Growth of Forecast Errors from Covariances Modeled by 4DVAR and ETKF Methods

CHIARA PICCOLO

Met Office, Exeter, United Kingdom

(Manuscript received 3 August 2009, in final form 29 July 2010)

ABSTRACT

Numerical weather forecasting errors grow with time. Error growth results from the amplification of small perturbations due to atmospheric instability or from model deficiencies during model integration. In current NWP systems, the dimension of the forecast error covariance matrices is far too large for these matrices to be represented explicitly. They must be approximated.

This paper focuses on comparing the growth of forecast error from covariances modeled by the Met Office operational four-dimensional variational data assimilation (4DVAR) and ensemble transform Kalman filter (ETKF) methods over a period of 24 h. The growth of forecast errors implied by 4DVAR is estimated by drawing a random sample of initial conditions from a Gaussian distribution with the standard deviations given by the background error covariance matrix and then evolving the sample forward in time using linearized dynamics. The growth of the forecast error modeled by the ETKF is estimated by propagating the full nonlinear model in time starting from initial conditions generated by an ETKF. This method includes model errors in two ways: by using an inflation factor and by adding model perturbations through a stochastic physics scheme. Finally, these results are compared with a benchmark of the climatological error.

The forecast error predicted by the implicit evolution of 4DVAR does not grow, regardless of the dataset used to generate the static background error covariance statistics. The forecast error predicted by the ETKF grows more rapidly because the ETKF selects balanced initial perturbations, which project onto rapidly growing modes. Finally, in both cases it is not possible to disentangle the contribution of the initial condition error from the model error.

1. Introduction

Estimates of forecast error covariances are at the heart of any data assimilation system and yet the way they are modeled in any operational assimilation scheme is limited by the compromises made for practical implementation and the available knowledge of the statistical properties of the forecast error. In most operational assimilation schemes the forecast error covariance is assumed stationary, homogeneous, and isotropic to overcome the difficulty of estimating the full covariance matrix. These assumptions fail to characterize the true forecast error covariance.

The purpose of this paper is to compare the growth of forecast errors from covariances modeled by the Met Office operational four-dimensional variational data assimilation (4DVAR) and ensemble transform Kalman filter (ETKF) methods over 24 h and examine whether the contribution of the initial condition error can be separated from the model error contribution. This is achieved by examining the underlying assumptions made to generate the data samples required to define the forecast error covariance in both methods and finally comparing the forecast error growth predicted by these two methods with the climatological error growth.

The growth of forecast errors implied by 4DVAR is estimated by drawing a random sample of the initial conditions from the Met Office operational background error covariance and then evolving that sample forward in time using a linearized version of the full model. This is valid under the assumption that the error growth is linear and the model error is negligible, which implies that only the initial condition errors are evolved by 4DVAR. The relaxation of the perfect model assumption, to include a model error contribution within the framework of 4DVAR, is now of great interest but it is still difficult in practice (Trémolet 2006; Carrassi and Vannitsem 2010).
The forecast error growth predicted by the implicit evolution of the background error covariance within 4DVAR is compared with the growth from stationary ensemble-based forecast error covariances generated using the Met Office Global and Regional Ensemble Prediction System (MOGREPS). These covariances are estimated by propagating model states with a full nonlinear model rather than under the assumption of linearity. MOGREPS initial conditions are generated using an ETKF, which includes model error through a variable inflation factor to compensate for the underspread ensemble and by a stochastic physics scheme to represent the effects of model uncertainties. When the ensemble size is small, the forecast error covariance tends to be noisy and to include spurious long-distance correlations. This requires localization methods, which were found to be a cause of imbalances (Mitchell et al. 2002; Hamill et al. 2001).

As a benchmark of forecast error growth, the climatological estimate of the forecast error statistics is obtained by accumulating statistics of the forecast error measured against the verifying analysis over a 1-month period. At the beginning, the forecast error is zero by definition while at later times it includes both the contribution of the initial condition error and the model error.

In summary, the covariances modeled by 4DVAR and ETKF methods rely on different assumptions and approximations. The 4DVAR approach relies on the perfect model assumption mainly for reasons of complexity and expense within an operational framework. In the ETKF the model error is included and its specification is crucial for the performance of the system. The ETKF is one of the many different ensemble Kalman filter (EnKF) formulations. For a detailed description of differences between EnKF and 4DVAR approaches, we refer to Kalnay et al. (2007) and Lorenc (2003b).

The present paper is organized as follows. In section 2 we recall the various assumptions and limitations of the different methods and describe the relevant techniques to estimate the forecast error statistics. In section 3 we present the results for temperature at 500 hPa and for zonal winds at 850 and 250 hPa; we also discuss possible explanations for differences in forecast error evolution and magnitude and the potential implications of these results for improving either the Met Office operational 4DVAR or MOGREPS. Finally, in section 4 we summarize the conclusions of this study.

