1749–1772.
- Källén, E. and Huang, X.-Y. 1988. The influence of isolated observations on short-range numerical weather forecasts. *Tellus* **40A**, 324–336.
- Leith, C. 1974. Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.* **1974**, 409–418.
- Leutbecher, M. and Palmer, T. N. 2008. Ensemble forecasting. *J. Comput. Phys.* **227**, 3515–3539.
- Lindzen, R. S. and Farrell, B. 1980. A simple approximate result for the maximum growth rate of baroclinic instabilities. *J. Atmos. Sci.* **37**, 1648–1654.
- Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.
- Lorenz, E. N. 1965. A study of the predictability of a 28-variable atmospheric model. *Tellus* **67**, 321–333.
- Lorenz, E. N. 1969. The predictability of a flow which possesses many scales of motion. *Tellus* **21**, 289–307.
- Lorenz, E. N. 1982. Atmospheric predictability experiments with a large numerical model. *Tellus* **34**, 505–513.
- Magnusson, L., Källén, E. and Nycander, J. 2008. Initial state perturbations in ensemble forecasting. *Nonlinear Processes Geophy.* **15**, 751–759.
- Magnusson, L., Leutbecher, M. and Källén, E. 2008. Comparison between singular vectors and breeding vectors as initial perturbations of ECMWF Ensemble Prediction System. *Mon. Wea. Rev.* **136**, 4092–4104.
- McLay, J. G., Bishop, C. H. and Reynolds, C. A. 2008. Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.* **136**, 1093–1108.
- Mureau, R., Molteni, F. and Palmer, T. 1993. Ensemble prediction using dynamically conditioned perturbations. *Quart. J. R. Meteorol. Soc.* **119**, 299–323.
- Murphy, A. H. 1973. A new vector partition of the probability score. *J. Appl. Meteorol.* **12**, 595–600.
- Palmer, T. N. 1993. Extended-range atmospheric prediction and the Lorenz Model. *Bull. Am. Meteorol. Soc.* **74**, 49–66.
- Reinhold, B. 1986. Structural determinism of linear baroclinic waves and simple nonlinear equilibration. *J. Atmos. Sci.* **43**, 1484–1504.
- Reynolds, C. A., Teixeira, J. and McLay, J. G. 2008. Impact of stochastic convection on the ensemble transform. *Mon. Wea. Rev.* **136**, 4517–4526.
- Roulston, M. S. and Smith, L. A. 2002. Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.* **130**, 1653–1660.
- Schubert, S. D. and Suarez, M. 1989. Dynamical predictability in a simple general circulation model: average error growth. *J. Atmos. Sci.* **46**, 353–370.
- Szunyogh, I., Kostelich, E. J., Gyarmati, G., Kalnay, E., Hunt, B. R., and co-authors. 2008. A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus* **60A**, 113–130.
- Talagrand, O., Vautard, R. and Strauss, B. 1999. Evaluation of probabilistic prediction system, In: *Proceedings of Workshop on Predictability*, October 1997, ECMWF.
- Talagrand, O., Candille, G. and Descamps, L. 2007. On some aspects of validation of probabilistic prediction, In: *Proceedings of Workshop on Ensemble prediction*, November 2007, ECMWF.

- Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NCEP: the generation of perturbation. *Bull. Am. Meteorol. Soc.* **74**, 2317–2330.
- Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.* **125**, 3297–3319.
- Trevisan, A. 1993. Impact of transient error growth on global average predictability measures. *J. Atmos. Sci.* **50**, 1016–1028.
- Tribbia, J. J. and Baumhefner, D. P. 1988. Estimates of the predictability of low-frequency variability with a spectral general circulation model. *J. Atmos. Sci.* **45**, 2306–2317.
- Vallis, K. G. 2006, *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press, Cambridge, p. 745.
- Wang, X. and Bishop, C. 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**, 1140–1158.
- Wang, X., Bishop, C. and Julier, S. 2004. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.* **132**, 1590–1605.
- Wei, M. and Toth, Z. 2003. A new measure of ensemble performance: perturbation versus error correlation analysis (PECA). *Mon. Wea. Rev.* **131**, 1549–1565.
- Wei, M., Toth, Z., Wobus, R. and Zhu, Y. 2008. Initial perturbations based on the ensemble transform (ET) technique in the NCEP global ensemble forecast systems. *Tellus* **60A**, 62–79.
- Wilks, D. S. 1997. Resampling hypothesis tests for autocorrelated fields. *J. Clim.* **10**, 65–82.
- Wilks, D. S. 2006, *Statistical Methods in the Atmospheric Sciences* 2nd Edition. Academic Press, London, p. 627.
- Zhang, F. 2005. Dynamics and structure of mesoscale error covariance of a winter cyclone estimated through short-range ensemble forecasts. *Mon. Wea. Rev.* **133**, 2876–2893.