Resilience of Hybrid Ensemble/3DVAR Analysis Schemes to Model Error and Ensemble Covariance Error

Brian J. Etherton

RSMAS, University of Miami, Miami, Florida

Craig H. Bishop

UCAR, Boulder, Colorado, and Naval Research Laboratory, Monterey, California

ABSTRACT

Previous idealized numerical experiments have shown that a straightforward augmentation of an isotropic error correlation matrix with an ensemble-based error correlation matrix yields an improved data assimilation scheme under certain conditions. Those conditions are (a) the forecast model is perfect and (b) the ensemble accurately samples the probability distribution function of forecast errors. Such schemes blend characteristics of ensemble Kalman filter analysis schemes with three-dimensional variational data assimilation (3DVAR) analysis schemes and are called hybrid schemes. Here, we test the robustness of hybrid schemes to model error and ensemble inaccuracy in the context of a numerically simulated two-dimensional turbulent flow. The turbulence is produced by a doubly periodic barotropic vorticity equation model that is constantly relaxing to a barotropically unstable state. The types of forecast models considered include a perfect model, a model with a resolution error, and a model with a parameterization error. The ensemble generation schemes considered include the breeding scheme, the singular vector scheme, the perturbed observations system simulation scheme, a gridpoint noise scheme, and a scheme based on the ensemble transform Kalman filter (ETKF). For all combinations examined, it is found that the hybrid schemes outperform the 3DVAR scheme. In the presence of model error a perturbed observations hybrid and a singular vector hybrid perform best, though the ETKF ensemble is competitive.

1. Introduction

The ensemble transform Kalman filter (ETKF) (Bishop et al. 2001) is currently used at the National Centers for Environmental Prediction (NCEP) to select flight tracks for dropwindsonde-equipped weather reconnaissance aircraft (Szunyogh et al. 1999, 2000; Majumdar et al. 2001). From a theoretical perspective, the ability of the ETKF to select reconnaissance flight tracks that maximally reduce forecast error variance depends on the accuracy of the error covariances used by the ETKF as well as their similarity to those assumed by NCEP's data assimilation scheme. Currently, there is a mismatch between the low-rank flow-dependent error covariances employed by the ETKF and the flow-independent error covariances assumed by NCEP's data assimilation scheme.

The ETKF estimates error covariance matrices in terms of the outer product of a matrix containing trans-

formed ensemble perturbations. The transformation coefficients are obtained by repeatedly solving the Kalman filter error statistics equations within the ensemble subspace (Bishop et al. 2001). These ensemble-based error statistics have flow-dependent characteristics such as one-point error correlation patterns that stretch out along fronts.

In contrast, the one-point error correlation patterns used for data assimilation at operational centers such as NCEP (Parrish and Derber 1992) and (until recently) the European Centre for Medium-Range Weather Forecasts (ECMWF) [Courtier et al. (1998), which was superceded by Rabier et al. (2000)], and those run by the National Aeronautics and Space Administration (NASA) (Cohn et al. 1998) and by the U.S. Navy (Daley and Barker 2001), are generally horizontally isotropic and temporally invariant. The assimilation of targeted data with error covariances profoundly different to those assumed by the ETKF limits the ability of the ETKF to quantitatively predict the forecast error variance, reducing the effect of targeted observations. To overcome this limitation, one would need to assimilate data with an error covariance model qualitatively similar to that assumed by the ETKF. Schemes that would have qual-

Corresponding author address: Brian J. Etherton, University of North Carolina at Charlotte, Department of Geography and Earth Science, 9201 University City Blvd., 450 McEniry Building, Charlotte, NC 28223.
E-mail: betherto@uncc.edu

itatively similar error covariance models to the ETKF would include all ensemble-based data assimilation schemes and, perhaps, a four-dimensional variational data assimilation (4DVAR) scheme using a flow-dependent background error covariance model. However, operational centers would only be able to justify such a change of error covariance model if the schemes that assumed error covariances similar to the ETKF were superior to those that are currently used.

Work by Evensen (1994), Evensen and van Leeuwen (1996), Burgers et al. (1998), Houtekamer and Mitchell (1998), Lermusiaux and Robinson (1999), Pham (2001), Anderson (2001), Whitaker and Hamill (2002), and others indicate that ensemble-based error covariances lead to highly effective data assimilation schemes in the absence of model error.[1] However, there is less literature on the performance of ensemble-based error covariances in the presence of model error. Keppenne (2000), Blanchet et al. (1997), Pham (2001), and Mitchell and Houtekamer (2000) have examined the performance of ensemble-based error covariances in the presence of model error. In these studies, the model error was unbiased, and characteristics of the model error covariance were perfectly known, and this perfect knowledge was used to create ensemble-based error covariance models that performed well. In operations, little is known about model-error bias let alone model-error covariance.

In this study, we deliberately abstain from using any detailed information about the bias and covariance of model errors in our data assimilation schemes because, in our view, a lack of such information is the typical state in operational weather forecasting. We investigate two types of model errors: a parameterization error and a resolution error. We ignore the model-error bias and the model-error covariance associated with these errors. The effect of such errors on the performance of ensemble Kalman filters has received relatively little attention in the meteorological literature, and covariances directly generated by the ETKF have yet to be discussed in the literature.

Because of unknown model error, limitations of the techniques used at operational centers to generate ensembles, and inadequacies in our ability to accurately specify observation error and the observational error of representation (Daley 1991), it is impossible in practice to obtain an ensemble whose members represent a random sample of the true probability distribution function of forecast errors. The primary purpose of this paper is to report on some experiments that test the resilience to model error and poor ensemble formation of an analysis scheme using the hybrid ensemble Kalman filter/3DVAR forecast error covariance model developed by Hamill and Snyder (2000). A secondary objective is to illustrate how an ETKF ensemble generation scheme can be used with Hamill and Snyder's (2000) hybrid

error covariance formulation to form a competitive analysis scheme.

The ensemble generation techniques to be considered in this study include perturbed observations/system simulation (Houtekamer et al. 1996), singular vector/optimal perturbations (Molteni et al. 1996), the ETKF (Bishop et al. 2001), breeding (Toth and Kalnay 1993), and gridpoint perturbations. All achieve optimality under some set of ideal conditions. In this paper, we test their effect on the performance of the hybrid analysis scheme in systems where these ideal conditions are far from being met. Section 2 describes the methodology of our data assimilation experiments. Section 3 presents the results, and concluding remarks follow in section 4. Details of the various ensemble generation techniques employed are presented in the appendix.

2. Methodology

a. The two-dimensional turbulence model

Our experiments are performed using a doubly periodic barotropic vorticity model identical to that used in Bishop et al. (2001). The governing equation for this model is

$$\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)\nabla^2\Psi + \beta\frac{\partial\Psi}{\partial x}$$
$$= \delta(\nabla^2\Psi)^n + \kappa(\nabla^2\overline{\Psi} - \nabla^2\overline{\Psi}_r), \qquad (1)$$

where Ψ is the streamfunction, δ and n are parameters associated with numerical diffusion, κ is a relaxation parameter, and β is the planetary vorticity gradient (Holton 1992). The values for the parameters δ, n, κ, and β are 0.08, 3, 0.005, and 2.15×10^{-11} m^{-1} s^{-1} (the β value at 20° latitude), respectively. A relaxation scheme relaxes vorticity values at grid points by a quantity proportional to the difference between the zonally averaged vorticity, $\nabla^2\overline{\Psi}$, and a reference zonal average of vorticity, $\nabla^2\overline{\Psi}_r$. In this manner, local vorticity extrema are not preferentially eroded, and meridional gradients of zonally averaged vorticity are maintained. The model domain is 6400 km by 6400 km.

To generate pseudodata for use in data assimilation experiments, a pseudotruth run is performed. The pseudotruth run is initialized from a zonal vorticity field consisting of two areas of high positive vorticity that are barotropically unstable (Fig. 1). To tap the instability, small amplitude gridpoint noise is also included in the initial state. To ensure that the circulation around the doubly periodic domain is zero, negative vorticity lies between the strips of positive vorticity. The regions of positive vorticity crudely represent the intertropical convergence zone and midlatitude storm track.

For the first 9 days, the integration is performed with a grid cell size of 50×50 km^2. After 9 days, four coherent tropical-cyclone-like vortices emerge from the southern vortex strip. This state will hereafter be re-

---

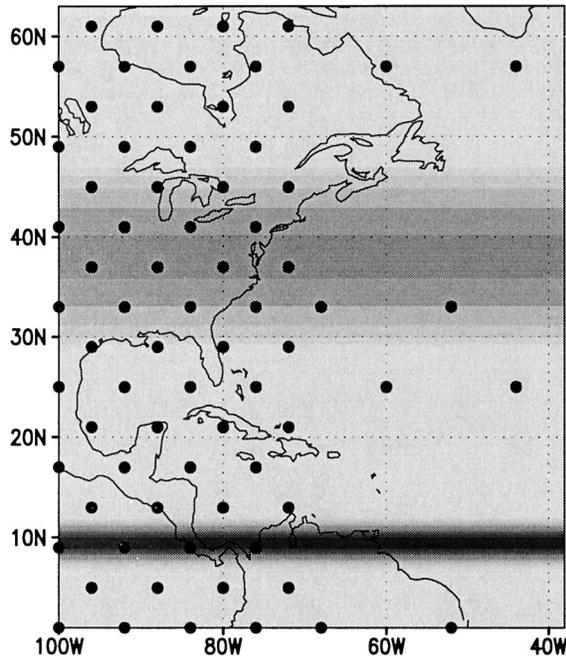[1] Tippett et al. (2003) point out that all these filters may be viewed as square root filters.

