Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter*

HERSCHEL L. MITCHELL AND P. L. HOUTEKAMER

Direction de la Recherche en Météorologie, Meteorological Service of Canada, Dorval, Quebec, Canada

GÉRARD PELLERIN

Centre Météorologique Canadien, Meteorological Service of Canada, Dorval, Quebec, Canada

(Manuscript received 9 October 2001, in final form 5 March 2002)

ABSTRACT

The ensemble Kalman filter (EnKF) has been proposed for operational atmospheric data assimilation. Some outstanding issues relate to the required ensemble size, the impact of localization methods on balance, and the representation of model error.

To investigate these issues, a sequential EnKF has been used to assimilate simulated radiosonde, satellite thickness, and aircraft reports into a dry, global, primitive-equation model. The model uses the simple forcing and dissipation proposed by Held and Suarez. It has 21 levels in the vertical, includes topography, and uses a 144×72 horizontal grid. In total, about 80 000 observations are assimilated per day.

It is found that the use of severe localization in the EnKF causes substantial imbalance in the analyses. As the distance of imposed zero correlation increases to about 3000 km, the amount of imbalance becomes acceptably small.

A series of 14-day data assimilation cycles are performed with different configurations of the EnKF. Included is an experiment in which the model is assumed to be perfect and experiments in which model error is simulated by the addition of an ensemble of approximately balanced model perturbations with a specified statistical structure. The results indicate that the EnKF, with 64 ensemble members, performs well in the present context.

The growth rate of small perturbations in the model is examined and found to be slow compared with the corresponding growth rate in an operational forecast model. This is partly due to a lack of horizontal resolution and partly due to a lack of realistic parameterizations. The growth rates in both models are found to be smaller than the growth rate of differences between forecasts with the operational model and verifying analyses. It is concluded that model-error simulation would be important, if either of these models were to be used with the EnKF for the assimilation of real observations.

1. Introduction

The standard Kalman filter provides a conceptual framework for atmospheric and oceanic data assimilation. However, in view of some of its underlying assumptions (e.g., linear dynamics) and its computational requirements, it ''can only be considered as a prototype algorithm'' (Cohn 1997) for these applications.

Evensen (1994) proposed a Kalman-filter-based Monte Carlo approach that is potentially feasible for large atmospheric and oceanic applications. This approach, termed an ensemble Kalman filter (EnKF), uses a forecast model to integrate an ensemble of model states from one analysis time to the next and employs ensemble-based covariances in the analysis step. It is well suited for parallel computation, since each ensemble member can be integrated independently (Evensen 1994; Keppenne 2000).

An efficient algorithm for the analysis step of the EnKF was proposed recently by Houtekamer and Mitchell (2001, hereafter HM01). That algorithm solves the Kalman filter equations by organizing the observations into batches that are assimilated sequentially. The small (and noisy) background-error covariances associated with remote observations are filtered using a Schur (elementwise) product of the covariances calculated from the ensemble and a correlation function having compact support (Gaspari and Cohn 1999). In agreement with Houtekamer and Mitchell (1998, hereafter HM98), it was found that the smaller the ensemble, the more severe was the localization required to cope with noisy co-

*This paper is dedicated to the memory of Dr. Roger Daley, whose scientific insight and congeniality will be sorely missed. The understanding of balance in atmospheric models and the application of Kalman filter theory to atmospheric data assimilation are two of the areas in which Dr. Daley made important contributions.

Corresponding author address: Dr. Herschel L. Mitchell, Direction de la Recherche en Météorologie, 2121 Route Trans-Canadienne, Dorval, QC H9P 1J3, Canada.
E-mail: Herschel.Mitchell@ec.gc.ca

variances. The sequential algorithm was shown to be computationally feasible for synoptic-scale analysis in an operational context, *if* the required number of ensemble members was $O(100)$.

The purpose of this paper is to test the sequential EnKF algorithm of HM01 in a primitive-equation context and to examine some of the issues that remain regarding its suitability for operational atmospheric data assimilation. In particular, how will the required number of ensemble members change as model-state vectors and numbers of observations approach values that occur in an operational context? Balance is also a concern in such a context. Daley (1991, section 6.3) discusses the type of problems that can occur if there is imbalance in a data assimilation cycle. Will any aspect of the EnKF (such as the use of the Schur product for localization) produce imbalance in a primitive-equation context? The Kalman filter framework includes a model-error term and its proper specification is crucial (Dee 1995). A third focus of this paper is how to account for model error in such a way that the balance in a primitive-equation context be maintained.

With regard to the issue of ensemble size, the results of HM98 and HM01 are encouraging. In both studies, it was found that, while analysis error decreased as the number of ensemble members increased, representative ensembles could be maintained even when the number of ensemble members was modest [i.e., $O(100)$]. In part, this was because of the use of a pair of ensembles. Moreover, it was found that the required number of ensemble members was rather similar in the two studies (cf. Fig. 3 of HM98 with Fig. 3 of HM01) even though the number of analysis grid points quadrupled and the number of observations increased by more than a factor of 5 in going from the earlier to the more recent study.

The study by Cohn et al. (1998) of the effects of data selection seems relevant with respect to the issue of balance in the EnKF. They found that local analysis methods (in particular, optimum interpolation) produce analysis increments having an unrealistically large ratio of divergence to vorticity. They concluded that imbalance in the analyses produced by such methods is partly due to the local nature of these methods. This is consistent also with the results of comparisons (e.g., Derber et al. 1991; Gauthier et al. 1999b) between optimum interpolation and (global) variational analysis methods. It follows that we might expect that the imbalance due to the Schur product localization would be largest when the localization is very severe (i.e., when the length scales for the Schur product correlation function are very short). As localization is relaxed, the amount of imbalance can be expected to decrease. It follows that the issues of ensemble size and balance are related; that is, if the use of small ensembles necessitates severe localization, we can expect a significant amount of spatial incoherence, which will manifest itself as imbalance. If, on the other hand, the ensemble is large, then no localization is necessary and the ensemble statistics can be used without modification. In that case, if errors are small enough for their dynamics to be approximately linear, the analyses will be very nearly balanced (Cohn and Parrish 1991). In the case of nonlinear error dynamics, a data assimilation method using a sum of Gaussian kernels (Anderson and Anderson 1999) may be required.

Mitchell and Houtekamer (2000, hereafter MH) developed a method of accounting for model error in an EnKF context by adding an ensemble of realizations of model error to the ensemble of model predictions. Following Dee (1995), the method involved parameterizing the model error and using innovations to estimate the model-error parameters. The study was performed with a three-level quasigeostrophic model. In the present study, that method of accounting for model error is extended to a primitive-equation context. However, the model-error statistics (like the observation-error statistics) are assumed to be known. This eliminates the need to (adaptively) estimate the model error (and the observation error).

In some recent papers (Pham 2001; Anderson 2001; Whitaker and Hamill 2002), the ensemble-based covariances are multiplied by a tunable factor to account for both inbreeding effects in the data assimilation algorithm (HM98) and model error. Here we use a pair of ensembles to reduce the inbreeding effects. To account for model error, we use an additive representation [Ghil et al. 1981, Eq. (2.10a); Daley 1991, Eq. (13.3.7)]. Such a formulation is more general than multiplication by a tunable factor, since the properties of the model error, such as the typical length scale, can be represented independently of the dynamical properties of the ensemble of predictions.

It is known (Petersen 1973; Cohn and Parrish 1991) that if the model error is balanced, then the Kalman filter state estimate will also be balanced. This result from linear theory motivates the approach to be taken in this paper, where a method of generating (approximately) balanced model perturbations will be described and will be used to generate the initial ensemble and to simulate model error. We note that an alternate approach (Ghil et al. 1981; Cohn 1982) involves representing model error as a sum of balanced and unbalanced components. In that case, a projection operator must be incorporated into the assimilation procedure to project the Kalman filter state estimate onto the slow manifold at each analysis time.

To examine the issues of ensemble size, balance, and model-error representation, simulated observations will be assimilated into a dry, multilevel, global, primitive-equation model using a sequential EnKF. The model uses the simple forcing and dissipation proposed by Held and Suarez (1994). This forcing and dissipation were originally proposed to foster the intercomparison of dry dynamical models, but they could also facilitate an intercomparison between different candidate 4D data assimilation methods.

In the next section, we describe the experimental environment including the forecast model, the observational network, and the representation of model error.

2. The experimental environment

Observations will be assimilated every 6 h into a global primitive-equation model. The observations are simulated by interpolating the true state to the observation locations and adding random perturbations. The true state is obtained from a sequence of 6-h integrations with the model. While the experimental environment is conceptually similar to that used in HM98, the forecast model and observational network used here are more like those to be found in an operational environment.

a. The model and the simulated true atmospheric state

Since we envision an eventual operational implementation of the EnKF, we choose to employ a lower-resolution version of the global forecast model that is used operationally at the Canadian Meteorological Centre (CMC). This model is known as the Global Environmental Multiscale (GEM) model. Its hydrostatic formulation has been described by Côté et al. (1998a,b). Here, a 144 × 72 longitude–latitude grid is used in the horizontal. This implies a 2.5° grid spacing. As described by Côté et al., the GEM model integrates the hydrostatic primitive equations using a two-time-level, implicit/semi-Lagrangian time discretization scheme. Our version uses a 1-h time step. To simplify the interpretation and design of the experiments and to facilitate a potential intercomparison between different 4D data assimilation methods, we decided to use a dry version of the GEM model with the forcing proposed by Held and Suarez (1994). The original Held and Suarez calculations did not include topography. Here, as in other more recent studies (e.g., B. Dugas 1999, personal communication; Fox-Rabinovitz et al. 2000), the Held and Suarez forcing is used with topography.