2. Formulation of forecast error covariances

a. Forecast error covariances in 4DVAR

Variational data assimilation (VAR) schemes are statistical methods of estimating the state of the atmosphere for weather prediction by minimizing a cost function. This cost function combines in an optimal way the a priori information of the atmospheric state, called the background state, with new information from a variety of observations over an assimilation period. In operational practice, the background is typically a 6-h short-range forecast from the previous analysis. An optimal analysis relies on accurate estimates of covariance matrices (Lorenc 1986). In this paper we will focus on the Met Office 4DVAR in its incremental form, which became operational in October 2004. For details we refer to Rawlins et al. (2007) and Lorenc and Rawlins (2005). The 4DVAR method minimizes a four-dimensional cost function in which a forecast model is used as a constraint to propagate the initial model state to the times of the observations. The minimization is performed at a lower resolution than the main forecast, with simpler physics and using a cost function evaluated by integrating a linearized forecast model rather than the full nonlinear model. Details on 4DVAR and its incremental formulation can be found in Courtier et al. (1994). Within each assimilation window the model is assumed to be perfect; background and observation errors are assumed to be mutually uncorrelated and unbiased, so that they can each be represented by a zero mean Gaussian distribution.

The perfect model assumption is limited by the model resolution and by the physical parameterizations in resolving exactly the governing equations. On the other hand the perfect model constraint reduces the complexity of 4DVAR to produce an operationally feasible computational algorithm. This is because, under this assumption, the evolution of the atmosphere is entirely determined by the initial conditions. There is interest in relaxing the perfect model constraint and allowing for model error, although it is not clear how to separate model error from other sources of error. Moreover, important simplifications are required anyway to estimate the model error contributions since there is not enough information available to get meaningful statistics (Trémolet 2006).

The assumption of unbiased errors is deficient in practice because there often are significant biases in the background fields and in the observations. The hypothesis of uncorrelated errors is usually valid since error sources in the background and in the observations are supposed to be completely independent. An exception to this is seen in the correlations between observation errors (especially for satellite data), which are usually ignored.

The assumption of linearity is strong. Incremental 4DVAR relies on the linearization of weakly nonlinear operators, at the expense of the optimality of the analysis. For strongly nonlinear problems, there is no general and simple way to calculate the optimal analysis.

In current operational meteorological models, the dimension of the model state is of order $10^7$, and the
dimension of the observation vector is of order $10^6$ per analysis. 4DVAR requires the specification of covariance matrices of the background errors and observational errors, respectively, of order $10^7$ and $10^6$. The elements of the matrices are estimated statistically. A correct specification of observation and background error covariances is crucial to the quality of the analysis, because they determine the weight of the observations and the background fields in the analysis and to what extent the background fields are corrected to match the observations. The observation covariance matrix depends on individual instrument error characterizations, which are usually well known and spatially uncorrelated. This often includes representativeness errors although they depend on the model discretization and not on instrumental problems. Estimation of the background error covariance is generally more difficult. Its full structure cannot be determined or even stored with modern computers, and, even if this was possible, there is insufficient statistical information to determine all its elements (Dee 1995).

It is usual to denote background error covariances by $P$:

$$P = \langle (x^b - x)(x^b - x)^T \rangle,$$

assuming that $\langle x^b - x \rangle = 0$. \hfill (1)

Here, the angled brackets denote the expected value of forecast errors $x^b - x$, where $x$ is the true state of the atmosphere, which is unknown, and $x^b$ is the background state. The value of $P$ is unknown and is prohibitive to calculate. For this reason the background error covariance used in 4DVAR, denoted by $B$, is a simplification of $P$. The possible ways to estimate $B$ without the knowledge of the true state $x$ will be discussed in section 2d.

One way to overcome the difficulty of treating such an enormous matrix is to deduce its structure from the knowledge of the atmospheric dynamics and physics and build a model of typical error covariances. The covariance model is based on a sequence of variable transformations, which use physical relationships between variables (e.g., balance relationships), and assumptions of homogeneity and isotropy of the background error to reduce the amount of information needed to characterize $B$ (Lorenc 2003a). This is a big advantage of variational approaches because this set of variable transformations helps to simplify the background term of the cost function and therefore the minimization process. Bannister (2008b) presents a review of modeling the forecast covariance statistics.

4DVAR performs better than 3DVAR. This is widely attributed to the implicit use of an evolving flow-dependent covariance matrix of the background error, the use of observations at correct times, and the propagation of the background departures back to the start of the analysis window (Lorenc and Rawlins 2005). The 4DVAR algorithm itself does not provide an estimate of the analysis covariance matrix, and the forecast error covariance matrix cannot be computed as a source of input for the next cycle of analysis. In practical implementations of 4DVAR, no cycling of covariances takes place and at each analysis cycle the forecast error covariance at the beginning of the assimilation period is replaced by the static background error covariance $B$. However, the covariance matrix $B$ is implicitly propagated in time according to the linearized dynamics to generate flow dependency at later times within the assimilation window. Under the assumptions of linear error growth and small model error, the short-range forecast error covariance within 4DVAR is given by $P = \text{MBM}^T$, where $\text{M}$ is the linearized forecast model integrated over the period of the assimilation. Here, we consider only the linear implementation of 4DVAR in order to replicate the current Met Office operational 4DVAR scheme, which does not yet include multiple outer-loop iterations to take into account nonlinearities of the system (Rawlins et al. 2007).