FIG. 1. Initial condition vorticity field, with darker shades indicating higher vorticity values (2.5×10^{-5} s^{-1} at 9°N, 1×10^{-5} s^{-1} at 37°N) and lighter shades indicating small or negative values (-0.5×10^{-5} s^{-1} at 21°N). The locations of observation sites in the routine observing network are denoted by the dots.

ferred to as day 0. For the sake of computational efficiency, after day 0, the pseudotruth run was continued at the larger grid cell size of either 100×100 km^2 or 200×200 km^2. As described below, these two different resolutions are used for the pseudotruth in different experiments. The agency model forecasts were run at 200 $\times$ 200 km^2. Apart from some runs to check the sensitivity of our results to the day 0 initial condition error, the day 0 initial condition error for agency model forecasts is set equal to the difference between the day 1 and day 0 200×200 km^2 pseudotruth fields. In other words, the agency model forecasts are started 1 day ahead of the pseudotruth at day 0. This results in initial condition error at day 0 similar to that which would be produced by a model that systematically propagated disturbances too rapidly.

At 100×100 km^2 resolution, the day 0 vortices begin to rotate around each other and merge together. By day 5, two of the original four vortices have merged, leaving only three (Fig. 2a). Later, in interacting with the northern vortex strip, the remaining vortices are sheared apart and absorbed into the midlatitude shear zone. This regime change is complete by about day 25 of the pseudotruth forecast. After this regime change, the vorticity structures were more ribbonlike, with vorticity fronts rather than individual vortices. A sample vorticity field after the regime change is shown in Fig. 2b. The lower-resolution pseudotruth run is similar to the 100×100
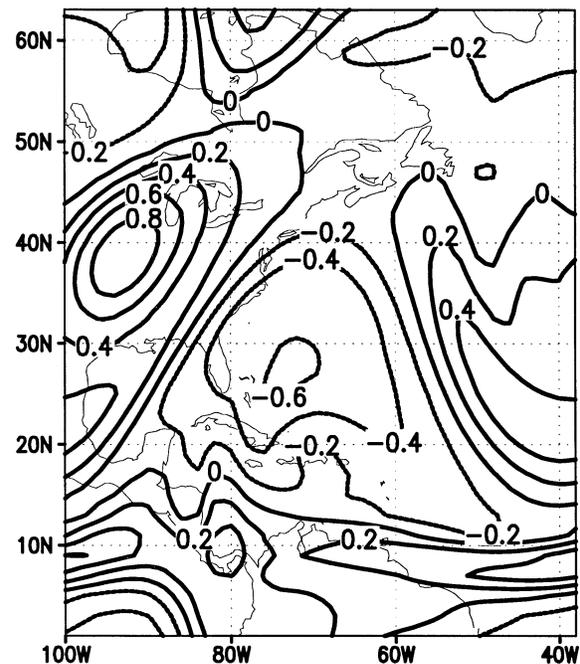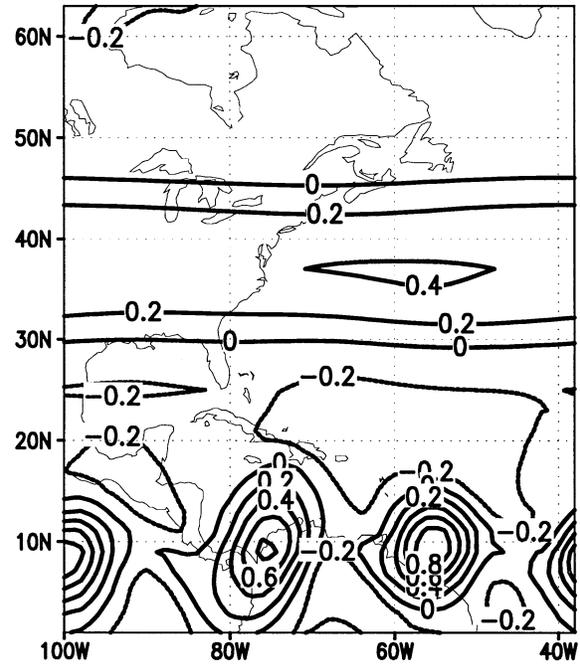

FIG. 2. The vorticity field of the pseudotruth run on days (a) 5 and (b) 30. The pseudotruth run was performed using a barotropic vorticity model of 100-km resolution, initialized after a 9-day evolution of the field shown in Fig. 1. Vorticity is in units of 10^{-5} s^{-1}, and contours are every 0.2×10^{-5} s^{-1}.

km^2 resolution run, though the initial vortices are weaker and the regime change occurs a few days earlier.

b. Experimental configurations

To test the performance of the data assimilation scheme with model errors, agency forecast models with (i) no model error, (ii) a parameterization error, and (iii) a resolution error were used. These experiments test the ability of the 200 × 200 km^2 resolution model to assimilate data provided from a different model, one which represents the truth. Fixing the agency forecast model resolution at 200 × 200 km^2 ensures that the ratio of degrees of freedom in the model to either the number of ensemble members or the number of observation sites, values that will be shown later, is the same for all simulations.

The perfect agency model used both the same grid spacing and same reference vorticity field for relaxation as used in the pseudotruth run. The only difference between the pseudotruth and the agency model was the 1-day initialization error. For the parameterization error agency model, the vorticity relaxation scheme relaxed to a vorticity field different than the field relaxed to in the pseudotruth run. Instead of relaxing the agency's model to the reference zonal average of vorticity field used in the pseudotruth run, it was relaxed to the vorticity values along the western boundary at the start of each forecast. The reference values from the westernmost grid points in the model, being determined by the meandering flow trajectory, were often very different from the fixed vorticity values toward which the pseudotruth was relaxing. For the resolution error experiments, the pseudotruth was run at a higher resolution, 100 × 100 km^2, rather than the 200 × 200 km^2 resolution used by the agency model. However, the initial vorticity field shown in Fig. 1 was used in the relaxation scheme for both the pseudotruth run and agency forecast model. The resolution and parameterization errors were not combined.

c. The hybrid analysis scheme

As given in Cohn (1997), the minimum analysis error variance increment

$$\mathbf{x}^a - \mathbf{x}^f = \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^f), \quad (2)$$

is used to combine observations of vorticity, $\mathbf{y}$, with the first-guess field, $\mathbf{x}^f$, to produce the new analysis $\mathbf{x}^a$. The observation operator, $\mathbf{H}$, converts streamfunction observations to vorticity values but does no interpolation, as observations are taken at model grid points.

The hybrid analysis scheme approximates the forecast error covariance matrix $\mathbf{P}^f$ with a mix of parameterized covariances, $\mathbf{B}^f$, and flow-dependent, ensemble-based covariances, $\mathbf{F}^f$. The forecast error covariance matrix $\mathbf{P}^f$ from Hamill and Snyder (2000) is given by

$$\mathbf{P}^f = (1 - \alpha)\mathbf{F}^f + \alpha\mathbf{B}^f, \quad (3)$$

where α is a constant, $\mathbf{B}^f$ is a 3DVAR-like time-invariant forecast error covariance matrix based on a quasi-isotropic correlation function, and

$$\mathbf{F}^f = \mathbf{Z}^f \mathbf{Z}^{fT}, \quad (4)$$

where each column of the matrix $\mathbf{Z}^f$ is proportional to the difference between an ensemble member forecast and a reference forecast at the time at which observations are to be assimilated. Depending on the ensemble generation scheme employed, the reference forecast is either the ensemble mean or a "control" forecast. With α equal to 1, $\mathbf{P}^f$ is $\mathbf{B}^f$, the 3DVAR correlation matrix. With α between zero and one, $\mathbf{P}^f$ is a mixture of flow-dependent and time-invariant error statistics. For a further discussion of ensemble-based versus isotropic covariance matrices, see Bishop et al. (2001).

A problem with the formulation given by (3) is that the weights on the error correlations associated with $\mathbf{B}^f$ and $\mathbf{F}^f$ not only depend on the parameter α but also the magnitude of the variances assigned to $\mathbf{B}^f$ and the evolving ensemble-generated variances of $\mathbf{F}^f$. For example, suppose that the ensemble variances became smaller for some reason. After such a change, isotropic error correlations would dominate the correlation formulation more than before the change. In section 2h, Dee's (1995) maximal likelihood theory is used to ensure that the parameter α gives a clear indication of the relative weights of flow-dependent and isotropic error correlation models.

d. Construction of the observation vector, y, and the observation error covariance matrix, R

Error prone observations were simulated by adding random values from a Gaussian distribution to values of the pseudotruth. For the perfect model and parameterization error experiments, the diagonal elements of the observation error covariance matrix, $\mathbf{R}$, listed the exact variances used to generate the random observation errors. The matrix $\mathbf{R}$ was diagonal, as observation errors at one location were assumed to be uncorrelated with observation errors at all other sites. The choice of observation error variance was based on an estimate of forecast error. The root-mean-square gridpoint vorticity error of the initial first-guess field is about 7×10^{-7} s^{-1}. The root-mean-squared instrumentation error is taken to be one-fifth of this value, 1.4×10^{-7} s^{-1}, which squared gives $\sim 2.0 \times 10^{-14}$ s^{-2}, the value used as the variance of the instrumentation error.