The GEM model uses a terrain-following vertical coordinate defined as

$$\eta = \frac{p - p_T}{p_s - p_T}, \tag{1}$$

where p is pressure, p_s is surface pressure, and p_T is the constant pressure at the top of the model. In this study, p_T is set to 50 hPa. A 21-level version of the GEM model is used here, with levels at $\eta = 0$ (the top level), 0.05, 0.09, 0.14, 0.19, 0.245, 0.304, 0.366, 0.43, 0.494, 0.558, 0.622, 0.684, 0.744, 0.8, 0.85, 0.894, 0.932, 0.964, 0.99, and 1 (the bottom level). In addition to the surface pressure, the model variables are the horizontal wind components, u and v, and the temperature, T. The latter three variables are carried at each model level. The model-state vector, therefore, has 663 552 (=64 × 144 × 72) gridpoint variables.

As described by Côté et al. (1998a, Table 1), the GEM model includes a Laplacian horizontal diffusion on all model variables. Following a suggestion by M. Roch (2001, personal communication), this horizontal diffusion has been removed in the present study, except on the winds in the vicinity of the North and South Poles. Of the many physical parameterizations available in the GEM model, three are used here in conjunction with the Held–Suarez forcing: a weak vertical diffusion (the constant diffusion coefficient equals 0.1 $m^2 s^{-1}$), a weak surface drag, and a dry convective adjustment.

To begin the experiments, we start with the CMC operational analysis for 0000 UTC 2 September 1998, a synoptic time chosen at random. A 29-day integration from this initial state, using our version of the GEM model, brings us to 0000 UTC 1 October 1998. This is the initial time, denoted t_0, for the data assimilation experiments.

The simulation of the true atmospheric state, denoted $\mathbf{X}_t$ (where the subscript t denotes truth), is obtained by extending the 29-day integration 6 h at a time for 14 days. Data assimilation cycles will be performed over this 14-day period. The initial time, t_0, for these cycles is denoted 0000 UTC of day 1 and the final time is 0000 UTC of day 15.

b. The observational network

Every day, observations are taken by radiosondes, satellites, and aircraft at 0000, 0600, 1200, and 1800 UTC. To obtain a realistic network at each of these synoptic times, we employ the following procedure.

For each synoptic time, starting at time t_0, we extract from the CMC operational files a list of the locations of the radiosondes, satellite thickness profiles (SATEMs), and aircraft reports (AIREPs), which had been offered to the operational analysis for that time. For each radiosonde or SATEM location, a simulated radiosonde or SATEM profile is interpolated at that position from $\mathbf{X}_t$. Similarly, at the location of each AIREP, we generate a simulated AIREP. To simulate the (imperfect) observations that are available in a real situation, a random perturbation that has been sampled from a prescribed observation-error distribution is added to each interpolated value [as in Eq. (6) of HM98]. In this study, observation-error distributions for all observations are taken to be independent.

Table 1 shows the number of radiosondes, SATEMs, and AIREPs available on 1 October 1998, the first day of the data assimilation experiments. This allows an impression of the amount of data used in the experiments. As implied by the table, each day during the data assimilation period, there are more than 30 000 observations to be assimilated at 0000 and 1200 UTC and about one-third as many at 0600 and 1800 UTC.

TABLE 1. Number of radiosonde and SATEM profiles and number of aircraft reports used by the EnKF on the first day of the assimilation experiments. Each radiosonde profile yields 40 observations, each SATEM profile yields 5 observations, and each aircraft report yields 3 observations.

	Time (UTC)			
Observations	0000	0600	1200	1800
Radiosondes	638	144	572	98
SATEMs	966	878	865	1087
Aircraft	845	766	1070	975

1) RADIOSONDES

Radiosondes report u, v, and T at the 13 mandatory levels between 50 and 1000 hPa, that is, at 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 925, and 1000 hPa. To facilitate the observation handling, the observation profiles are assumed to be complete irrespective of the height of the topography. Although all observations will be assimilated, we note that the ''analysis below the surface'' will largely be filtered in the interpolation from pressure to η levels. The vertical interpolations are discussed in section 2e.

The assumption that observation errors have no vertical correlation is quite reasonable for u, v, and T observations [Hollingsworth and Lönnberg (1986, Fig. 4); Lönnberg and Hollingsworth (1986, Fig. 4); Mitchell et al. (1990, Figs. 6 and 11)] and implies that the corresponding observation-error covariance matrices are diagonal. The observation-error standard deviations are given in Table 2. The observation errors for wind are from Gauthier et al. (1999b, Table 3, extratropics north winter), while those for temperature were obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF; E. Andersson 2000, personal communication).

Each radiosonde profile also contains a surface pressure observation. The standard deviation of the observation error for this observation is set to 1 hPa following Parrish and Derber (1992, Table 1).

2) SATELLITES

Each SATEM report, that is not located over land, provides five thickness observations. The layers are specified in Table 3. This data usage is consistent with the way SATEM reports were used for many years at ECMWF (Kelly and Pailleux 1988) and the CMC (Mitchell et al. 1996; Gauthier et al. 1999b). The observation-error standard deviations are also given in Table 3. These were taken from Gauthier et al. (1999b, Table 5, clear conditions).

3) AIRCRAFT

Aircraft report u, v, and T at a pressure that corresponds to the flight level. Since our 2.5° horizontal resolution corresponds to the synoptic (rather than the meso-) scale, we do not use the highly automated densely spaced Aircraft to Satellite Data Relay (ASDAR) and Aircraft Communications Addressing and Reporting System (ACARS) reports for generating simulated aircraft reports. Only the more traditional AIREP reports are used. [See Rukhovets et al. (1998) for a recent discussion about the different types of aircraft reports.] The observation errors for the u, v, and T observations are specified to be the same as the corresponding wind and temperature observations for 200-hPa radiosonde observations given in Table 2. These values are indicative of highly accurate aircraft reports.

TABLE 2. Standard deviation of observational error for wind components (m s^{-1}) and temperature (K) observed by radiosondes.

Pressure (hPa)	u; v (m s^{-1})	T (K)
50	2.47	1.4
70	2.47	1.3
100	2.47	1.2
150	2.62	1.2
200	2.62	1.1
250	2.69	1.1
300	2.69	1.1
400	2.47	1.0
500	2.33	1.0
700	2.19	1.0
850	2.19	1.1
925	2.19	1.2
1000	2.19	1.4

c. Description and generation of random perturbation fields

To account for additive model error, we need a procedure to generate a random perturbation field. Such a perturbation field actually consists of u, v, and T perturbations at each model level and a p_s perturbation. An attempt is made to generate these perturbations in such a way that their addition to the true state will have a minimal impact on model balance. To accomplish this, we first generate a random streamfunction perturbation, ψ', as a realization of a multivariate probability distribution. Then u, v, T, and p_s perturbations are derived from ψ' so that the relationship between these perturbations is similar to that assumed for the balanced component of the background errors in 3D variational

TABLE 3. Standard deviation of observational error for SATEM thicknesses (m).

Pressure layer (hPa)	Thickness error (m)
50–100	33
100–300	53
300–500	25
500–700	17
700–1000	27

schemes (Parrish and Derber 1992; Gauthier et al. 1999a,b). This same procedure and statistical description are also used for the generation of an ensemble of first-guess fields valid at t_0.

1) STATISTICAL DESCRIPTION OF STREAMFUNCTION PERTURBATIONS

In general, a realization of ψ' can be generated in terms of M horizontal random fields, where M denotes the number of model levels. The M fields (defined on a 144 × 72 Gaussian grid) can themselves be obtained using the spectral algorithm described in appendix A of MH and the horizontal correlation given by a (degenerate) third-order autoregressive function; that is,

$$\rho_0(r) = \left(1 + \alpha r + \frac{\alpha^2 r^2}{3}\right)\exp(-\alpha r), \tag{2}$$

where r is distance and α is a length-scale parameter. The M horizontal fields, ψ'_m, can then be combined using a vertical cosine expansion:

$$\psi' = \frac{g}{f_0}\sum_{m=0}^{M-1} a_m \psi'_m \cos(m\pi\eta). \tag{3}$$

Such an expansion is motivated by the cosinelike vertical structure of solutions to the linearized primitive equations under certain simplifying conditions [Simmons 1982, Eq. (2); Bartello and Mitchell 1992, Eq. (8a)]. The factor g/f_0, where $f_0 = 2\Omega \sin(45°)$, transforms from height to streamfunction. (Note that Ω is the earth's rotation rate, 7.292×10^{-5} rad s^{-1}, and $g = 9.81$ m s^{-2}.) Note that if α takes the same value for each of the M horizontal fields, we obtain a separable 3D covariance structure [Hollingsworth and Lönnberg 1986, Eq. (4.4)].