b. Forecast error covariances in the ETKF

For a linear system, the Kalman filter provides an optimal cycling of error covariances. The Kalman filter evolution of covariances can be divided in an analysis step at time $k$ and a forecast step from time $k$ to $k + 1$:

$$P^a_k = (I - K_k H_k)P^b_k \quad \text{and} \quad P^b_{k+1} = M_{1-k}^{-1}P^a_{k}M_{1-k}^{-1} + Q_{k+1}, \quad \hfill (2)$$

where $P^a_k$ is the covariance matrix of the background error, which is reduced according to the Kalman filter gain matrix during the analysis step, and $H_k$ is the linearized observation operator. In the forecast step, the covariance matrix of the analysis error, $P^a_k$, is propagated in time by the linear forecast model $\text{M}$ over the interval $(k, k + 1)$ as described by Eq. (3). Here, $Q$ represents the covariance matrix of the model error. For details on the Kalman filter method, see Kalman (1960).

It is well known (e.g., Ménard and Daley 1996) that for the linear case and a perfect model ($Q = 0$), the 4DVAR analysis at the end of the assimilation period is equivalent to the Kalman filter analysis over the same interval, given the same inputs and specifically when $P^a_k = B$. However, the Kalman filter is impractical for large systems, since it requires handling $10^7 \times 10^6$ covariance matrices and it has to calculate $10^7$ integrations of the linear forecast model $\text{M}$ to propagate the covariance matrix of the analysis error.

For large systems, like NWP systems, approximate Kalman filters or reduced-rank methods have been developed to restrict the covariance equations to a small
subspace. EnKFs are an example of reduced-rank Kalman filters where the background covariance matrix $P_k^b$ at time $k$ is constructed from a limited sample:

$$
P_k^b \approx \frac{1}{N-1} \sum_{i=1}^{N} [x_k^{b(i)} - \bar{x}_k^b][x_k^{b(i)} - \bar{x}_k^b]^T, \quad (4)
$$

where the index $i$ refers to the ensemble member, $N$ is the number of ensemble members, and $\bar{x}_k^b = (1/N)\sum_{i=1}^{N} x_k^{b(i)}$ is the ensemble mean forecast. In this way the sample covariance matrix represents an estimate of the true covariance matrix restricted to the subspace of the ensemble. The analysis error covariance may be propagated using only $N$ ($N \sim 100$) model integrations and $P_{k+1}^b$ may be projected onto a new subspace to generate an approximate covariance matrix to be used in the next analysis cycle. The EnKF method still assumes that all the errors are Gaussian.

The EnKF is an approximation of the full Kalman filter and it is not optimal, because the analyses are restricted to the space spanned by the ensemble members and the approximated $P_k^b$ presents spurious long-distance correlations. However, ensemble methods are attractive because they do not require the tangent linear or adjoint operators like 4DVAR. In the EnKF method the covariances are estimated using a limited-size random sample equal to the number of the ensemble members and by propagating model states with a fully nonlinear model rather than under the assumption of linearity. The sample of background states is generated from a sample of analysis states, $x_k^{(1)}, \ldots, x_k^{(N)}$:

$$
x_k^{b(i)} = \mathcal{M}_{k-i_k}^{-1}[x_k^{a(i)}], \quad (5)
$$

where the full nonlinear model $\mathcal{M}_{k-i_k}$ is used to propagate the analysis states from $f_k$ to $f_{k+1}$. The ensemble of analysis $x_k^{b(1)}, \ldots, x_k^{b(N)}$ may be generated in different ways.

MOGREPS is an ETKF where the ensemble of analyses is calculated by transforming the background states of the previous cycle (Bowler et al. 2008). Defining the $N$ forecast perturbations by $x_k^{b(i)} - \bar{x}_k^b$, the ETKF method transforms the forecast perturbations into analysis perturbations by a transformation matrix such that the analysis perturbations for each cycle are a linear combination of the forecast perturbations and therefore lie in the subspace of the forecast perturbations. The transform matrix in the ETKF rotates and rescales the forecast perturbations according to the geographical variations of the observation density and accuracy. For details on the transform matrix, see Bowler et al. (2008), Wang and Bishop (2003), and Whitaker and Hamill (2002). In MOGREPS the analysis perturbations are then added to the Met Office operational 4DVAR analysis to provide the initial conditions for the ensemble member forecasts.