For the resolution error experiments, there is an additional source of observation error, the error of representation. Each grid box in the 200 × 200 km^2 domain contains four grid boxes of the 100 × 100 km^2 pseudotruth run. However, observations taken from a 100 × 100 km^2 grid box are used to represent the value of the 200 × 200 km^2 grid box. To quantify this error, the average squared difference between the value of each of the four pseudotruth 100 × 100 km^2 grid boxes with-
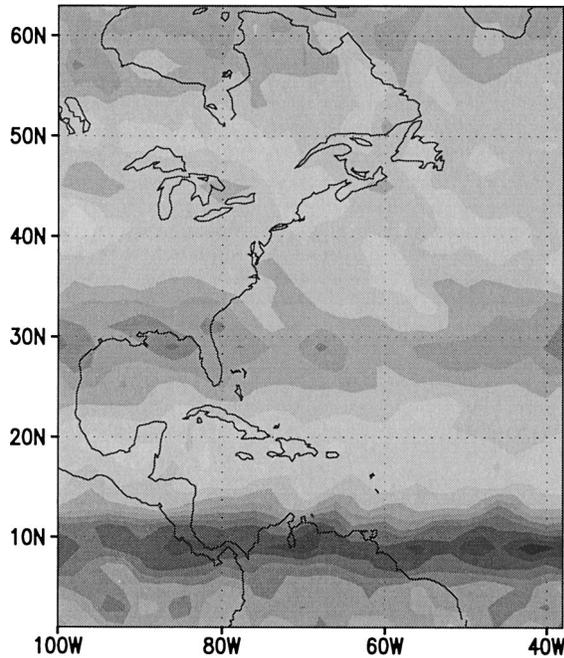
FIG. 3. The representation error (shaded) for the resolution error agency model. Values range from approximately 2.5×10^{-7} s^{-1} in the lightest shade to approximately 5.8×10^{-7} s^{-1} in the darkest shade.

in a 200×200 km^2 grid box and the mean of those four values are calculated for each day of the pseudo-truth run for every grid point. The average value over the final 69 days (after the regime change) of the 99 days of integration of this squared difference was used as the representation error variance in the data assimilation schemes for the resolution error agency model. The error of representation is highest in the regions that have the highest average vorticity gradients (Fig. 3). The range of the squared error of representation is between 6.5×10^{-14} s^{-2} and 25.9×10^{-14} s^{-2}, with a mean value of 17.5×10^{-14} s^{-2} or about 9 times larger than the instrumentation error variance.

The observational network was constructed to crudely mimic the observation density fluctuations found on earth. The accuracy of the mainly satellite-based observational network over the oceans is likely to be less accurate than the land-based network, which includes a multitude of in situ data. To crudely simulate this observational accuracy difference, more observations were placed over the western part of the domain (over "land") than on the eastern half (over "ocean"). Figure 1 shows the observational network of 72 observations in the $n = 1024$ gridpoint model domain.

The time interval between taking observations of the two-dimensional turbulent flow was selected to be crudely consistent with the ratio of error growth to intervals between data assimilation cycles used at operational centers. Using ECMWF data (Lorenz 1982) and NCEP data (Kistler et al. 2001), the size of 1-day error

growth in large-scale models is about 1.5 and 1.7, respectively. Assuming exponential (linear) growth over the 24-h period, the growth rate of errors between 6-h data assimilation cycles in operations is estimated to be about $1.7^{1/4}$ $(1 + 0.7/4)$, or approximately 1.14 (1.18). For the parameterization error and resolution error agency forecast models, the square root of the ratio of the average daily 24-h squared streamfunction forecast error to the average daily squared streamfunction analysis error is 1.4. For the perfect agency forecast model, the ratio is lower, 1.2. Thus, the ratio of error growth over one data assimilation interval is slightly higher in our experiments than it is in operations.

Since nonlinear error propagation can lead to highly non-Gaussian error statistics, we desire that the degree of nonlinearity of error propagation in our experiments would be similar to that in operational centers. We have estimated the importance of nonlinear error propagation in our experiments by calculating the ratio of the nonlinear terms to the linear terms. At the forecast time, the average value over 99 days is 0.192 for the perfect agency model, 0.300 for the resolution error agency model, and 0.299 for the parameterization error agency model. These values indicate that errors in the second significant figure and possibly the first significant figure could be expected to arise from a straightforward attempt to propagate error covariance information with a purely linear model.

e. Isotropic error covariance model

At many operational centers, the forecast error covariance matrix $\mathbf{P}^f$ is approximated using a time-invariant covariance matrix (e.g., Courtier et al. 1998; Cohn et al. 1998; Parrish and Derber 1992; Daley and Barker 2001). Such time-invariant forecast (background) error covariances will also be called 3DVAR covariances. For the two-dimensional turbulent flow used in this experiment, vorticity determines the entire state; so only vorticity correlations are needed. Following Daley (1991), in this paper we set the time-invariant covariance matrix, $\mathbf{B}^f$, to be equal to the product of a scalar variance with an error correlation matrix. The terms in the correlation matrix are obtained from an error correlation function describing how errors at one location are related to errors at all other locations in the model domain. For vorticity, a reasonable error correlation matrix is the Laplacian of the streamfunction correlation function given in Daley [1991, chapter 4, P. 117, Eq. (4.3.21)]. The use of this Gaussian function was developed by Sasaki (1960) and Barnes (1964). For the $\mathbf{B}^f$ matrix, the first-guess error correlation function for vorticity takes the form

$$f(r) = [1 + \ln(0.1) \times (r/D)^2] \times \exp[\ln(0.1) \times (r/D)^2], \qquad (5)$$

where r is the distance away from a location, and D is

TABLE 1. $\mathbf{B}^f$ matrix specifications.

Model error	Length scale (km)	Variance (10^{-12} s^{-2})
Perfect	800	0.299
Parameterization	1000	1.263
Resolution	1100	0.765

a correlation length scale. While this is the same equation as was used in Bishop et al. (2001), D was tuned to a different value in that paper. Here, D is tuned by running a 99-day simulation, using straight 3DVAR to assimilate observations of the pseudotruth every 24 h. The value of the length scale D that produces the lowest daily average 24-h squared vorticity forecast error was selected. Table 1 shows the different "optimal" length scales used for the 3DVAR covariances for the three different model-error regimes.

Having selected a correlation length scale for each particular agency forecast model, the next step was to scale the $\mathbf{B}^f$ matrix error variance. A parameter was tuned so that the magnitude of the error covariances was approximately the same as the magnitude of the forecast errors. For the first run, the trace of $\mathbf{B}^f$ was chosen subjectively, being roughly equal to the magnitude of the first-guess-field squared vorticity error on the first day of the simulation. After performing a 99-day simulation, the trace of $\mathbf{B}^f$ was modified such that the trace of $\mathbf{HB}^f\mathbf{H}^T + \mathbf{R}$ was equal to the average daily magnitude of the squared innovation vector (the squared difference between the observed values and the first-guess-field values). The values used for the trace of $\mathbf{B}^f$ are given in Table 1. As will be shown in section 2h, the variance value assigned to the $\mathbf{B}^f$ matrix is not as significant as the correlation length scale, as the trace of $\mathbf{B}^f$ is rescaled every time an increment is made so as to match the magnitude of the innovation vector of that day.

f. Ensemble-based error covariances

As noted in the introduction, the ensemble-based flow-dependent forecast error covariance matrix $\mathbf{F}^f$ is given by the outer product $\mathbf{Z}^f\mathbf{Z}^{fT}$. The columns of $\mathbf{Z}^f$ are proportional to ensemble perturbations about a reference forecast. For ensemble generation schemes that stochastically estimate error covariances, such as the perturbed observations technique and random gridpoint perturbations, the reference forecast is taken to be the ensemble mean. For ensemble generation techniques that deterministically estimate error covariances such as the singular vector and the ETKF techniques, the reference forecast is a control forecast initialized with the analysis field produced by Eq. (2).