In future experiments, we envision using all vertical modes in the expansion (3). However in this paper, the vertical expansion will have only one nonzero coefficient a_m for $m = 0$, 1, or 2. Mode 0 is a barotropic mode that is used to describe background error at the initial time. Modes 1 and 2 are baroclinic modes, which, having baroclinic structure, would seem appropriate for the description of model error or unstable initial perturbations. The length-scale parameter, α, will be set to 10 rad^{-1}, as in HM01. In the data assimilation experiments, the nonzero a_m will be set to 10 m.

2) GENERATION OF AN APPROXIMATELY BALANCED PERTURBATION TO A MODEL STATE

A random perturbation, ψ', can be used to generate perturbations of u, v, T, and p_s that are approximately balanced. The procedure consists of the following steps.

- Working in spectral space, compute wind perturbations, u' and v', from ψ'. Transform to the Gaussian grid and from there interpolate to the 144 × 72 longitude–latitude grid that is used by the model.
- The mass field is obtained through the intermediary of the mass variable, P, defined as

$$P \equiv \Phi + RT^* \ln\frac{p_s}{p_s^*}, \tag{4}$$

  where T^* and p_s^* are a constant reference temperature and surface pressure. Compute P' in spectral space from ψ', using the linear balance equation:

$$\Delta P' = \nabla \cdot f\nabla\psi', \tag{5}$$

  where Δ is the horizontal Laplacian operator and $f = 2\Omega \sin(\text{latitude})$. Transform to the Gaussian grid and from there interpolate to the 144 × 72 longitude–latitude grid. Note that, from (4),

$$P' = \Phi' + RT^* \ln\left(1 + \frac{p'_s}{p_s}\right). \tag{6}$$

- At the surface, $\Phi' \equiv 0$. Equation (6) then reduces to

$$\ln(1 + p'_s/p_s) = P'(\eta = 1)/RT^*, \tag{7}$$

  from which p'_s can be obtained.
- Equation (6) can now be used to obtain the geopotential perturbation, Φ'.
- Finally, T' can be obtained from Φ' using the hydrostatic equation:

$$\frac{d\Phi'}{dp} = -\frac{RT'}{p}. \tag{8}$$

  A multistep method suitable for unequally spaced model levels [developed by J. Côté and described in Béland and Beaudoin (1985, appendix B)] is used to solve (8).

In this way, a perturbation of ψ can be used to generate perturbations of u, v, T, and p_s, which are quasigeostrophically consistent.

The observation error (section 2b) and model error, as described by (2)–(8), are the only sources of error simulated in our system.

d. Generation of ensembles of perturbations

Error statistics are obtained, as needed, from ensembles. So that these remain representative, we utilize a pair of ensembles, as in our earlier work. A brief description of the generation of these ensembles is given here for completeness.

To begin assimilating data, we require a model state $\mathbf{X}_c^f(t_0)$ that represents the best available estimate of the true state $\mathbf{X}_t(t_0)$. In a real situation the latter is, of course, unknown. The state $\mathbf{X}_c^f(t_0)$, where the superscript f indicates that this state is to be used as a forecast (first guess), is generated by adding a random perturbation to the true state [as in Eq. (7) of HM98]. The random perturbation is generated as discussed in section 2c. The model state $\mathbf{X}_c^f(t_0)$ will serve as a central state for the

pair of initial ensembles of first-guess fields, hence the subscript *c*. To obtain the pair of *N*-member ensembles of first-guess fields at time t_0, we add $2N$ sets of random perturbations to $\mathbf{X}_c^f(t_0)$, as in Eq. (8) of HM98. The spread among the ensemble members reflects the uncertainty in $\mathbf{X}_c^f(t_0)$.

Model error, denoted Q, will be simulated in the experimental environment as described in MH (section 2d). The statistical description used to generate $\mathbf{X}_c^f(t_0)$ and the initial ensembles of guess fields will be used also (but possibly with different parameter values a_m) for the model error. Realizations of the model error will be generated every 6 h and added to the pair of ensembles of first-guess fields prior to the assimilation of observations.

Sets of perturbed observations to be assimilated into the ensembles of guess fields are generated every 6 h from time t_0 onward [as in Eq. (9) of HM98]. In fact, two sets of *N* perturbed observations are generated for each simulated observation. As in HM98 (see also Burgers et al. 1998); prior to their addition to the simulated observations, the mean of each set of *N* perturbations is adjusted to zero so that the mean of each set of perturbed observations is equal to the corresponding simulated observation.

It remains now to describe the data assimilation procedure.

e. Data assimilation procedure

To assimilate data, we use the sequential EnKF algorithm described in HM01. This algorithm utilizes a pair of *N*-member ensembles to deal with a problem of inbreeding. Having two ensembles allows the Kalman gain used for the assimilation of data into one ensemble to be computed from the other ensemble.

The analysis is performed on pressure levels as was done for many years at various centers, including the CMC. Such a strategy necessitates vertical interpolations in order to interface the analysis with the forecast model. Figure 1 gives an overview of the entire data assimilation procedure and shows when the vertical interpolations are performed.

1) INTERPOLATIONS

For both the truth integration and the ensemble integrations, the GEM model performs a horizontal interpolation from the analysis (Arakawa A) grid to a staggered model (Arakawa C) grid and back. Using the same experimental procedure, consisting in both cases of a sequence of 6-h integrations (rather than using a long uninterrupted integration for the truth), is a necessary part of a perfect-model design.

Two measures were taken to minimize the effect of the vertical interpolation from η to p. First, an attempt was made to retain the atmospheric vertical structure by using sufficient vertical resolution for the analysis. We

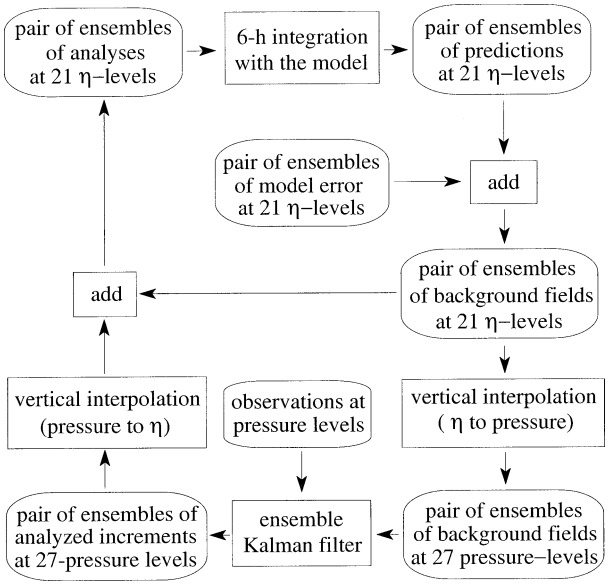

FIG. 1. The procedure used to perform one data assimilation cycle with the sequential EnKF. The coordinate systems employed at each stage of the procedure are indicated.

decided to perform the analysis at the following 27 levels: 50, 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 730, 760, 790, 820, 850, 875, 900, 925, 950, 975, 1000, and 1025 hPa. Second, rather than interpolate the complete analyses from p to η levels, only the analysis increments are interpolated. This has a number of advantages, as discussed for example by Lönnberg and Shaw (1987) and Mitchell et al. (1993). It not only allows features present in the background to be better retained in general, but also ensures that the background remains unchanged in regions where there are no observations.

To go from η to p levels, the algorithm from the GEM model output program is used. This algorithm interpolates u, v, and T linearly in pressure. To generate values below the topography, u and v are assumed to be constant below $\eta = 1$, so the value at the surface is used at any pressure level below the surface. The extrapolation of temperature is performed using a constant lapse rate of 6.5×10^{-3} K m^{-1}.

In general, all horizontal interpolations are performed using bilinear interpolation. However, as explained in HM01, near the edges of the tiles that are introduced to parallelize the algorithm, we perform an extrapolation. (A computationally more expensive alternative would be to employ halo techniques.) The extrapolation proved to be problematic for the surface pressure field, which is highly variable since it depends strongly on the topography. It was decided to use nearest-neighbor (i.e., zero order) interpolation for this field. This is done both for the interpolation of the truth to the observation locations (when generating surface pressure observations) and when performing the forward interpolation of the ensemble members to the surface pressure ob-

servations (when the latter observations are assimilated).

Where there are observations, there is some, perhaps inevitable, interpolation error. We suspect that this is responsible for both the small ''model error'' that seems to appear in our so-called perfect-model experiments (section 4a) and perhaps also for a lack of balance in some analysis increments (section 3b).[1]

2) THE SEQUENTIAL ENKF

Horizontal localization is achieved, in the EnKF algorithm of HM01, by performing a Schur (elementwise) product of the covariances calculated from the ensemble and a correlation function having compact support. No other localization, either in the vertical or between variables, has been added in the current multilevel multivariable context; all interlevel and intervariable covariances are used directly as calculated from the ensembles.