When the ensemble size is small, the covariance estimates tend to be noisy and to include spurious correlations with increasing distance between geographical locations. A way around this problem is achieved by multiplying the background error covariance by some localization function, which increases the effective rank of the covariance matrix (Hamill et al. 2001). Mitchell et al. (2002) noted that the smaller the ensemble, the more severe the localization was required to be, which introduced substantial imbalances in the analysis. Localization in MOGREPS (Bowler et al. 2009) is included in the calculation of the transform matrix by using a limited set of observations within a domain centered on a set of points known as localization centers. The transform at a given point is calculated by using those transform matrices whose localizations centers are within 2000 km. These matrices are linearly combined using the distance between the grid points and the localization centers as weights. Localization in MOGREPS is not sufficient to compensate for the lack of the representation of the background error covariance outside the subspace defined by the ensemble forecasts. Therefore, to improve the background covariance estimate, MOGREPS uses a variable (per forecast cycle) inflation factor to ensure that the forecast ensemble spread matches the error of the ensemble mean forecast at $T+12$ h and at the radiosonde observation locations. The magnitude of the MOGREPS inflation factor is of the order of 70–80, which means that the ETKF perturbations are 70–80 times smaller than the innovations (Bowler et al. 2008). The large magnitude of the inflation factor is necessary to compensate for the limited ensemble size, inbreeding effects, and an incomplete treatment of model error (e.g., Whitaker and Hamill 2002).

In addition to the representation of model error resulting from the inflation factor, a model error term $\mathbf{Q}$ is also used in MOGREPS. This is represented by using stochastic-physics schemes: a random parameter scheme and a stochastic convective vorticity scheme (Bowler et al. 2008). These schemes aim to represent the effects of physical processes occurring at scales not resolved by the model and to account for uncertainties associated with empirical model parameters. The stochastic physics adds model perturbations to the ETKF initial condition perturbations, which makes it difficult to disentangle initial condition errors from model errors.

c. Verification of forecast errors

The verification of operational deterministic forecasts usually relies on the assumption that the unknown true state of the atmosphere is well represented by the observations or by the analysis verifying at the forecast...
lead time. The verification against the observations does not provide global information on the forecast error statistics, whereas the verification against the analysis provides a global estimate of the forecast error statistics.

The root-mean-square forecast error (rmse) of the high-resolution deterministic forecast $\mathbf{x}'(t)$ is given by

$$\text{rmse} = \sqrt{\langle [\mathbf{x}'(t) - \mathbf{x}(t)]^2 \rangle_t},$$

where $\mathbf{x}'(t)$ is the verifying analysis valid at time $t$ and $\langle \rangle_t$ denotes the time average. The main limitation of this approach is imposed by the sample size and by the fact that the statistical sample characterizes a climatological forecast error in which each realization represents a single weather condition. Therefore, we have a single measurement of the forecast error for a particular weather condition and the averaging is done over 1 month of single weather cases instead of sampling the same weather condition with an ensemble of forecasts.

d. Estimation of forecast error covariances in this study

The forecast error statistics depend on the physical processes governing the meteorological situation, on the observing system, and on how the physical processes are modeled. In particular, forecast error variances reflect the uncertainty in the background’s features or in the observations. As has already been mentioned, the estimation of the covariance matrix of the background error is difficult. This is because the true background state is unknown and a surrogate state must be used instead, the size of the $\mathbf{B}$ matrix is unfeasibly large, and there is not enough accurate data to calibrate background error statistics.

A way to estimate background error statistics for variational assimilation is to assume that they are uniform over a domain and stationary over a period of time. In this way one can estimate empirical statistics by taking a number of error realizations. Additional approximations used in modeling the background error covariance, such as isotropy, homogeneity, and closeness to balance, partially compensate for the missing information needed to estimate such a large covariance matrix (Bannister 2008b; Fisher 2003).

Useful information on the background error statistics can be gathered from diagnostics of an existing data assimilation system using the innovation method (described in Hollingsworth and Lönnberg 1986) and the National Meteorological Center [NMC, now known as the National Center for Environmental Prediction (NCEP)] method (described in Parrish and Derber 1992). Details on methods for estimating $\mathbf{B}$ can be found in Bannister (2008a).

The Met Office operational $\mathbf{B}$ matrix is estimated using the NMC method, where the forecast error is characterized by differences between 30- and 6-h forecasts valid at the same time. In this way the proxy of background errors include 24-h model error through the different time lengths of the model integration. In addition, when calculating forecast difference statistics, the mean value is not removed from the statistics in order to be consistent with the analysis algorithm, which does not explicitly correct for biases in the background (Lorenc et al. 2000).

So strictly speaking, the operational Met Office $\mathbf{B}$ matrix is a cross-product matrix rather than a covariance matrix.

Here, we describe only the ensemble method used in this study for generating a proxy for background errors. This method uses the MOGREPS ensemble and it is sketched in Fig. 1. Equation (4) in section 2b gives an estimate of the $\mathbf{P}^b$ matrix by taking the covariance statistics of the differences between each ensemble member and the mean at a particular time. The unknown true state is replaced by the ensemble mean and the ensemble of forecasts is assumed to have the same spread as the true forecast error.