There is just one "control" analysis made via Eq. (2) for all of the schemes except for the perturbed observations (PO) ensemble generation technique. The PO ensemble generation technique requires that separate analyses be performed for each of the ensemble members. To prevent the ensemble members from converg-

ing to the same solution, each ensemble member is updated using ensemble-based error covariances formed from the set of ensemble members that does not include the member being updated (Houtekamer and Mitchell 1998). One attractive feature of the PO ensemble generation technique is that, in the absence of model error, it provides a true random sample of the error probability density function regardless of the type of forecast error covariance model used in the data assimilation scheme. In contrast, even with a perfect model, the ETKF ensemble generation scheme outlined below is only capable of producing perfectly accurate error covariances when the forecast error covariances used in the data assimilation scheme are entirely ensemble based; that is, when $\alpha = 0$. Here, we outline the ETKF ensemble generation technique in detail, as it has not yet been introduced to the literature. Details of the other schemes are given in the appendix.

g. ETKF ensemble generation scheme

For the ETKF scheme (Bishop et al. 2001), forecast perturbations $\mathbf{Z}^f$ are transformed into analysis perturbations $\mathbf{Z}^a$ by a transformation matrix $\mathbf{T}$, that is, $\mathbf{Z}^a = \mathbf{Z}^f\mathbf{T}$. These analysis perturbations are then added to the control analysis to give the initial conditions for the next ensemble. The transformation matrix $\mathbf{T}$ is chosen in order to ensure that the covariance matrix associated with the transformed perturbations $\mathbf{F}^a = \mathbf{Z}^a\mathbf{Z}^{aT}$ would be precisely equal to the true analysis error covariance $\mathbf{P}^{at}$ if $\mathbf{F}^f = \mathbf{Z}^f\mathbf{Z}^{fT}$ were precisely equal to the true forecast error covariance matrix $\mathbf{P}^{ft}$. Since

$$\mathbf{P}^{at} = \mathbf{P}^{ft} - \mathbf{P}^{ft}\mathbf{H}^T(\mathbf{HP}^{ft}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{HP}^{ft}, \quad (6)$$

(see, e.g., Daley 1991), $\mathbf{T}$ must satisfy

$$\mathbf{P}^a = \mathbf{Z}^a\mathbf{Z}^{aT} = \mathbf{Z}^f\mathbf{T}\mathbf{T}^T\mathbf{Z}^{fT}$$
$$= \mathbf{Z}^f\mathbf{Z}^{fT} - \mathbf{Z}^f\mathbf{Z}^{fT}\mathbf{H}^T(\mathbf{HZ}^f\mathbf{Z}^{fT}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{HZ}^f\mathbf{Z}^{fT}. \quad (7)$$

As shown in Bishop et al. (2001), this is achieved by letting

$$\mathbf{T} = \mathbf{C}(\Gamma + \mathbf{I})^{-1/2}, \quad (8)$$

where $\mathbf{C}$ contains the eigenvectors, and Γ the eigenvalues, of the matrix product $\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HZ}$. Tippett et al. (2003) have shown that this transformation is equivalent to the ensemble update methods subsequently proposed by Anderson (2001) and Whitaker and Hamill (2002), provided that no error covariance localization is employed. Having transformed forecast perturbations into analysis perturbations in a manner consistent with (6), the analysis perturbations are added to the control analysis to create a new set of initialized ensemble members. These ensemble members are then propagated forward in time using the nonlinear model. At the next observation time, the process is repeated. Thus, the ETKF continuously recycles ensemble perturbations in a manner consistent with (6). Note that Eqs. (6)–(8) assume that the data assimilation

scheme uses a purely ensemble-based error covariance model ($\alpha = 0$). Consequently, the mismatch between ETKF ensemble covariances and the error covariances associated with assimilating data with the hybrid error covariance formulation given by (3) is likely to increase as α is increased from 0 to 1.

To generate the initial perturbations for the runs involving recycling the error covariances, the eigenvectors of the climatological $\mathbf{B}^f$ covariance matrix were used. For the ETKF ensemble, the top 16 eigenvectors of the $\mathbf{B}^f$ matrix, rescaled by their eigenvalues, were used as initial perturbations. For the perturbed observations method, the initial perturbations were formed by randomly sampling the top 256 directions of $\mathbf{B}^f$.

h. Maximal likelihood scaling of error covariance and initial perturbation amplitude

Ideally, the forecast error covariance matrix used in data assimilation accurately represents the magnitude and direction of errors of the first-guess field. However, because of small ensemble size and model error, a priori models of error variance are likely to be inaccurate. From maximal likelihood theory (Dee 1995), the most likely value for the trace of ($\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}$) is the squared magnitude of the innovation vector, ($\mathbf{y} - \mathbf{Hx}^f$). We use this fact to make online estimates of the trace of the flow-dependent $\mathbf{F}^f$ and isotropic $\mathbf{B}^f$ estimates of error covariance. To be specific, at each data assimilation time we compute λ and ρ such that the trace of ($\mathbf{H}\lambda\mathbf{F}^f\mathbf{H}^T + \mathbf{R}$) equals the trace of ($\mathbf{H}\rho\mathbf{B}^f\mathbf{H}^T + \mathbf{R}$) equals $|(\mathbf{y} - \mathbf{Hx}^f)|^2$. Using these rescaling parameters, the Hybrid forecast error covariance formulation given by Eq. (3) was replaced by

$$\mathbf{P}^f = (1 - \alpha)\lambda\mathbf{F}^f + \alpha\rho\mathbf{B}^f, \qquad (9)$$

for our experiments. Equation (9) ensures that α gives a clear indication of the relative weightings of ensemble based and isotropic error correlations. The value of ρ depends completely on the magnitude of the innovation vector, as the trace of $\mathbf{B}^f$ is fixed. The value of λ is not as strongly tied to the magnitude of the innovation vector, as ensemble spread varies daily.

Model error, small ensemble size, and imperfections in the various ensemble generation schemes considered in this paper made all of the schemes incapable of producing an $\mathbf{F}^f$ whose trace was consistent with the observed innovation vector magnitude. To partially correct this problem, we let $\mathbf{Z}^a = \mathbf{Z}^a_{\text{raw}}\lambda^{1/2}$ instead of $\mathbf{Z}^a_{\text{raw}}$, where $\mathbf{Z}^a_{\text{raw}}$ represents the perturbations that the ensemble generation scheme would produce in the absence of maximal likelihood rescaling. This rescaling of the initial perturbations attempts to correct the spread of the next set of forecast ensemble perturbations by using the factor that would have produced a perfect spread had it been applied to the previous perturbations.

i. Checks for dependency of results to differing sets of initial conditions

It was found that on any given day, the performance of any particular data assimilation scheme was sensitive to the sequence of random observation errors that had been used in that particular data assimilation cycle. A small error in a critical location could project that error onto the increment, degrading future forecasts. To help separate the skills of the various forecasting schemes used in our experiments, sets of ten 99-day runs were always performed for each hybrid data assimilation technique used. Each of the ten 99-day cycles was given a different seed for the random number generator used to produce the observation error. By doing 10 runs, the beneficial or detrimental effects of random observation errors (which accounted for about 10% of our errors for the ETKF 16-member ensemble) on analysis scheme performance were averaged out.

3. Results

a. The perfect agency model

Since the focus of this study is the performance of hybrid forecast error covariance formulations in the presence of model error, we restrict our discussion of the performance of perfect model hybrids to just two types of ensemble generation schemes: the ETKF and PO.

Table 2 shows the average errors at the analysis and forecast times for the two generation schemes over a range of α values. The error values in the table are the mean of the daily domain average errors from ten 99-day runs, each covering the same 99 days, but started with a different seed value for the random number generator producing observation error. For the 16-member ETKF-generated ensemble, it is the analysis and forecast errors of a control run that are shown. For the 17-member PO-generated ensemble, it is the errors of the mean analysis and errors of a forecast from an integration of the mean analysis that are shown. Thus, both the PO ensemble and ETKF ensemble used the integration of one analysis to make a forecast.

Figure 4 shows large errors and large error variability at the beginning of the simulation. These errors are a result of our choice of initial conditions to begin the 100-day run. The initial conditions for the 100-day simulation had an error appropriate in magnitude for the resolution error and parameterization error simulations. The initial condition error was, however, too large for the perfect model simulation. Included in Table 2 are results from days 11–99, after the impact of the initial condition error had started to wane. A striking difference between the PO and ETKF hybrid results is that while the analysis and forecast errors produced by the PO hybrid become very large for a pure ensemble-based covariance ($\alpha = 0$), the corresponding $\alpha = 0$ performance for the ETKF hybrid is about as good as the PO

TABLE 2. Analysis and forecast squared vorticity errors per grid point for the perfect agency forecast model as a function of alpha.

Vorticity errors 10^{-14} s^{-2}	$\alpha = 0.0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$	$\alpha = 1.0$
Analysis error											
ETKF16	1.934	1.182	1.309	1.406	1.553	1.748	1.973	2.500	2.852	4.502	34.86
ETKF64	1.406	1.221	1.172	1.201	1.221	1.182	1.191	1.221	1.436	1.904	34.86
PO16	39.90	2.568	2.119	1.855	1.865	1.914	1.836	1.865	1.865	2.139	26.40
Forecast error											
ETKF16	2.783	1.797	1.963	2.109	2.354	2.646	3.027	3.789	4.326	6.572	42.16
ETKF64	2.090	1.875	1.807	1.846	1.846	1.787	1.807	1.855	2.197	2.861	42.16
PO16	49.12	3.584	3.115	2.705	2.725	2.764	2.695	2.744	2.783	3.125	35.73
Forecast error—Days 11 to 99											
ETKF16	0.896	0.373	0.406	0.423	0.537	0.626	0.851	1.274	1.641	3.589	36.34
ETKF64	0.338	0.329	0.316	0.324	0.337	0.342	0.371	0.432	0.607	1.093	36.34
PO16	44.98	1.349	1.124	0.785	0.800	0.856	0.834	0.939	0.980	1.186	29.06

hybrid performance for its best α value. While the ETKF hybrid outperforms the PO hybrid for small α values, at larger α values ($\alpha > 0.5$) the ETKF hybrid is clearly worse than the PO hybrid. Recalling that $\alpha = 1$, gives a pure 3DVAR-type scheme with isotropic error correlations, it is clear from Table 2 that both types of hybrid schemes outperform 3DVAR.