The sequential EnKF has a number of adjustable parameters. Several of these relate to the formation of batches of observations. In this study, at most three regions of observations are permitted per batch; r_0, the radius of the observation regions, is set to 500 km; and $p_{\max}$, the maximum number of observations in a region, is set to 200. All of the experiments are performed with the parallel sequential EnKF using 8×8 tiles. The effect of varying the remaining adjustable parameters, in particular the ensemble size and the length scale used for localization, will be a focus of the experiments. Following HM01, we define r_1 to be the distance at which the correlation function used in the Schur product vanishes.

f. Total energy norm

To facilitate the presentation of the results, it is convenient to introduce a total energy norm, similar to that which has been used in a number of previous studies, for example, Ehrendorfer and Errico (1995). This permits the results for the different model levels and variables to be concisely summarized. We define the total energy norm as the square root of E:

$$E = \frac{1}{2S} \int_S \int_0^1 \left[u^2 + v^2 + \frac{c_p}{T_r} T^2 + R_a T_r \left(\frac{p_s}{p_r} \right)^2 \right] d\eta \, dS, \quad (9)$$

where c_p and R_a are the specific heat at constant pressure and gas constant of dry air, respectively (with numerical values of 1005.7 and 287.04 J kg^{-1} K^{-1}, respectively). The constants T_r and p_r are a reference temperature and pressure, respectively (with numerical values of 270 K and 1000 hPa). The integration extends over the entire globe, S, and vertical dimension, η. As noted by Ehrendorfer and Errico, (9) should not be interpreted as an actual energy, but rather as a simple quadratic expression that looks like an energy and facilitates the presentation of the results. The root of E has units of meters per second.

3. Balance

Since 4D data assimilation methods are intrinsically more consistent with the atmospheric flow than 3D methods, they improve the possibility of producing balanced analysis increments. It follows that a balancing operator (such as a digital filter or nonlinear normal mode initialization), which is perhaps not entirely consistent with the model dynamics, may not need to be applied. However approximations introduced in applying 4D methods, and in particular here in the EnKF, could compromise this potential advantage. In this section, we first present the methodology to be used to evaluate the degree of balance/imbalance produced by the EnKF. This methodology is then applied to examine how varying the localization length scale in the EnKF affects the degree of imbalance in the analyses. As discussed in section 1, we expect that imbalance may be substantial when localization is severe, but will decrease as localization is relaxed.

a. Methodology

Given a model state, we want to be able to determine to what extent high-frequency gravity wave noise will be excited if that state is used as the initial condition for a model forecast.

To do this, we will make use of a digital-filter finalization technique (Fillion et al. 1995) that is available in the GEM model environment. More specifically, we will perform a 12-h forecast from the given model state using a 30-min time step. (This shorter time step should allow for better resolution of the gravity wave dynamics.) This 12-h forecast yields 25 model states at 30-min intervals. Digitally filtering this sequence of model states yields a filtered central estimate, valid 6 h after the initial time, that should be free of high-frequency noise. This filtered central estimate is then compared with the unfiltered forecast valid at this same time and the magnitude of the difference between the two states will be taken to be a measure of the degree of imbalance in the original model state.

A 6-h cutoff period is used for the digital filter and a Lanczos window is used to reduce the Gibbs oscillations due to the use of a truncated Fourier series. The implementation is therefore the same as that described in Fillion et al. (1995). Figure 1 of that paper shows the digital filter coefficients and the transfer function for a 12-h forecast with a 30-min time step and a 6-h

[1] As part of the ongoing software development for the EnKF, and subsequent to the experiments reported here, we have moved to an analysis procedure that is performed on η levels, rejects observations below the model topography, and uses bilinear (forward) interpolation of the logarithm of surface pressure. This configuration suffers from a qualitatively similar small ''model error.''

cutoff period. As can be seen in that figure, the transfer function is not perfectly sharp; that is, it is only an approximation to a unit step function with a cutoff frequency of 6 h. Therefore, oscillations having frequencies of less than 6 h are not completely removed by the filter and, conversely, oscillations having frequencies of greater than 6 h will be somewhat affected by the filter.

In practice, it is not necessary to store the 25 model states in order to apply the filter. As noted by Fillion et al., ''the filter coefficients are computed prior to the integration and the sum [of the model states] is accumulated in a workspace vector of the same dimension as the model-state vector to produce the filtered estimate valid 6 h after the initial time.'' Also, the unfiltered 6-h forecast is saved during the 12-h integration, so that only a single integration needs to be performed.

The experimental procedure consists of a number of steps. First, a pair of 32-member ensembles is generated at 0000 UTC 1 October 1998, as described in section 2d. The random perturbation fields used to generate this pair of ensembles have the statistical structure described in section 2c. [Experiments will be performed for the three vertical modes $m = 0$, 1, and 2, in turn, with $a_m = 40$ m in (3). A large value of a_m is used here in order to provoke an easily detectable imbalance.] Second, the EnKF is used to assimilate the observations available at 0000 UTC 1 October 1998 into the pair of ensembles. Finally, the 32 members of the first ensemble of the pair are integrated for 12 h and the digitally filtered and unfiltered 6-h forecasts are compared, as described above. Performing the procedure for an ensemble of 32 forecasts yields stable statistics. In fact, for each of the three vertical modes, a series of experiments will be performed in order to examine the effect of varying the length scale used for localization in the EnKF.

b. Results

The three curves in the top panel of Fig. 2 show the root-mean-square (rms) differences between the digitally filtered and the unfiltered 6-h forecasts for vertical modes $m = 0$, 1, and 2 as a function of r_1, the distance at which the correlation function used in the Schur product vanishes. The results are presented in terms of the total energy norm defined in (9). Note that r_1 determines the ''radius of influence'' of each observation. It follows that when $r_1 = 0$, no analysis is performed. In this case, the procedure for measuring imbalance is being applied directly to the ensemble of first-guess fields. Perceived imbalance in this case is due to differences between the approximations used to define balance in generating the ensemble of initial fields, on the one hand, and in measuring imbalance, on the other hand. In the first case, quasigeostrophic relationships are used to approximate balance, while in the second case, an imperfect approximation of a 6-h frequency cutoff is used. It follows that the values plotted in Fig. 2 for $r_1 = 0$ are a baseline for the precision of the procedure.

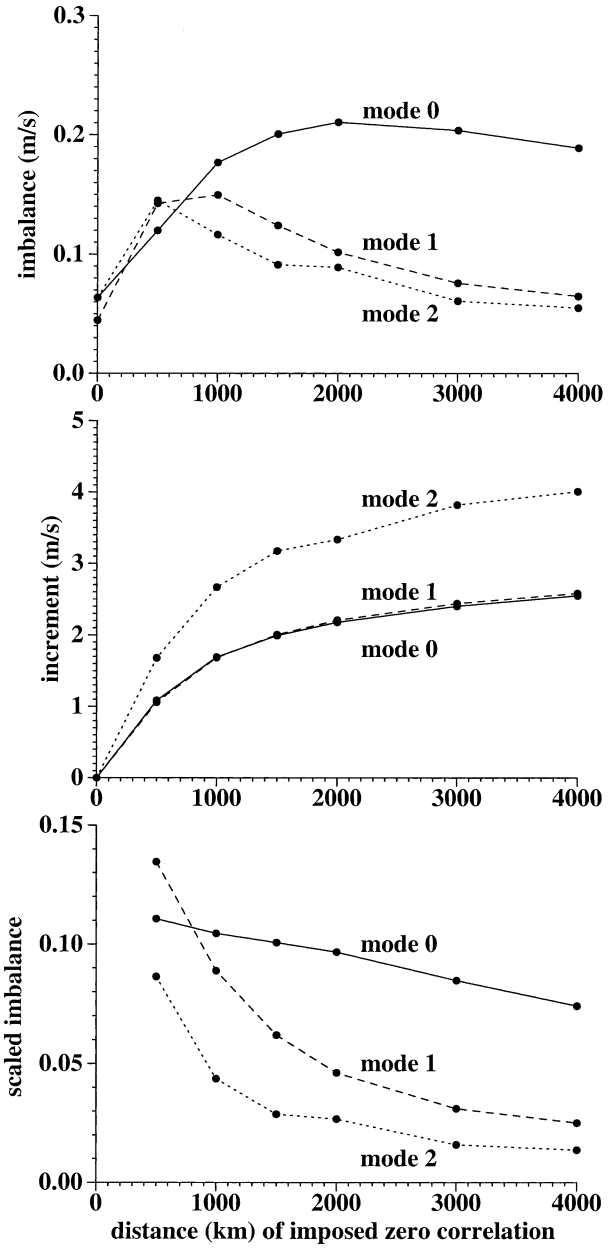

FIG. 2. (top) The imbalance, calculated as described in the text, as a function of r_1, the distance (km) of imposed zero correlation. The imbalance is presented in terms of the total energy norm and averaged over 32 ensemble members. The solid, dashed, and dotted curves show the imbalance when the vertical structure of the background fields is specified using modes 0, 1, and 2, respectively. (middle) As in (top) but for the analysis increment and (bottom) as in (top) but for the imbalance divided by the analysis increment.

It can be seen in the top panel of Fig. 2 that the perceived imbalance for all three of these vertical modes increases fairly sharply as r_1 becomes nonzero. The values of r_1 for which the curves reach their maximum values are 2000 km for mode $m = 0$, 1000 km for mode $m = 1$, and 500 km for mode $m = 2$. As r_1 increases beyond these values, there is a gradual decrease in the

perceived imbalance. In the case of mode $m = 2$, the perceived imbalance has dropped back to its baseline value for $r_1 = 3000$ km. For mode 0, there is little indication of a return to baseline values for large values of r_1. For this mode, unlike what is observed for modes 1 and 2, the surface pressure term (not shown) is fairly large. For surface pressure, the imbalance does not have a peak at small r_1 for any of the three vertical modes studied here. We therefore conclude that the imbalance observed in the surface pressure component is not caused by the localization procedure.