For the estimation of the climatological $\mathbf{B}$ matrix, the averaging is performed over many assimilation cycles:

$$\mathbf{B} = \langle \langle (\mathbf{x}^b - \mathbf{x})(\mathbf{x}^b - \mathbf{x})^T \rangle_{\text{ens}} \rangle_{\text{time}},$$

where the angled brackets $\langle \rangle_{\text{ens}}$ and $\langle \rangle_{\text{time}}$ represent the averaging over $N$ ensemble members at a particular time.
and over many assimilation cycles, respectively. Here, \( x' \) represents the “truth,” which can be chosen to be the ensemble mean or the verifying analysis for that particular assimilation cycle considered.

When we assume \( x' = x^a \), the climatological \( B \) matrix is

\[
B = \langle (x^b - x^d)(x^b - x^d)^T \rangle_{\text{ens}}/\text{time},
\]

\[
= \langle (x^b - \bar{x}^b + \bar{x}^b - x^d)(x^b - \bar{x}^b + \bar{x}^b - x^d)^T \rangle_{\text{ens}}/\text{time},
\]

\[
\approx \langle (x^b - \bar{x}^b)(x^b - \bar{x}^b)^T \rangle_{\text{ens}}/\text{time}
\]

\[
+ \langle (\bar{x}^b - x^d)(\bar{x}^b - x^d)^T \rangle_{\text{time}},
\]

where \( \bar{x}^b = \langle x^b \rangle_{\text{ens}} \). Equation (10) states that under the assumption that the forecast error and the analysis error are uncorrelated, \( B \) can be expressed as the sum of two terms. The first term is the time average of the ensemble spread, which includes model error as stochastic physics and as an inflation factor, and the second term is the time average of the square of the rmse of the ensemble mean. The second term can be thought of as being a systematic component of the model error, given by the time average of the departure of the ensemble mean from the analysis.

Ensemble methods for the generation of surrogate background errors are very attractive when studying flow-dependent forecast errors and for providing a flow-dependent background error covariance for each assimilation cycle (Wang et al. 2008; Lorenc 2003b; Hamill and Snyder 2000).

3. Growth of forecast errors

a. Results and discussion

Here, we compare the results of the growth of the forecast errors predicted by the Met Office operational MOGREPS ensemble and by the implicit evolution of the Met Office operational 4DVAR approach.

The MOGREPS system consists of both global and regional forecasts; although, in this study we consider only the global prediction system. The global ensemble has 24 members of which 23 are perturbed using the ETKF and stochastic physics and 1 is the unperturbed forecast starting from the operational 4DVAR analysis. The global ensemble is run twice a day for 72 h. Here, the comparison is done for the first 7 days of December 2006, which implies hundreds of error realizations for the calculation of the forecast error covariance matrix.

Figure 2a shows the growth of forecast errors for temperature at 500 hPa over 24 h, calculated using the ensemble method described in section 2d. Figure 2b shows the error growth for zonal winds at 850 hPa in black and 250 hPa in gray. The error growth is split into six latitude bands: with equator–20°N, 20°–40°N, and 40°–90°N shown in the top 3 panels and equator–20°S, 20°–40°S, and 40°–90°S in the bottom 3 panels. In these and the following figures, the diamonds describe the growth of the ensemble spread, which includes the model error through the stochastic physics scheme and the inflation factor [i.e., first term in Eq. (10)], and the triangles describe the growth of the ensemble total rmse, that is, the sum of the ensemble spread and the rmse of the ensemble mean under the assumption that \( x' = x^a \) [see Eq. (10)]. The verifying analysis \( x^a \) comes from the operational 4DVAR system.

At \( T + 0 \) h, the initial condition error is the same, as expected, because the ensemble initial condition perturbations are constructed to sum to 0. As the time evolves, the forecast error grows faster when the systematic component of the model error is included (triangles). The ensemble spread and the ensemble total rmse evolutions are similar for the first 6 h and then they tend to separate. It is not possible to tell whether the forecast error comes from the initial condition error, the model error or, more likely, both. MOGREPS uses a variable inflation factor to ensure that the ensemble spread matches the ensemble mean error at \( T + 12 \) h at the radiosondes locations. This implies that at \( T + 12 \) h the ensemble total rmse should be \( \sqrt{2} \) of the ensemble spread, which follows from Eq. (10) when the first term representing the square of the ensemble spread is equal to the second term representing the mean squared error of the ensemble mean. This happens in the equatorial regions and the southern polar region for temperature at 500 hPa and only in the equatorial regions for zonal winds at both 850 and 250 hPa. In the other latitude bands the ensemble tends to be overspread at \( T + 12 \) h. This may be explained partly by the use of overlapping observation domains in the localized ETKF for the calculation of the inflation factor and partly by the fact that the verifying analysis as a proxy of the truth may underestimate the forecast error, because at \( T + 12 \) h the forecast error and analysis error may be correlated.