We speculate that the ETKF hybrid outperforms the PO hybrid at small α values because the PO ensemble's error covariance estimates are very noisy. Calculating ensemble statistics from just 16 ensemble members and a perfect random sample ensemble error, variance estimates would follow a chi-square distribution with just 16 degrees of freedom. With 16 degrees of freedom the 95% confidence intervals for the variance estimate are about 60% of the true value. In contrast, the ETKF ensemble estimates are not disturbed by random observational error. A relatively poor performance of PO ensemble against square root filter ensembles such as the ETKF was also documented by Whitaker and Hamill (2002). However, the extreme differences between the PO and ETKF for the $\alpha = 0$ case may also be the result of neglecting to set the mean of the perturbations for each observation to zero, or not using a Schur product or other means to localize error correlations. Thus, this result may not necessarily reflect skill of the PO covariances for purely ensemble-based error statistics. Results from Mitchell et al. (2002) and Keppenne and Rienecker (2002) showed that a purely ensemble-based data assimilation scheme using a perturbed observations ensemble produced far better results than indicated here.

The superior performance of the PO hybrid for large α values is, presumably, a consequence of the fact that PO ensembles give noisy error covariance estimates, but with an accurate mean, for all α values. In contrast, the ensemble update equation used by the ETKF assumes that the data assimilation is being performed with an error covariance model based purely on ensemble-based error statistics. For small α values, this is reasonable. For larger α values, this becomes a poor assumption, and the performance of the ETKF hybrid suffers because

the ETKF ensemble error covariance matrix becomes increasingly different to the true forecast error covariance matrix.

In another PO ensemble study, Hamill and Snyder (2000) found that $\alpha = 0.1$ gave the best PO hybrid performance. For our PO ensemble, a larger α value (0.3 for days 11–99 of the simulation, 0.6 for all 99 days) gave the best PO hybrid performance. This difference may simply be due to differences in the models and in observational networks used in the two studies or it may be because our hybrid error covariance formulation was based on (9) whereas Hamill and Snyder's error covariance formulation was based on (3). [Recall that (9) ensures that α gives an unambiguous measure of the relative contributions of ensemble-based and isotropic error correlations whereas (3) does not.] When we reran the experiment using (3) instead of (9), we found $\alpha = 0.1$ to be the optimal α value, the same optimal value as Hamill and Snyder (2000).

To better illustrate the superiority of the hybrid error covariance formulation in the absence of model error, a comparison of the time evolution of the domain-averaged 24-h vorticity forecast error for the 3DVAR and 16-member ETKF hybrid is shown in Fig. 4. Figure 4a gives the result for 3DVAR, while Fig. 4b gives the result for the 16-member ETKF hybrid [using Eq. (9)] with the optimal (for the ETKF ensemble) α value of 0.1. With the hybrid error correlation models, the forecast error reduces by nearly two orders of magnitude in 25 days. In contrast, it takes the isotropic error correlation model (3DVAR) more than 50 days to achieve a similar forecast error reduction. Note also the rapid error reduction rate produced by the ETKF hybrid in the first 10 days of the simulation. This early error reduction is even more striking when one recalls that the ETKF ensemble was initialized at day 1 by random perturbations to an analysis that was 1 day ahead of the true state of the atmosphere. Thus, in this case, the superior performance of the ETKF hybrid over 3DVAR appears to be resilient to poor ensemble initialization.

To summarize, for the perfect model experiment, the

(a)

**Mean Squared Gridpoint Vorticity Forecast Error
Perfect Agency Model -- 3D Var**

(b)

**Mean Squared Gridpoint Vorticity Forecast Error
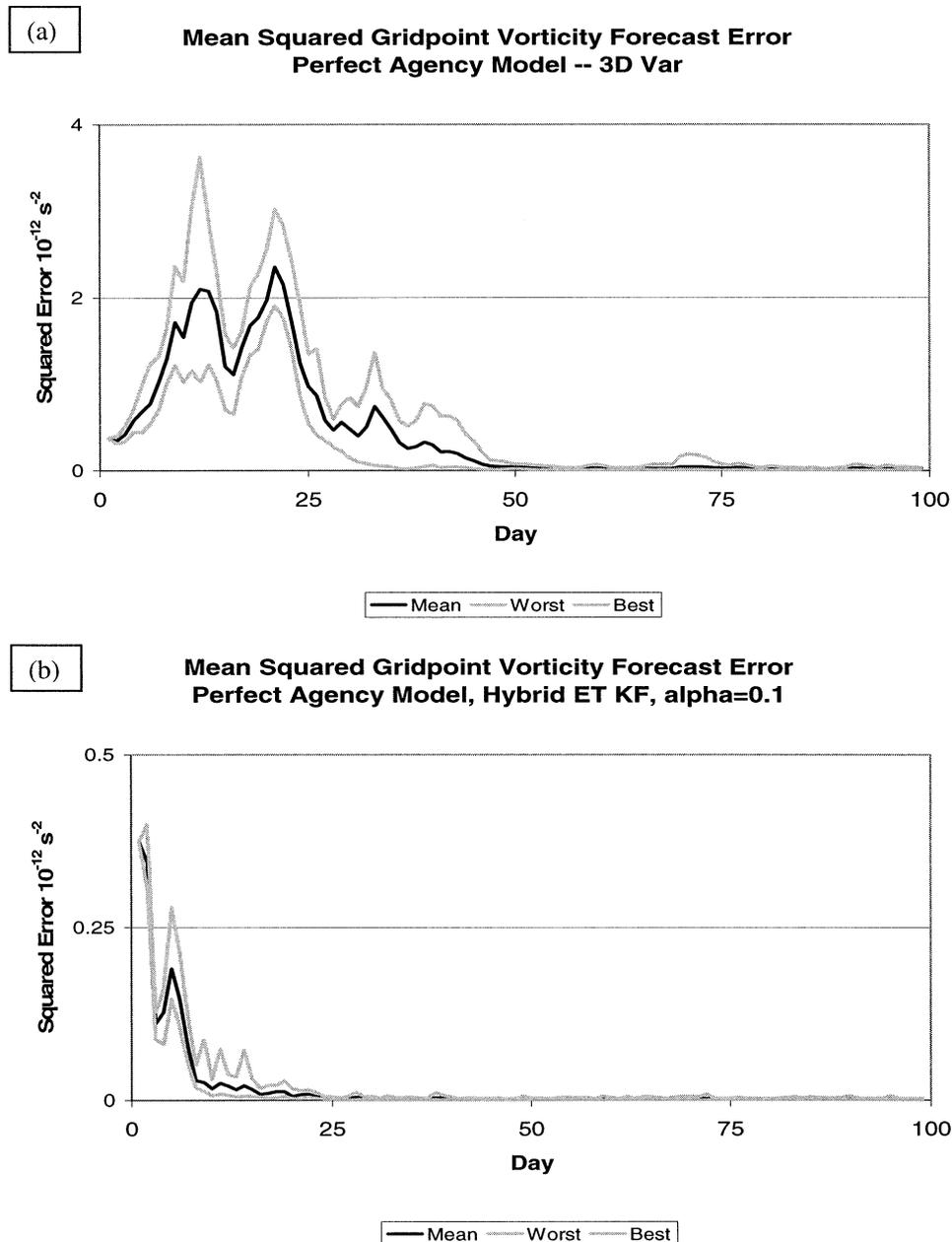Perfect Agency Model, Hybrid ET KF, alpha=0.1**

FIG. 4. Average gridpoint mean squared vorticity error for the perfect agency model run when (a) 3DVAR and (b) the hybrid ETKF with $\alpha = 0.1$ are used for data assimilation. The mean value per day, the greatest value per day, and the smallest value per day for the 10 different runs are presented.

average 24-h forecast error using 3DVAR is 0.444×10^{-12} s^{-2}, using the hybrid data assimilation scheme [$\alpha = 0.1$ for the ETKF ensemble, $\alpha = 0.3$ for the PO ensemble, both using Eq. (9)], the mean-square errors are 0.018×10^{-12} s^{-2} for a control run using a 16-member ETKF ensemble for data assimilation error statistics, and 0.036×10^{-12} s^{-2} for the mean of a 17-member PO ensemble. These values represent mean-square error reductions of 96% and 92%. Removing the first 50 days of the run, when the system was correcting

for the 24-h initialization error of the agency forecast model, the average 24-h forecast using 3DVAR is 2.64×10^{-14} s^{-2}. Using the hybrid data assimilation scheme (again, $\alpha = 0.1$ for the ETKF ensemble, 0.3 for the PO ensemble), this error value is 0.198×10^{-14} s^{-2} (ETKF) and 0.524×10^{-14} s^{-2} (PO). These results represent mean square error reductions of 93% and 80% for the ETKF and PO 16-member ensembles, respectively.