As r_1 increases, each observation has a greater ''radius of influence'' and therefore the magnitude of the overall analysis increment can also be expected to increase. This effect can be seen in the middle panel of Fig. 2, where the total analysis increment is plotted as a function of r_1 for each vertical mode.

An examination of the ratio of the imbalance to the total analysis increment allows one to obtain an impression of the unbalanced fraction of the total analysis increment. This ratio is plotted in the bottom panel of Fig. 2 for the three vertical modes, $m = 0$, 1, and 2. It can be seen that for each of these modes the largest ''scaled imbalance'' occurs at $r_1 = 500$ km, the severest localization for which the calculation was performed. For this value of r_1, we find that 8%–14% of the total analysis increment is unbalanced. As the distance of imposed zero correlation increases beyond this value, the unbalanced portion of the analysis increment is seen to decrease. This decrease is rather rapid for modes $m = 1$ and 2 and more gradual for mode $m = 0$. By the time $r_1 = 3000$ km, the unbalanced portion of the analysis increment is less than 4% for modes $m = 1$ and 2 and is about 8% for mode $m = 0$.

The results of Fig. 2 confirm our expectation that severe localization causes imbalance in the EnKF and that this imbalance can be reduced by relaxing the localization. The results also indicate that as r_1 increases to about 3000 km, the amount of imbalance caused by localization becomes too small to be measured with our experimental procedure. Consequently, no explicit dynamic balancing (such as digital filtering) is used with the Held–Suarez-forced version of the GEM model in the remainder of this study.

4. Data assimilation results

In this section, the performance of the sequential EnKF is examined in a series of data assimilation experiments. A first experiment is performed in a perfect-model context. In subsequent experiments, model error is simulated and the effects of varying the ensemble size and the correlation length scale used in the Schur product are examined.

a. Results in a perfect-model context

In this first data assimilation experiment, N, the number of members in each ensemble of the pair, is set to 32, and r_1, the critical distance beyond which horizontal correlations vanish, is set to 2800 km. To generate the central first-guess state at the initial time, $\mathbf{X}_c^f(t_0)$, and the pair of N-member ensembles about that state, we specify a barotropic vertical structure by setting a_0 to 10 m in (3).[2]

The results of the 14-day data assimilation cycle are summarized in Fig. 3 in terms of the square root of the contributions to the energy norm of (9) due to the wind (top panels), temperature (middle panels), and surface pressure (bottom panels). It can be seen that the wind contribution dominates the contributions due to temperature and surface pressure. It follows that the total energy norm itself (not shown) looks very much like the wind contribution, in this case.

Two measures of error are shown in each panel of Fig. 3: (i) the rms difference between the ensemble mean and the true state (dashed curve) and (ii) the rms spread in the ensemble (solid curve). Results for both ensembles of the pair are shown, since the two ensembles give somewhat different impressions of filter performance during the latter part of the assimilation period. In particular, while all three panels relating to the first ensemble indicate an increasingly serious problem of filter divergence, of the three panels relating to the second ensemble only the surface pressure panel gives this impression. These results suggest that there is some difference in the treatment of the surface pressure between the data assimilation cycle (Fig. 1) and the simulation of the true atmospheric state. This would cause the ensemble spread to differ from the ensemble mean error as observed in Fig. 3 and would justify the addition of a small model-error term to account for this difference.[3]

A number of further observations can be made about the results in Fig. 3. All variables, and especially the rms ensemble spread of wind and temperature, exhibit the typical sawtooth pattern indicative of error growth during the forecast and error reduction at analysis times. It can be seen that the error reduction is significant at 0000 and 1200 UTC, but is very much smaller at 0600 and 1800 UTC. This is consistent with Table 1 and the importance of radiosonde observations in our simulated network. Also, since the initial ensembles of first-guess fields are generated using only the mode $m = 0$ in (3), the initial temperature contribution to the energy norm is zero (from the hydrostatic equation), as can be seen in the middle panels of Fig. 3. Within about two days,

[2] Under fairly general conditions, including linear dynamics, the Kalman filter is asymptotically stable (Cohn and Dee 1988). The initial specification of the covariances (i.e., the choice of m and the value of a_m) has no lasting effect on filter performance (Daley 1991, p. 382). This is consistent with Fig. 4 of HM98.

[3] An alternative explanation for the filter divergence is that it is caused by inbreeding that may occur in a data assimilation cycle (Houtekamer and Mitchell 1999) or even during the sequential assimilation of different batches of observations at each analysis time (HM01).

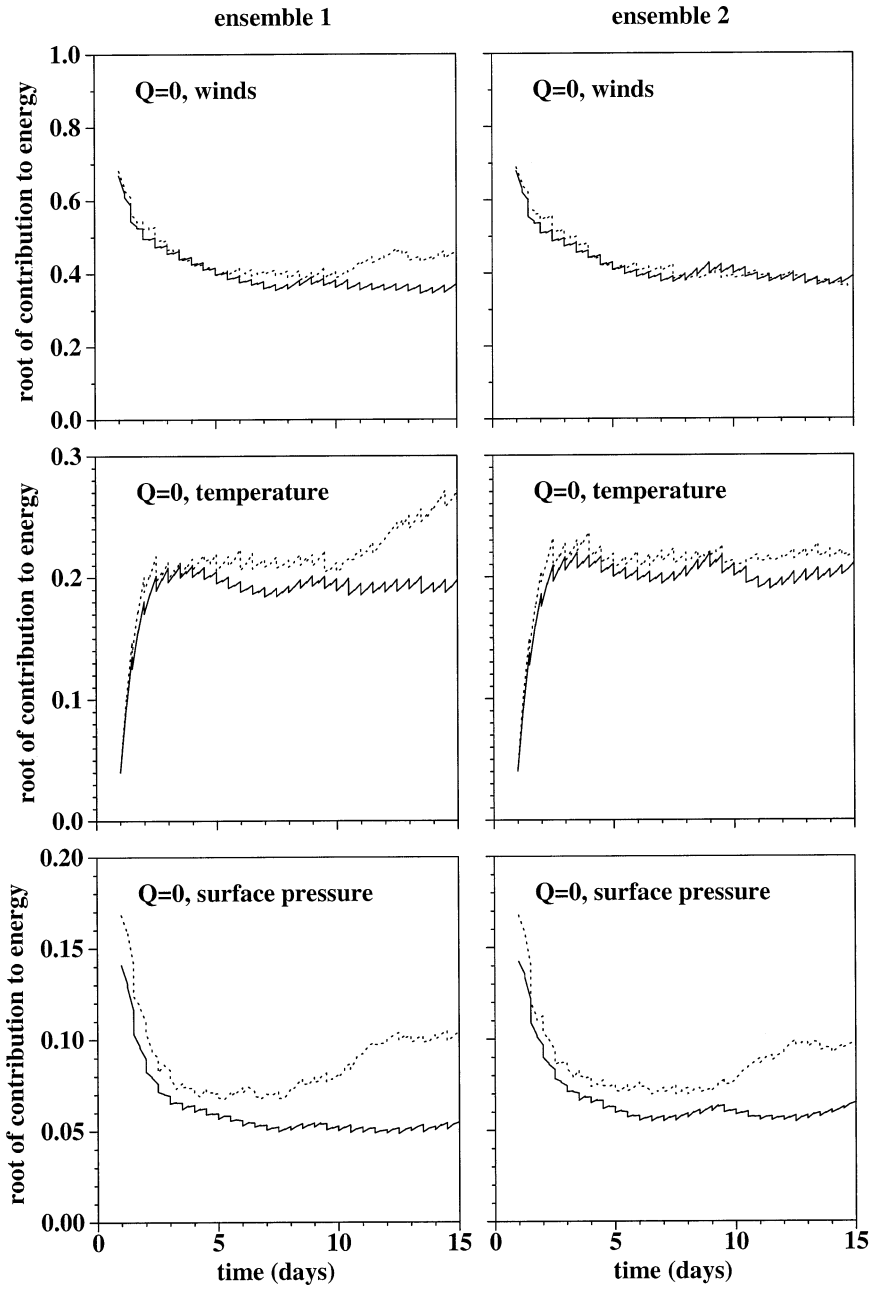

FIG. 3. Square root of the contribution to the total energy for (top) winds, (middle) temperature, and (bottom) surface pressure over the 14-day data assimilation period. Units are m s^{-1}. For this perfect-model experiment, $N = 32$ and $r_1 = 2800$ km. In each panel, the rms error of the ensemble mean is indicated by the dashed curve and the rms spread in the ensemble is indicated by the solid curve. The left-hand panels are for the first ensemble of the pair, while the right-hand panels are for the second ensemble.

the temperature contribution is seen to reach a level commensurate with its final value.