In both cases the error evolution includes the stochastic physics perturbations. This is because the MOGREPS system is a filtering method that linearly mixes and scales forecast errors from the previous cycle and therefore feeds back model uncertainties in the initial condition perturbations. To estimate the contribution of the stochastic physics to the forecast error, the ETKF initial condition perturbations have been removed and only the model perturbations have been used to generate the sample of background states to be integrated in time with the full nonlinear model. This error evolution, illustrated in Fig. 2 by the squares, represents the growth of both the random
component of the model error coming from stochastic physics scheme and the systematic component of model error coming from the time-average departure of the ensemble mean from the verifying analysis. The stochastic physics is embedded in the model by perturbing model parameters during the model integration. At $T = 0$ h, the stochastic physics does not contribute to the initial condition error. As time evolves, the contribution of the

Fig. 2. (a) Temperature at 500 hPa and (b) zonal wind at 850 hPa in black and at 250 hPa in gray for the latitude bands indicated. The diamonds represent the ensemble spread, the triangles the total rmse of the ensemble, and the squares the stochastic physics error (see text for more details).
stochastic physics varies from 25% to 50% depending on latitude bands. In the equatorial regions the stochastic physics generate most of the growth in the first 6 h: in these regions the model is more sensitive to the random sampling of the convective-scale parameter.

Figure 3 shows the evolution of the ratio of unbalanced pressure over total pressure errors over 24 h at 500 hPa. This ratio represents the remaining unbalanced component of the flow in the system after imposing the geostrophic balance relationship (Bannister 2008b). In the extratropics the imbalance in the forecast of pressure at 500 hPa is due to the fact that the stochastic physics perturbations are the only perturbations in the model evolution. On the other hand, when the stochastic physics perturbations are added to the ETKF perturbations, the total perturbations are in balance and stay in balance over the 24-h period. In the equatorial region, as expected, the ETKF initial conditions are unbalanced and the total perturbations stay unbalanced in both cases.

Imbalances make the forecast error covariance less accurate at initial times, while as time evolves it does not contribute to the forecast error because it is dissipated quickly by the model. It is known in the literature that the model error term is one of the causes of the imbalance in the EnKF and that in general it is hard to represent (Mitchell et al. 2002). It is also known from linear theory that if the model error is balanced, then the Kalman filter state is also balanced. Therefore, one way of improving the balance issue in the EnKF is using methods that generate approximately balanced model perturbations.

Now we want to compare the error growth estimated using the ensemble method with the growth of the forecast error implied by 4DVAR. The evolution of the forecast error covariance is implicit in the 4DVAR system. The randomization method, described in Fisher and Courtier (1995) and Andersson et al. (2000), provides a diagnosis of the background errors at the beginning of the assimilation period, when $B$ is of the form $B = UU^T$.

If $j_l$ is a set of $L$ random vectors in control variable space drawn from a Gaussian distribution with zero mean and unit variance, then applying $U$ to this set of random vectors gives a distribution with the covariance matrix $B_{i_l} = \frac{1}{L} \frac{1}{C_2} \sum_{l=1}^{L} (U_j l) (U_j l)^T$. The standard deviations produced by using the randomization method are noisy if $L$ is small. With Gaussian statistics and sample size $L$, the randomization noise in the estimated standard deviations is $1/\sqrt{2L}$ (e.g., 10% for $L = 50$). Here, we use 50 random vectors and evolve them with the linearized model dynamics for 24 h to simulate the flow dependency of 4DVAR. The comparison of the error growth predicted by 4DVAR and by MOGREPS over 24 h is shown in Fig. 4a for temperature at 500 hPa, and in Figs. 4b and 4c for zonal winds at 850 and 250 hPa, respectively. In these figures, the open symbols represent the nonlinear model evolution and the filled symbols the evolution with the linearized model. The diamonds represent the growth...
of the MOGREPS spread while the circles show the growth of random initial conditions sampled from the Met Office operational background error covariance matrix. Finally, the stars represent the linear growth of random initial conditions sampled from a static background error covariance matrix estimated from the European Centre for Medium Range Weather Forecasts (ECMWF) ensemble of analyses (Fisher 2003). The forecast error started from the Met Office operational random initial conditions (circles) in general does not grow except in the region $40^\circ$–$90^\circ$S for temperature at 500 hPa and zonal winds at 850 hPa. In other regions the forecast error actually decreases. In the tropical regions the forecast error increases in the first 6 h only for temperature at 500 hPa. This may be related to the detailed structure of the $B$ matrix but needs to be investigated in more detail. The evolution of a random sample of ECMWF background errors (stars) is in general similar to the forecast error evolution starting from random samples of the operational $B$ matrix (circles), although the forecast errors magnitudes are much smaller. The ECMWF background error statistics are generated from forecasts of the same length as the background, that is, 6 h, starting from the ensemble of 4DVAR analyses that sample the analysis error distribution. Therefore, the ECMWF background error statistics represent a more appropriate proxy for 6-h forecast error than the Met Office operational background error statistics, which includes 24-h model error integration from differences between forecasts of different lengths valid at the same time. In the extratropics the growth of random initial conditions sampled from static $B$ matrices is smaller than the growth of the ensemble spread for both temperature and zonal winds, while generally the opposite occurs in the tropics. The lack of error growth from random initial condition perturbations may be explained by the covariance model, defined by a series of transformations represented by the operator $U$, which imposes homogeneity and isotropy. The initial condition perturbations selected in this way do not project onto rapidly growing structures, while MOGREPS selects fast-growing modes that quickly develop in forecast errors. This is not a property of the different sampling techniques: Magnusson et al. (2009) showed that random initial condition perturbations which have the same dynamical properties as the variability of the atmosphere exhibit a pattern of growth similar to the ones generated by more sophisticated methods such as singular vectors and ETKF methods.