To study the effect of ensemble size, a second set of runs using the ETKF for ensemble generation was done
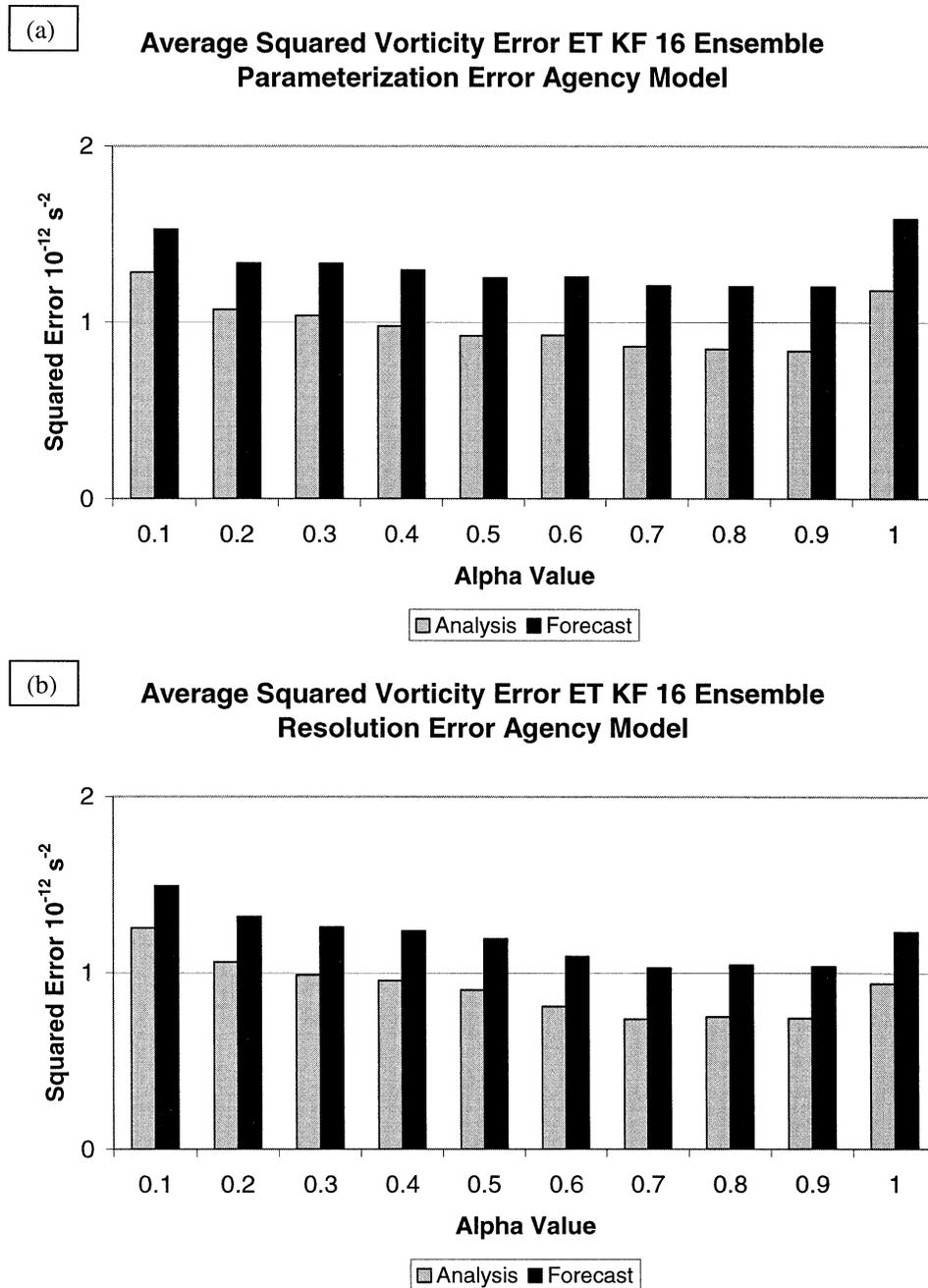
(a)

Average Squared Vorticity Error ET KF 16 Ensemble Parameterization Error Agency Model

(b)

Average Squared Vorticity Error ET KF 16 Ensemble Resolution Error Agency Model

FIG. 5. The daily average squared vorticity error as a function of the parameter alpha ($\alpha = 0$ being pure ETKF, $\alpha = 1$ being pure 3DVAR) for when the agency's forecast model had (a) parameterization error or (b) resolution error. A 16-member ETKF-generated ensemble was used for the construction of the flow-dependent error statistics.

using 64 ensemble members. Neglecting the first 10 days (when the experiment was coming into balance) the 64-member hybrid ETKF outperforms the 16-member hybrid ETKF for all alpha values (Table 2). In addition, the use of a 64-member ETKF ensemble for data assimilation results in the ETKF outperforming perturbed observations over all alpha values.

b. Forecast model with either parameterization or resolution error

The average analysis and 24-h squared forecast error produced with the 16-member ETKF hybrid over a range of α values in the presence of either parameterization or resolution error is shown in Fig. 5. As in

Table 2, values are domain averages over ten 99-day runs, where each of the 10 runs was started with a differing observation error seed. Recall that the parameterization error is produced by a faulty relaxation scheme that controls the way in which meridional variations in zonally averaged vortices are maintained and that the resolution error is a consequence of the ''pseudotruth'' run having been made at half the gridpoint spacing as the agency forecast model.

For the parameterization error (Fig. 5a), the hybrid forecast error correlation model produces superior performance to the purely isotropic correlation model (α = 1) for all α values from 0.1 to 0.9. The best performance was obtained with $\alpha = 0.9$. As previously discussed, the isotropic error correlation model provides a crude representation of forecast error covariance due to model error. Comparing the $\alpha = 0.1$ result with the $\alpha = 0.9$ result shows the substantial improvement that can be obtained in the presence of model error by increasing the weight of the isotropic error correlation model in the hybrid forecast error covariance model. Interestingly, increasing α by just 0.1 from its optimal value takes it to the 3DVAR value of 1.0 that delivers the worst performance of all α values. Thus, the ensemble-based error correlation model produced a substantial reduction in analysis and forecast error even when the weighting given to it in the hybrid scheme was small.

For the resolution error (Fig. 5b), only α values greater than 0.4 produce superior performance to the $\alpha = 1.0$ (3DVAR) performance. An α value of 0.7 produced the smallest average squared forecast error in this case. Again, we see a pronounced difference between $\alpha = 0.9$ and $\alpha = 1.0$ performance, again indicating that a relatively small contribution of ensemble-based error correlations to the hybrid error correlation model can significantly improve data assimilation performance.

Figure 6 shows the time evolution of daily mean square forecast error with $\alpha = 0.75$ for the parameterization error (Fig. 6a) and the resolution error (Fig. 6b). For the parameterization error agency model, there is a maximum in error magnitude near day 15, another around day 30, then a general decrease in error magnitude for the remainder of the run, though not to the extent of the perfect agency model. When 3DVAR is used (not shown), the peak near day 15 is roughly double the value shown in Fig. 6a, a result consistent with the 3DVAR error statistics being poorly tuned for the early flow regime. For the resolution error model, there is again an error magnitude maximum near day 15 (the regime change) and another peak near day 90. Comparison of Fig. 4b with Fig. 6 shows how model error significantly compromises the ability of sparse observational networks to reduce analysis and forecast error.

In considering the relatively large α values required to achieve optimal performance under the two types of model error, it is worthwhile remembering that in the presence of unknown model error, it is impossible to generate an ensemble that is guaranteed to accurately

represent the probability distribution function of forecast errors. Furthermore, there is no rigorous basis to identify an optimal ensemble generation scheme. Indeed ensemble generation schemes that are good at spanning the model subspace in the absence of model error could produce bad error statistics in the presence of model error and vice versa.

Currently, the performance of most operational ensemble generation schemes is evaluated most stringently at lead times of 1–10 days and not at the 6–12-h lead times relevant to data assimilation. These operational schemes include the breeding technique (Toth and Kalnay 1993, 1997), the singular vector (SV) technique (Molteni et al. 1996), and the PO (system simulation) approach (Houtekamer et al. 1996). With the exception of the PO technique, the performance of these schemes in providing an ensemble basis for error covariance modeling has yet to be tested. Here, we test these ensemble generation schemes together with the ETKF and a gridpoint perturbations scheme for their ability to reduce analysis and forecast errors when applied in a hybrid data assimilation scheme. To save time computationally, a fixed value of 0.75 was chosen for α for comparisons between ensemble generation schemes. This value did well for both types of model error and was small enough to allow a significant contribution from ensemble-based error correlations. Note that for the 64-member ETKF ensemble, the optimal α value is smaller than it was for the 16-member ETKF ensemble, 0.5. However, 0.75 was used for α in all comparative experiments.

For the faster ETKF, breeding, and gridpoint perturbations ensemble generation methods, ensembles of 16 and 64 members were used. For the other two ensemble generation methods, only 16 ensemble members were used. Ten 99-day runs were done using each method for both the parameterization and resolution error agency forecast models for the chosen α value of 0.75. The daily mean forecast and analysis squared vorticity gridpoint errors of the 10 runs for each ensemble generation technique were then calculated. Results (Table 3) show that the average 24-h forecast errors are smaller for any ensemble-based data assimilation compared to 3DVAR. This is the case for both forms of model error.