An impression of the vertical profiles of the wind and temperature rms forecast errors toward the end of this data assimilation cycle can be obtained from the left-hand-side panels (denoted Q = 0) of Fig. 4. These profiles of the rms of the ensemble mean error and of the rms spread in the ensemble were obtained by averaging these quantities over both ensembles of the pair over the last 4 days of the cycle. The magnitude of the corresponding surface pressure rms errors is indicated in Table 4. The vertical profiles can be compared with the corresponding profiles from operational data assimilation systems, for example Rabier et al. (1998, Fig. 1),

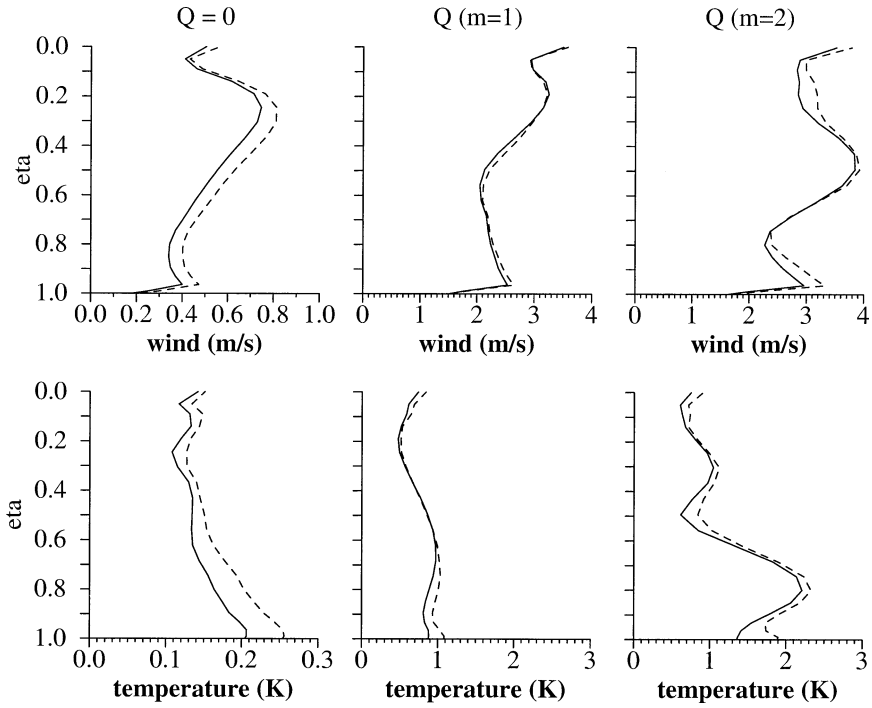

FIG. 4. Vertical profiles of forecast error for vector wind and temperature averaged over days 11–14 of three data assimilation cycles performed with $N = 32$ and $r_1 = 2800$ km. In each panel, the rms error of the ensemble mean is indicated by the dashed curve and the rms spread in the ensemble is indicated by the solid curve. The left-hand panels are for the perfect-model experiment. The center and right-hand panels are for the experiments with simulated model error on vertical modes $m = 1$ and $m = 2$, respectively.

Gauthier et al. (1999a, Fig. 8). Such a comparison shows that the profiles of Fig. 4 have a shape that is quite similar to those from the operational systems: in all cases the temperature profiles (below 50 hPa) are fairly constant, with a relative maximum at the surface, while the vertical wind profiles are dominated by a broad and pronounced maximum in the vicinity of the tropospheric jet. With vector wind errors of less than 1 m s^{-1}, temperature errors of less than 0.3 K, and surface pressure errors of not more than 0.5 hPa, it can be seen that the magnitude of the rms forecast errors in the present experiment is quite small.

The convergence of the EnKF with 2×32 ensemble members to such low error levels in the present perfect-model environment is a noteworthy result. It implies that, to the extent that our forecast model is realistic, the atmospheric flow could be determined quite accurately, given the basic observation network, mostly consisting of radiosondes, that is being simulated here and an EnKF with 2×32 members. To obtain more realistic error levels, a model-error term will now be included.

TABLE 4. Effect of model error (Q) on the surface pressure (hPa) for three data assimilation cycles performed with a pair of 32-member ensembles and distance of imposed zero correlation, r_1, equal to 2800 km. For each of the three cycles, the rms spread and the rms error of the ensemble mean averaged over both ensembles of the pair and over days 11–14 are shown.

	Q = 0	Q(m = 1)	Q(m = 2)
Rms spread	0.28	1.82	1.91
Mean error	0.50	1.81	2.03

b. Effect of simulated model error

To examine the effect of simulated model error of the same order of magnitude as the deficiencies of operational forecast models, we repeat the previous data assimilation cycle but with the addition of model error every 6 h. Thus as indicated in Fig. 1, each time the ensemble pair is integrated 6 h in time, it is now perturbed by the addition of simulated model error, as described in section 2d. In this experiment, the model error is parameterized by setting a_1 to 10 m in (3). All other parameters are the same as those used in the data assimilation experiment in section 4a.

The results of this experiment differ in a number of ways from the results in Fig. 3 and are presented in Fig. 5. Both ensembles of the pair now exhibit similar behavior. As noted and discussed with respect to Fig. 3 of HM98, the ensemble spread behaves ''in a more stable fashion than the more erratic error in the ensemble mean.'' This follows from the fact that the latter is much more susceptible to sampling error, depending as it does

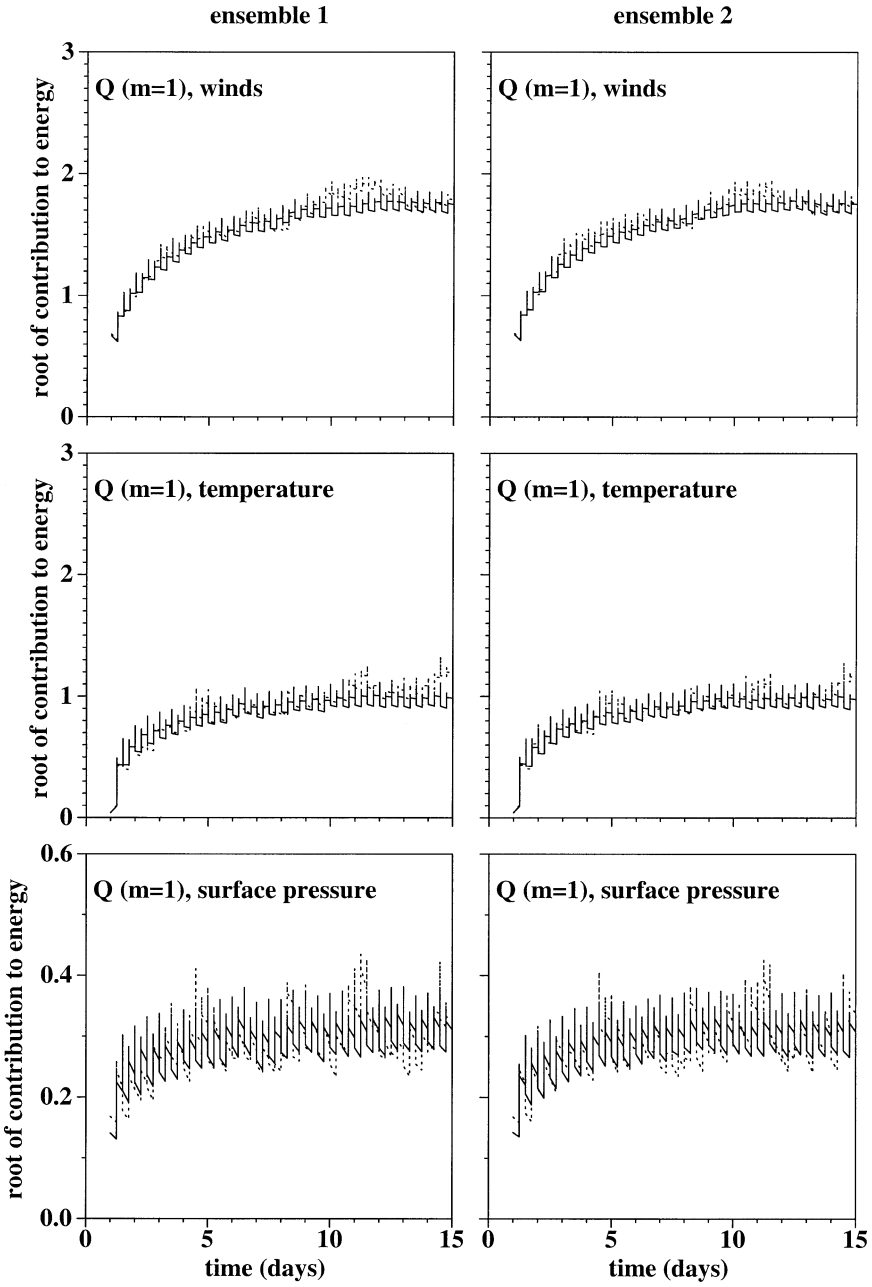

FIG. 5. As in Fig. 3 but including the effect of simulated model error having the vertical structure of mode $m = 1$. Note also the differences in scale of the vertical axes of these panels as compared to the corresponding panels in Fig. 3.

on a single realization of model error and observation error, than is the ensemble spread, which is based on an ensemble of realizations.

The effect of the model error is clearly seen in Fig. 5 as a spike that perturbs the contributions to the energy norm every 6 h in each panel. As in Fig. 3, the subsequent analysis substantially reduces the error levels at 0000 and 1200 UTC, while the error reduction at 0600 and 1800 UTC is more modest. In contrast to the generally decreasing error levels that were observed in Fig. 3, all three contributions to the total energy are seen to increase from their initial values in Fig. 5. By about day 10, the errors in each panel (especially of the more stable rms spread) have achieved a quasi-steady-state level. Looking at local values of the rms spread (not shown), as opposed to a global norm, significant day-to-day variations can be seen. Such flow-dependent statistics have already been shown by HM98 and are not a focus of the present paper.