The tangent linear approximation is generally good for both temperature and zonal winds when the initial conditions are generated by the randomization method (filled versus open circles) except for equatorial regions where the nonlinear effects of the physics are stronger for light winds. When the initial conditions are generated by the ETKF, the linear and nonlinear evolution (filled versus open diamonds) of the ensemble spread is similar for the Northern Hemisphere extratropics and for $40^\circ$–$90^\circ$S in the case of temperature at 500 hPa and zonal winds at 850 hPa. In the case of zonal wind at 250 hPa, the growth patterns of the ensemble spread predicted by the nonlinear and the linear dynamics are different mainly in the extratropics. The discrepancy may be caused by different effects: most of the physics of the nonlinear model are omitted in the linearized model (Rawlins et al. 2007), which leads to the enhancement of the effects of convection in the Southern Hemisphere summer regions; the stochastic physics is absent in the linearized dynamics and finally the resolution of the full nonlinear model is 3 times higher than the linearized model’s resolution.

Finally, although in theory a strong constraint 4DVAR only allows for perturbation growth, assuming Gaussian and unbiased error statistics, in practice $B$ includes model error. Therefore, for a more realistic comparison, in Fig. 5 we compare the growth of the ensemble total rmse (triangles) and the linear growth of random initial conditions sampled from the Met Office operational $B$ matrix (filled circles) and from the ECMWF $B$ matrix (stars) with the verification error (crosses). Figure 5a shows the comparison for temperature at 500 hPa and Fig. 5b for zonal wind at 250 hPa only, since the results at 850 hPa are similar. The verification error of the operational deterministic forecast is averaged over 1 month of cases. It includes the model error due to model deficiencies in the model integration, such as model parameterization and resolution. The growth of the verification error is smaller but rapid in the first 12 h compared to the linear growth of random samples of background errors, for both the operational and the ECMWF $B$ matrices. The early rapid growth is the growth of the unbalanced motions, which is due to geostrophic adjustments; the growth stops between $T + 12$ h and $T + 24$ h and resumes after $T + 24$ h (not shown here), due to the growth of the balanced motions. As expected in the equatorial regions, the error growth behaves differently since the geostrophic adjustment does not happen. In the case of the linear evolution of the random samples of the background errors, the forecast error does not grow but actually decreases over 24 h, independently of the calibrating datasets used for the estimation of $B$. The ensemble total rmse exhibits a pattern of rapid growth in the balanced flow due to the fact that the initial conditions are very well balanced, as shown in Fig. 3. It shows an offset compared to the verification error, because it implicitly includes the rmse of the ensemble mean. Equation (10) can be thought of as the sum of the ensemble spread and the verification error when the ensemble mean
coincides with the deterministic forecast. Therefore, the ensemble total rmse and the verification error of the deterministic forecast are comparable when the ensemble spread gets very small. Comparing Fig. 2 with Fig. 5, the contribution of the rmse of the ensemble mean, represented by the difference between the triangles and the diamonds, is less than the verification error. This may be due to the limited sample size of the ensemble or because the assumption of uncorrelated forecast error and analysis error is not valid.

Fig. 4. (a) Temperature at 500 hPa, and zonal wind at (b) 850 and (c) 250 hPa. Open (filled) diamonds show nonlinear (linear) growth of the ensemble spread, open (filled) circles show nonlinear (linear) growth of the random initial conditions sampled from the operational B matrix, and stars show the linear growth of the random initial condition sampled from the ECMWF B matrix (see text for more details).
b. Implications

The results in this paper depend on the operational assumptions and approximations made in the Met Office 4DVAR and ETKF systems and are limited by the sample of weather regimes and the period of the year analyzed. Nevertheless, they provide interesting information on both systems and possible ways to improve each.

The background error assumed in 4DVAR comes from the natural growth of small unbalanced perturbations to the model attractor, while that assumed in the ETKF comes from the fastest-growing balanced perturbations that stay on the model attractor. The growth of the deterministic forecast error, shown by crosses in Fig. 5, initially reflects the geostrophic adjustments, which return the system to a state on the model attractor, and later reflects the growth of the balanced motions.

At short lead times, the background error assumed in 4DVAR is comparable with the error in the deterministic forecast in the extratropics. This happens in particular when the sample of perturbations is drawn from the ECMWF background error statistics (shown by stars in Fig. 5), which represents a more appropriate proxy of 6-h forecast error than do the NMC method error statistics. Other diagnostics also show that ECMWF background error statistics have smaller length scales than NMC background error statistics. This can be explained by the fact that the ensemble statistics are computed from forecasts of the same lead time as the background, (i.e., 6 h), while the NMC method includes the effects of the saturation of small-scale errors and the growth of large scales over the 24-h model integration. These two results suggest that ECMWF error statistics better represent the short-range forecast error.