Note the dependency of the effectiveness of the ensemble generation scheme on the type of model error. For the parameterization error runs, the optimal or Hessian SV ensemble produced the lowest mean squared errors, whereas for the resolution error runs the PO ensemble produced the smallest errors. Although our gridpoint perturbations were unlike typical analysis errors in that their amplitude generally decayed with time,[2] they nevertheless produced an ensemble-based error correlation model that was highly effective at reducing analysis and forecast error in the presence of model

---

[2] Errico and Baumhefner (1987) have noted that gridpoint perturbations in the atmosphere also decay.

(a)

Daily Mean Squared Gridpoint Vorticity Forecast Error
Parameterization Error Agency Model
Hybrid ET KF(alpha=0.75)

(b)

Mean Squared Gridpoint Vorticity Forecast Error
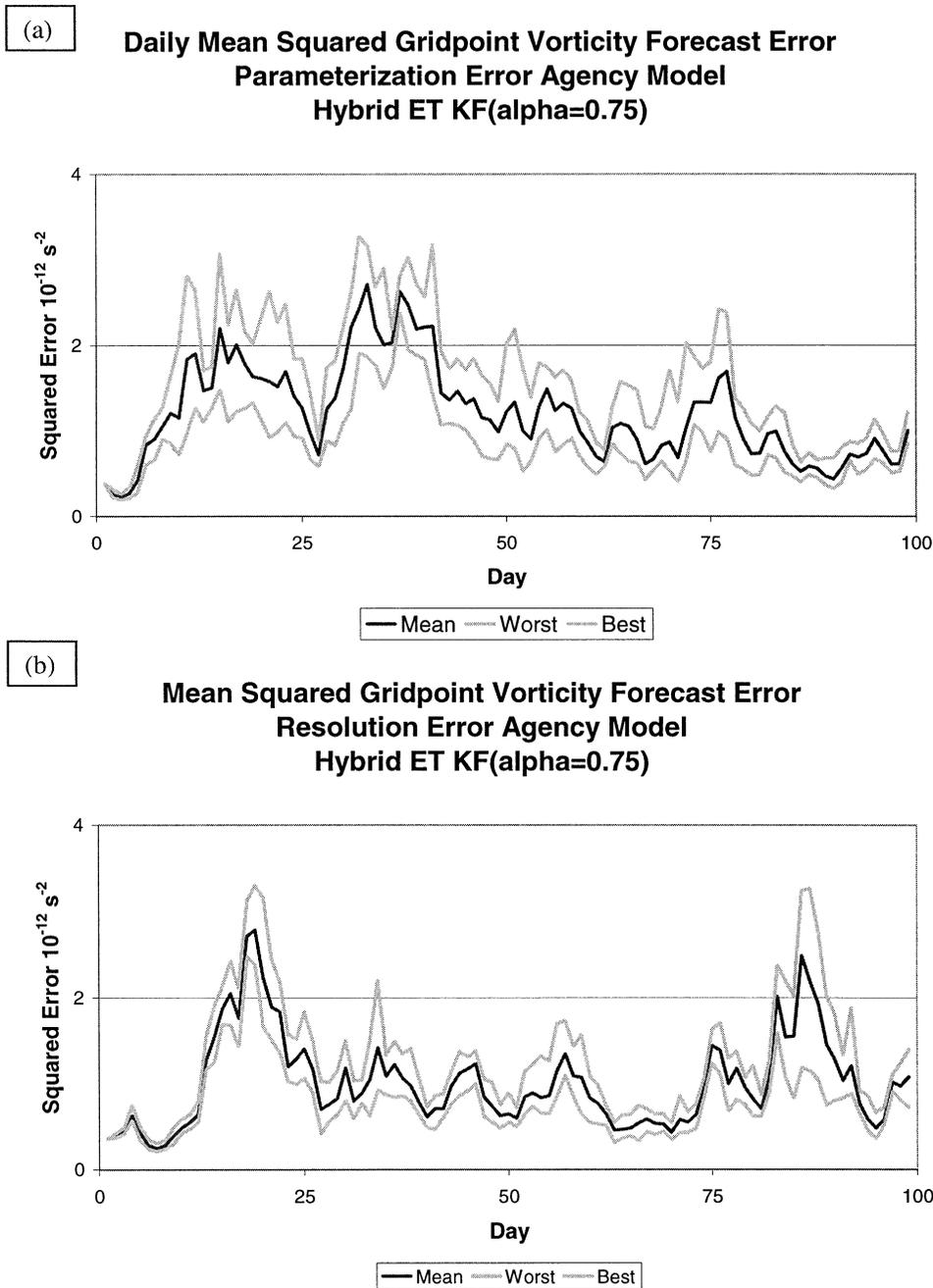### Resolution Error Agency Model
Hybrid ET KF(alpha=0.75)

FIG. 6. Average gridpoint mean squared vorticity error for the (a) parameterization error and (b) resolution error agency model runs when the hybrid ETKF ($\alpha = 0.75$) is used for data assimilation. The mean, largest, and smallest value per day of the 10 different runs is presented.

error. In considering this, recall that our use of maximal likelihood rescaling ensures that the trace of each ensemble-based covariance matrix for a given day is the same, and thus it is the ensemble-based correlations and not the covariances that are tested by our experiments.

For the resolution error experiment, the inclusion of the error of representation in the data assimilation scheme improved forecasts. Average forecast errors for the 64-member ETKF ensemble were 1.068×10^{-12} s^{-2} with no error of representation, 0.961×10^{-12} s^{-2} using the same value for the error of representation at all points, and 0.939×10^{-12} s^{-2} (the value shown Table 3) using the geographically dependent error of representation shown in Fig. 3.

TABLE 3. Analysis and forecast squared vorticity errors and percent improvements over 3DVAR for the experiments that include model error. (Alpha = 0.75.)

Vorticity errors 10^{-12} s^{-2}	3DVAR	ETKF16	ETKF64	Grid 16	Grid 64	Bred 16	Bred 64	Op perts	Pert obs
Parameterization error									
Analysis	1.168	0.865	0.689	0.985	0.667	0.993	0.990	0.638	0.617
Forecast	1.569	1.212	1.013	1.343	0.987	1.369	1.366	0.916	0.940
Error change percent									
Analysis		−26.0%	−41.0%	−15.6%	−42.9%	−15.0%	−15.3%	−45.4%	−47.2%
Forecast		−22.8%	−35.4%	−14.4%	−37.1%	−12.8%	−12.9%	−41.6%	−40.1%
Resolution error									
Analysis	0.930	0.731	0.651	0.719	0.594	0.817	0.792	0.716	0.412
Forecast	1.220	1.017	0.939	1.010	0.898	1.137	1.103	0.988	0.712
Error change percent									
Analysis		−21.3%	−30.0%	−22.7%	−36.1%	−12.1%	−14.8%	−23.1%	−55.7%
Forecast		−16.6%	−23.0%	−17.2%	−26.4%	−6.8%	−9.5%	−19.0%	−41.6%

4. Concluding remarks

The performance of a forecast error correlation model that combines an ensemble-based error correlation model with an isotropic error correlation model has been assessed for three different types of forecast model: a perfect forecast model and forecast models suffering from parameterization or resolution errors. The hybrid error correlation model has been shown to consistently outperform a purely isotropic error correlation model of the type that is typically used in today's operational data assimilation schemes. This result has been found to be resilient to the type of model error considered and to the type of ensemble generation scheme used to generate the ensemble-based error correlations.

With a perfect forecast model, the ETKF ensemble generation scheme yields a hybrid error correlation model that produces mean square errors that are about 33% smaller than the errors produced by a hybrid error correlation model based on the perturbed observations (PO) ensemble generation scheme. Depending on the flow regime considered and whether the PO or ETKF ensemble generation scheme is used, application of the hybrid error correlation model for data assimilation results in mean forecast errors 80%–96% smaller (Table 2) than when an isotropic error correlation is used to assimilate data. In the perfect model experiments, we found that the hybrid error correlation model performed optimally when the weighting ratio of the isotropic error correlation model over the ensemble-based error correlation model was 1/9 ($\alpha = 0.1$); in other words, the optimal hybrid error correlation model was dominated by the ensemble-based error correlation model.

When either parameterization or resolution error was introduced into the forecast model, it was found that for the ETKF hybrid the optimal hybrid error correlation model became dominated by the isotropic error correlation model. Presumably, this is because the isotropic error correlation model provides a better representation of model-error correlation than the ETKF ensemble-based error correlation.

relations. Nevertheless, even with ensemble-based correlations contributing just 10% to the total hybrid error correlation model ($\alpha = 0.9$), the hybrid error correlation model was significantly better than the isotropic error correlation model at reducing forecast and analysis error. Unsurprisingly, the usefulness of any given generation scheme for the ensemble-based error correlations was found to be sensitive to the type of model error present. With $\alpha = 0.75$, an ensemble-based on Hessian SVs was found to be the best for the model with parameterization error, whereas a PO ensemble was found to be best for the model with resolution error (Table 3). These results indicate that any statements about which ensemble generation method and which alpha value give the best results are model and model-error dependent, and cannot be extrapolated to more complex systems.