Similar results are obtained in a second experiment

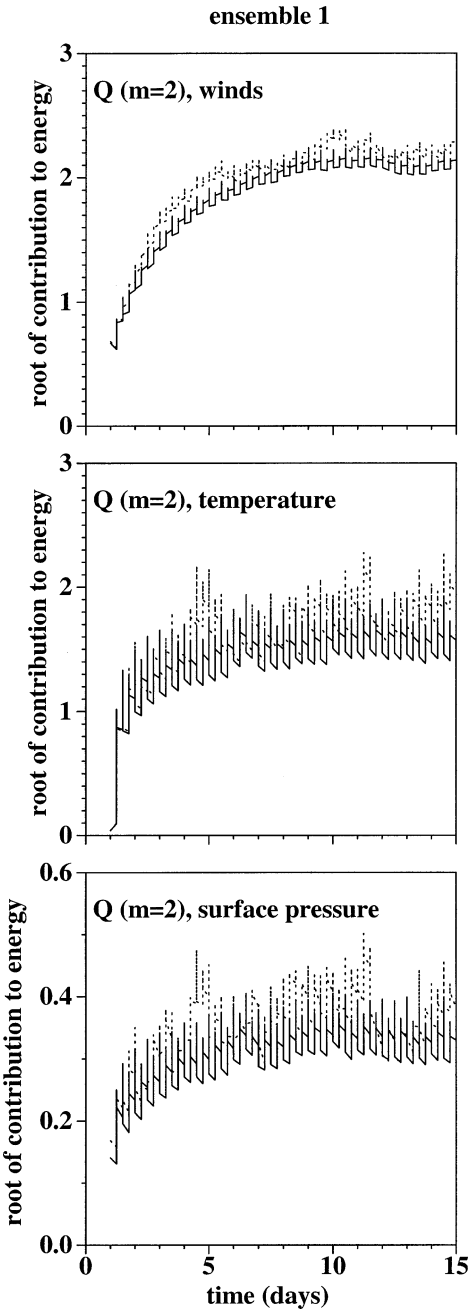

FIG. 6. As in Fig. 5 but the simulated model error is generated having the vertical structure of mode $m = 2$. All results are for the first ensemble of the pair.

with simulated model error. The results (for the first ensemble) are shown in Fig. 6. For this experiment, the vertical structure of the model error was changed by setting a_2 to 10 m in (3). In comparison with Fig. 5, it can be seen that this change in the model error structure results in somewhat increased error levels overall and a tendency for the ensemble spread to underestimate the error in the ensemble mean. It is interesting to note that error levels (especially of temperature and surface pressure) tend to decrease during the 6-h forecast in both Figs. 5 and 6. This indicates that the model-error component projects only weakly on the unstable modes[4] and is consistent with the initial decrease of temperature and surface pressure perturbations in Fig. 7 (to be discussed in section 5). This is in contrast to the perfect-model experiment (Fig. 3), where in agreement with Toth and Kalnay (1993), the analysis cycle is seen to be a ''breeding ground'' for fast-growing modes.

The vertical profiles of the wind and temperature rms forecast errors averaged over the last 4 days of each of these data assimilation cycles are presented as the middle and right-hand-side panels of Fig. 4. The magnitudes of the corresponding surface pressure rms errors are given in Table 4. As indicated above, it can be seen that the agreement between the rms spread and the error of the ensemble mean is substantially better for these two experiments than it was in the perfect-model experiment. Looking at the magnitudes of the errors, it can be seen that they are now significantly larger than was the case in the perfect-model experiment. The simulated model error with $a_2 = 10$ m is seen to produce generally larger errors than with $a_1 = 10$ m. This is a reflection of the larger temperature perturbations in the higher modes of the cosine expansion.

These forecast error vertical profiles can also be compared with those from the operational data assimilation systems (Rabier et al. 1998, Fig. 1; Gauthier et al. 1999a, Fig. 8). Such a comparison indicates a lack of agreement between some of the dominant features in the profiles based on $a_2 = 10$ m (e.g., the wind maximum centered at $\eta = 0.5$ and the pronounced temperature maximum centered at $\eta = 0.8$) and the operational profiles. The result is that profiles obtained with $a_1 \neq 0$ agree better with those from the operational systems than do the profiles obtained with $a_2 \neq 0$. For this reason, it was decided to adopt the $m = 1$ model-error profile for the subsequent data assimilation experiments. It might be noted in passing that the $m = 1$ error magnitudes in Fig. 4 are still smaller than those of the operational systems.

c. *Effect of varying N and r_1*

A series of data assimilation cycles was performed (with $m = 1$ model-error structure) to examine the effect of varying N, the size of each ensemble of the pair, and r_1, the distance beyond which correlations vanish. An examination of the time evolution of the rms errors (not shown) indicated a behavior similar to that shown in Figs. 5 and 6, that is, (i) generally good agreement between the rms spread and the error of the ensemble

[4] In fact, the apparent imbalance of the model-error perturbations suggests that the simplifying conditions of section 2c(1) leading to the cosine expansion and/or the value of the length-scale parameter, α, and/or the use of the linear balance equation are not entirely appropriate.

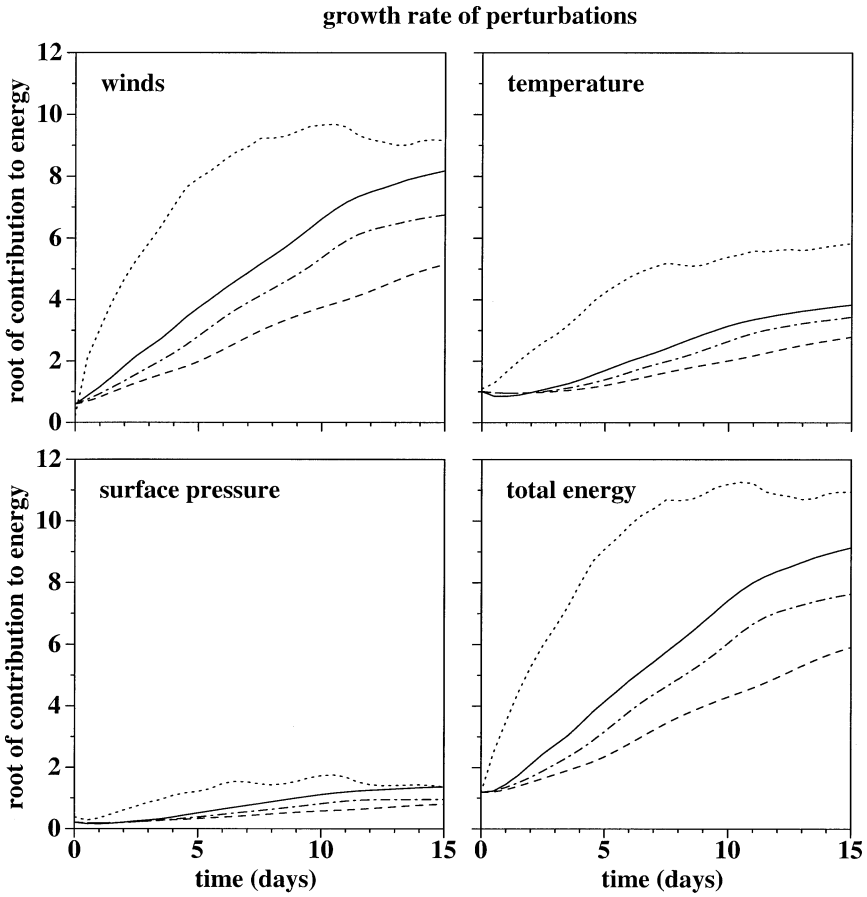

FIG. 7. The dashed curves show the evolution of the spread of a 16-member ensemble of forecasts with the 144 × 72 Held–Suarez-forced model about their mean over the course of a 15-day forecast. The results in the bottom-right panel are presented in terms of the total energy norm. The dashed curves in each of the other three panels show the square root of the wind, temperature, and surface pressure terms of (9). The dashed–dotted and solid curves are similar but for the 400 × 200 Held–Suarez-forced and operational models, respectively. The evolution of the actual forecast error of the ensemble mean of the forecasts with the operational model is shown by the dotted curves.

mean, and (ii) the achievement by day 10 of a quasi-steady-state error level. In view of this second observation, the results of these data assimilation cycles can be concisely summarized by presenting some pertinent statistics for each of these cycles.

Table 5 shows the rms spread and the error of the ensemble mean averaged over days 11–14 for five data assimilation cycles with different EnKF configurations. The configuration with 2 × 32 ensemble members and r_1 = 2800 km is the configuration for which results were presented in Fig. 5. The statistics for this configuration appear in the center of the table and serve as a reference for the other configurations. With respect to ensemble size, results are shown when the ensemble size used in the reference configuration is halved and doubled. Similarly with respect to localization, since r_1 determines the area over which the correlation function used in the Schur product is nonzero, decreasing r_1 from 2800 to 2000 km and then increasing r_1 from 2800 to 4000 km approximately halves and then doubles this area.

We look first at the effect of ensemble size by setting r_1 = 2800 km and comparing the results for the three data assimilation cycles with 2 × 16, 2 × 32, and 2 × 64 ensemble members. Two benefits of larger ensembles are indicated by the results in the table: (i) a decrease in error levels, and (ii) improved agreement between rms spread and ensemble mean error.