The 4DVAR method should also benefit from the use of flow-dependent perturbations generated by MOGREPS instead of using isotropic and homogeneous perturbations that do not project onto rapidly amplifying structures (see, e.g., Buehner et al. 2010; Berre and Desroziers 2010). However, the diagnostics show that MOGREPS characterizes the growth of balanced large-scale motions and not necessarily the growth of forecast error within the 4DVAR 6-h window. The MOGREPS inflation factor adjusts the ensemble spread at T + 12 h to match the rmse of the ensemble mean but it does not necessarily capture the slow growth of the systematic model error, which may lie in a subspace not spanned by the ETKF perturbations. MOGREPS has a small number of members; increasing it should also improve the sampling error. Alternatively, the ETKF method could be replaced by using an ensemble of 4DVAR as the ensemble initial conditions in which observations are randomly perturbed in each analysis cycle according to their error statistics.

The calculated deterministic forecast errors represent the real forecast error growth in the deterministic model at sufficiently large time scales. However, at T + 6 h, the calculated forecast error is questionable because the forecast error and analysis error are correlated. In addition, the
rapid growth of the errors at short range is due to geostrophic adjustment, which is difficult to represent as a model error term. Further work is needed to remove this undesirable feature of the system.

4. Conclusions

In this paper we have compared the growth of forecast errors from covariances modeled by the Met Office...
operational 4DVAR and ETKF methods and with the growth of the rmse of the Met Office operational deterministic forecast error.

The ensemble used in this study is provided by MOGREPS, where the initial conditions are generated by an ETKF in which the model error contribution is included via a tuning inflation factor that adjusts the ensemble spread to the ensemble mean error at $T + 12$ h and a stochastic physics scheme that randomly samples model parameters within specified errors. The ETKF requires a very large inflation factor to keep the ensemble spread close to the ensemble mean error. The tuning achieves this in the equatorial regions and in the southern polar hemisphere for temperature at 500 hPa and in the equatorial regions only for zonal winds at 850 and 250 hPa, while in the other latitude bands the ensemble tends to be overspread at $T + 12$ h. The contribution from the stochastic physics varies from 25% to 50% depending on the latitude bands. The stochastic physics perturbations mainly introduce imbalance into the forecasts of pressure at 500 hPa when they are the only perturbations in the model evolution. On the other hand, when the stochastic physics perturbations are added to the ETKF perturbations, the total perturbations are in balance and stay in balance over the 24-h period. In general, the forecast error predicted by the ETKF grows rapidly because the initial perturbations are balanced and the ETKF tends to select fast-growing modes that quickly develop in forecast errors.

The forecast error predicted by the implicit evolution of 4DVAR in general does not grow except in the region $40^\circ$–$90^\circ$S for temperature at 500 hPa and zonal winds at 850 hPa. In other regions the forecast error actually decreases except in the first 6 h only for temperature at 500 hPa in the tropical regions. This appears to happen in both cases when the random sample of background errors is drawn from the operational B matrix estimated using the NMC method and when it is drawn from a static background error covariance matrix estimated from the ECMWF ensemble of analyses. One explanation may be that the covariance model imposes homogeneity and isotropy on the selected initial condition perturbations, which do not project onto rapidly growing structures.

The growth of forecast errors predicted by MOGREPS and 4DVAR is compared with the verification. The verification error shows smaller amplitude but more rapid growth than the linear evolution of random samples of homogeneous and isotropic initial conditions in the first 12 h, then it gets larger but stops growing between $T + 12$ h and $T + 24$ h. When compared with the ensemble total rmse, the verification error is larger than the rmse of the ensemble mean, maybe due to the limited size of the ensemble or the invalidity of the assumption of uncorrelated forecast and analysis errors.

Finally, in all of the above cases it is not possible to tell if the forecast error comes from the initial condition error or model error. However, these results reflect the limitation of the period of the year and the sample of regimes used in this study.

Acknowledgments. The author would like to thank the many people at the Met Office for their help, suggestions, and encouragement: Sarah Beare, Neill Bowler, Adam Clayton, Mike Cullen, Andrew Lorenc, Stephen Oxley, Mike Thurlow, Martin Willet, and Marek Wlasak. Particular thanks to Mike Cullen for his inspirational guidance and valuable comments, and for reading the draft manuscript. Thanks to Anca Brookshaw for proofreading the manuscript. The author would also like to thank three anonymous reviewers and the editor for their careful reading and constructive comments, which helped to greatly improve this manuscript.

REFERENCES


Fisher, M., 2003: Background error covariance modelling. Extended Abstracts, Seminar on Recent Developments in Data Assimilation, Reading, United Kingdom, ECMWF, 45–64.