Houtekamer and Mitchell (2001), Anderson (2001), and Whitaker and Hamill (2002) have pointed out that ensemble-based error correlation models can be improved by localizing one-point error correlation maps via a Schur inner product with an error correlation matrix. Improving our ensemble-based error correlation models and ensemble generation scheme in this way would likely yield hybrid error correlation models that outperform a pure isotropic error correlation model by a larger margin than those considered here. They might also further increase the resilience of the hybrid error correlation model to forecast model error. Use of such error covariance localization might also change the relative rankings of the ensemble error correlation schemes presented here but they probably would not change our finding that the rankings are sensitive to the type of model error present. Indeed, none of the Hybrid schemes (PO, ETKF, etc.) was perfectly optimized, but our central hypothesis still holds: that the use of a hybrid ETKF/3DVAR data assimilation system is superior to either a purely 3DVAR or purely ensemble-based data assimilation scheme. We tested this hypothesis in three different simulations: a perfect model, a model with res-

olution error, and a model with a parameterization error. For all three simulations, the hybrid did indeed perform better than either purely ensemble-based or purely 3DVAR error statistics.

We found that including a geographically dependent error of representation in the data assimilation scheme led to improved analyses and forecasts. Given the strong correspondence between geographical variation of basic-state barotropic instability and the error of representation, it is hard to imagine that the strong geographical variations in hydrodynamic instability (Buizza and Palmer 1995) of the atmosphere do not also correspond to strong geographical variations in the error of representation. We are unaware of any operational center that attempts to estimate or account for such geographical variations in the error of representation. The fact that such variations are neither well known nor taken into account is yet another reason why ensemble error covariance information will remain inaccurate in the near term.

Our finding that the hybrid error correlation model is robustly superior to the isotropic error correlation model also has implications for the use of the ETKF as a tool for selecting optimal locations for supplemental observations. Currently, the accuracy of ETKF predictions of the error variance reducing properties of various possible configurations of the observational network is limited. The limitation is the mismatch between the purely ensemble-based error correlation model assumed by the ETKF and the purely isotropic error correlation model assumed by the schemes that actually assimilate the observations (Bishop et al. 2001; Majumdar et al. 2001). Our findings indicate that the operational centers might gain forecast accuracy by switching to a hybrid error correlation model. Such a switch would enhance the accuracy of the predictions from the ETKF of the error variance reducing potential of observational networks.

APPENDIX

Ensemble Generation Schemes in this Study

a. Perturbed observations

This approach uses 17 independent model runs, rather than 1 control run and 16 perturbations. Each model run is treated as an agency forecast and has its own increment made to produce 17 new analyses. Error covariances are computed using the 16 ensemble perturbations about the mean other than the one for the member being incremented. To represent observation error uncertainty in the ensemble, the innovation vector ($\mathbf{y} -$ $\mathbf{Hx}^f$) has additional random errors added to it for each ensemble member. These errors are in accordance with the values in the observation error covariance matrix, $\mathbf{R}$. The mean scores were taken by averaging 10 simulations, using a different seed for the random errors for each simulation. However, we neglected to adjust the random errors for each observation such that the sum of the 17 random numbers used, one for each member, was equal to zero. After the full collection of ensemble members has new analyses, the control run is set to the mean of the 17 analysis fields. Note that the separate analysis required for each member of the PO ensemble renders this technique more computationally time consuming than the ETKF ensemble generation method.

As with the ETKF approach to ensemble generation, without any representation of model error, the spread of the ensemble members would become smaller than the actual error of the first-guess fields as the analysis cycles continued. The perturbations about the ensemble mean are rescaled at the time the increment is made in accordance with the rescaling parameter. These rescaled perturbations are then added back onto the ensemble mean to generate the 17 different first-guess fields. From there, the analyses for each member are produced. Thus, the rescaling is applied at the analysis time, but in a slightly different manner than for the ETKF ensemble.

b. Breeding method

For each analysis cycle, the difference between the control run and each member are computed. Using maximal likelihood theory, the magnitudes of the ensemble perturbations are then adjusted to ensure that the spread of the 24-h forecast perturbations are consistent with the magnitude of the forecast error. These rescaled perturbations are then added to the control run to generate the ensemble. Like the ETKF, this method is very quick. Its disadvantages include a tendency for all the ensemble perturbations to become parallel to each other and the fact that, in its simplest form, it provides no representation of the distribution of observations. See Wang and Bishop (2003) for a more detailed comparison of the ETKF and breeding ensemble generation techniques.

c. Optimal or Hessian perturbations

Optimal or Hessian perturbations evolve into the leading eigenvectors of the forecast error covariance matrix. Here, we approximate the true optimal perturbations by finding the perturbations that evolve those leading eigenvectors if the distribution of analysis errors implied

by the isotropic error correlation matrix were propagated through time. This distribution is defined by the analysis error covariance matrix $\mathbf{B}^a$, where

$$\mathbf{B}^a = \mathbf{B}^f - \mathbf{B}^f\mathbf{H}^\mathrm{T}(\mathbf{H}\mathbf{B}^f\mathbf{H}^\mathrm{T} + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}^f,$$

and $\mathbf{B}^f$ is the forecast error covariance matrix obtained by applying maximal likelihood theory to determine the variances that should be applied to the isotropic correlation matrix. We used a brute force method to obtain the leading eigenvectors of the forecast error covariance matrix corresponding to $\mathbf{B}^a$. We simply propagated all of the orthonormal eigenvectors of $\mathbf{B}^a$ scaled by the square root of their eigenvalue through time using the nonlinear model to obtain the forecast error covariance matrix. We then performed a singular vector decomposition of this matrix within the ensemble subspace to find the initial perturbations that evolve into the leading 16 eigenvectors of the forecast error covariance matrix. Since experiments indicated that essentially the same optimal perturbations were obtained when we estimated the forecast error covariance matrix using just the top 256 eigenvectors of $\mathbf{B}^a$, for the sake of computational efficiency and for cycling data assimilation purposes, the 16 optimal perturbations were derived from the 256-member approximation to the forecast error covariance. Maximal likelihood rescaling is used to ensure that the spread of the initial conditions, when propagated forward by the forecast model, will equal the expected magnitude of the forecast error.

d. Gridpoint perturbations

A vector of random numbers from a Gaussian distribution, mean zero, variance 1, was formed. The length of this vector was equal to the number of elements in the state vector, such that each element serves as a streamfunction perturbation value at a grid point. The elements of the state vector were multiplied by a factor, based on maximal likelihood rescaling, such that at the forecast time the spread of the ensemble at the observation sites is roughly equal to the magnitude of the innovation vector. This vector was then transformed from streamfunction gridpoint perturbations into vorticity gridpoint perturbations.

REFERENCES

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

Barnes, S., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.,* **3,** 395–409.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.,* **129,** 420–435.

Blanchet, I., C. Frankignoul, and M. A. Cane, 1997: A comparison of adaptive Kalman filters for a tropical Pacific Ocean model. *Mon. Wea. Rev.,* **125,** 40–58.

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

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.,* **126,** 1719–1724.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan,* **75,** 257–288.

——, A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO Physical Space Statistical Analysis System. *Mon. Wea. Rev.,* **126,** 2913–2926.

Courtier, P., and Coauthors, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part I: Formulation. *Quart. J. Roy. Meteor. Soc.,* **124,** 1783–1807.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

——, and E. Barker, 2001: NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.,* **129,** 869–883.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123,** 1128–1145.

Errico, R., and D. Baumhefner, 1987: Predictability experiments using a high resolution limited-area model. *Mon. Wea. Rev.,* **115,** 488–504.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

——, and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeotrophic model. *Mon. Wea. Rev.,* **124,** 85–96.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D-variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905–2919.

Holton, J. R., 1992: *An Introduction to Dynamic Meteorology.* 3d ed. Academic Press, 511 pp.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126,** 796–811.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 123–137.

——, L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.,* **124,** 1225–1242.

Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 1971–1981.

——, and M. M. Rienecker, 2002: Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.,* **130,** 2951–2965.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.,* **82,** 247–268.

Lermusiaux, P. F., and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.,* **127,** 1385–1407.

Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus,* **34,** 505–513.

Majumdar, S. J., C. H. Bishop, I. Szunyogh, and Z. Toth, 2001: Can an ensemble transform Kalman filter predict the reduction in forecast error variance produced by targeted observations? *Quart. J. Roy. Meteor. Soc.,* **127,** 2803–2820.

Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 416–433.

——, ——, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 2791–2808.

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

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical Interpolation Analysis System. *Mon. Wea. Rev.,* **120,** 1747–1762.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.,* **129,** 1194–1207.

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

Sasaki, Y., 1960: An objective analysis for determining initial conditions for the primitive equations. Texas A&M University Tech. Rep. 60-16T, College Station, TX.

Szunyogh, I., Z. Toth, K. A. Emanuel, C. H. Bishop, C. Snyder, R. E. Morss, J. Woolen, and T. Marchok, 1999: Ensemble based targeting experiments during FASTEX: The impact of dropsonde data from the Lear jet. *Quart. J. Roy. Meteor. Soc.,* **125,** 3189–3218.

——, ——, R. E. Morss, S. J. Majumdar, B. J. Etherton, and C. H. Bishop, 2000: The effect of targeted dropsonde observations during the 1999 Winter Storm Reconnaissance program. *Mon. Wea. Rev.,* **128,** 3520–3537.

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

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317–2330.

——, and ——, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* **125,** 3297–3319.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.,* **60,** 1140–1158.

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