We now consider the effect of varying r_1, the critical distance for horizontal localization. For a fixed ensemble size (with 2 × 32 members), we compare results for data assimilation cycles with r_1 = 2000, 2800, and 4000 km. Consider the effect of decreasing r_1 from 2800 to 2000 km. It can be seen that this has the effect of increasing both the rms spread and the ensemble mean error, indicating that of these two values, r_1 = 2800 km is preferable for this ensemble size. Consider now the effect of increasing r_1 from 2800 to 4000 km. It can be

TABLE 5. Summary of the results of five data assimilation cycles with model-error simulation for vertical mode $m = 1$. The results are presented in terms of the total energy norm (m s^{-1}) averaged over days 11–14. For each cycle, the rms spread and the rms error of the ensemble mean are shown for both ensembles of the pair. The five data assimilation cycles explore EnKF performance as the ensemble size and the localization parameter, r_1, vary.

		r_1 = 2000 km		r_1 = 2800 km		r_1 = 4000 km	
		Rms spread	Mean error	Rms spread	Mean error	Rms spread	Mean error
2 × 16	Ensemble 1			2.320	2.464		
	Ensemble 2			2.289	2.366		
2 × 32	Ensemble 1	2.323	2.284	2.143	2.232	2.017	2.277
	Ensemble 2	2.347	2.279	2.138	2.172	2.011	2.245
2 × 64	Ensemble 1			2.010	2.073		
	Ensemble 2			1.996	1.995		

seen that the effect of this change is to decrease the rms spread, while *increasing* the error of the ensemble mean. This type of inconsistency between the rms spread and the ensemble mean error was observed when the localization length scales in Fig. 4 of HM01 were large. There too, the rms spread became unrepresentatively small when the length scale used for localization became too large for a given ensemble size. As shown in Fig. 4 of HM01 (see also Fig. 5 of HM98), for a given ensemble size there is an optimal value of the localization length scale. The present results suggest that for an EnKF with 2 × 32 members operating in the current circumstances, a value of $r_1 = 2800$ km is close to optimal.

In summary, the EnKF with 2 × 32 members performs very well. For this number of ensemble members, a value of $r_1 = 2800$ km seems close to optimal. Increasing the ensemble size results in decreased error levels and, in view of Fig. 4 of HM01, would permit an increase in the value of r_1.

5. Growth rates

The growth rates of errors, due to internal model dynamics, have been measured over a number of generations of the ECMWF operational forecast model (e.g., Simmons et al. 1995; Simmons 1996). It is generally seen (Simmons et al. 1995, Table 2; Simmons 1996, Fig. 16) that the intrinsic error growth rates of the model increase as an unavoidable consequence of the development of a more active and more realistic forecast model. In the ultimate case, in which the behavior of the atmosphere and the forecast model are indistinguishable, true forecast errors will grow at the same rate as differences between two model runs. If one could use such a perfect forecast model for the Kalman filter, one would not need a model-error term in the equations.

It would seem natural to expect that, as more and more small scales are included in the model, the dynamics will become higher dimensional (i.e., there will be more independent directions in phase space along which evolution may occur). Tippett et al. (2000) have argued that the true covariance matrices need to have a good low-dimensional approximation, if ensemble-based methods are to be successful. Thus, we would a priori expect to need more ensemble members or a more severe localization in the case of a more complex model.

Combining the above arguments, we might expect that an EnKF using a low-resolution model could function well with fairly small ensembles. However, the model-error representation would be a very important component of such an EnKF. For a higher-resolution model, more ensemble members and/or a more severe localization would be needed. However, the model-error term would become less dominant.

The experiments in section 4 show an EnKF that behaves well even when used with fairly small ensembles of only 64 members in all. Although this is in itself an encouraging result, it can also be seen as a further indication that the model-error parameterization would have to play a prominent role, if this configuration were to be used for the assimilation of real data.

In this section, we perform error-growth experiments with several different versions of the GEM model that are available to us. These versions include the Held–Suarez-forced model (on a 144 × 72 horizontal grid) used in the data assimilation experiments. The same model is also utilized on a higher-resolution 400 × 200 (i.e., 0.9°) horizontal grid. Finally we perform an experiment with the operational configuration of the GEM model, which also uses a 400 × 200 horizontal grid. This configuration includes the CMC operational physical parameterizations. Although, by June 2001, the operational model had evolved somewhat from the version described by Côté et al. (1998a,b), the placement of the model top at 10 hPa and the use of a 400 × 200 × 28 computational grid had not changed. Forecasts with the operational model have also been compared with the verifying analyses to investigate to what extent the growth of differences between forecasts and analyses resembles the growth of differences between different forecasts.

To examine the error growth of the Held–Suarez-forced model used in the data assimilation experiments, the model state at time t_0 was randomly perturbed by the addition of an ensemble of 16 different realizations

of the separable covariance structure defined by (2) and (3), with a_2 set to 10 m in the latter equation. Each perturbed model state was used as the initial conditions for a 15-day forecast and the spread of these forecasts about their mean was computed. As before, the results are presented in terms of the total energy norm. The dashed curve in the lower-right panel of Fig. 7 shows the evolution of this norm for the spread of the 16-member ensemble about its mean during the 15-day forecast period. The dashed curves in each of the other three panels show the square root of the wind, temperature, and surface pressure terms of (9).

The 16 perturbed model states at time t_0 were each interpolated to a 400 $\times$ 200 longitude–latitude grid and the experiment was repeated with the Held–Suarez-forced model, except now using the higher-horizontal-resolution grid. The spread of these forecasts about their mean was also calculated and is plotted as the dashed–dotted curves in each panel of Fig. 7. It can be seen that with increased horizontal resolution the Held–Suarez-forced model yields significantly faster growth rates. In addition, there is now evidence of error saturation during the last few days of the forecast. It is also interesting to note that, as is the case for the other curves in Fig. 7, the total energy norm (bottom-right panel) is very much dominated by the wind term (top-left panel).

To calculate the perturbation growth rate for the operational version of the GEM model, we obtained the CMC operational analysis for 0000 UTC 1 October 2000 and interpolated this analysis to the 144 $\times$ 72 horizontal grid. An ensemble of 16 model perturbations, identical to the previous ensemble except that the mode $m = 2$ vertical structure now extended from the surface to 10 hPa, was then generated. Each of these model perturbations was added to the operational analysis to yield a model state, which was then interpolated back to the 400 $\times$ 200 grid and then used as the initial conditions for a 15-day forecast. Calculating the spread of these forecasts about their mean and combining the results in terms of the total energy norm, as before, yielded the solid curves in Fig. 7. It can be seen that the use of the operational model results in a further significant increase in the observed growth rates. Some evidence of error saturation can now be seen by day 10 or 11 of the forecast.

Finally, the ensemble mean of the forecasts with the operational model was verified against the operational analyses for the entire forecast period. This yields the dotted curves in Fig. 7. It can be seen that the actual forecast error grows much more quickly (until it begins to saturate) than the perturbation growth rate of the model.

Although these results are only from a single case, and although they may depend on the parameters used to generate the initial conditions, they would seem to indicate that the parameterization of model error would be important, if any of these models were to be used with the EnKF for the assimilation of real observations.

6. Summary and concluding discussion

In this paper, the sequential EnKF proposed in HM01 has been used to assimilate simulated observations into a 21-level, dry, global primitive-equation model. The model uses the simple forcing and dissipation proposed by Held and Suarez (1994) and includes a realistic topography. In the data assimilation experiments, approximately 80 000 observations are assimilated per day. The present study has focused on three related issues that are relevant to the possible suitability of the sequential EnKF algorithm for atmospheric data assimilation. These issues are (i) the required ensemble size, (ii) the potential for imbalance due to localization, and (iii) model-error representation in a primitive-equation context.

As expected, it was found that there can be substantial imbalance in the analyses when localization is severe. However, the imbalance decreases as the localization is relaxed (Fig. 2). As the distance of imposed zero correlations increases to about 3000 km, the amount of imbalance due to localization becomes insignificant. This result guided the choice of the localization parameter for the data assimilation cycles.

A first data assimilation cycle was performed in a perfect-model context. As indicated (e.g., see the left-hand panels of Fig. 4), the error levels achieved are much lower than those typically observed in operational data assimilation systems. This result implies that, to the extent that our experimental environment is realistic, the atmospheric flow could be determined quite accurately given the basic observation network that is being simulated here and an EnKF with 2 $\times$ 32 members. However, even in the present experimental context, we are unable to account for all sources of error and signs of filter divergence become apparent (Fig. 3).

To increase the realism of the experimental context and to protect the EnKF against filter divergence, a simulated model-error component was included in the system. The model error is parameterized, following Dee (1995) and MH. It is obtained by generating an ensemble of random streamfunction perturbations as realizations of a multivariate probability distribution. Then wind, temperature, and surface pressure perturbations are derived from the streamfunction perturbations so that the relationship between these perturbations is similar to that assumed for the balanced component of the background errors in 3D variational schemes.

The initial data assimilation cycle was then repeated but with simulated model error. The results (Figs. 5 and 6) indicate that the EnKF, with a total of 64 ensemble members, performs well in this experimental context. Error levels (middle and right-hand panels of Fig. 4) are now similar to (although still smaller than) those typically observed in operational systems. In future experiments with real observations, the parameters for the model-error description will have to be (adaptively) estimated. For such an application, we envision using a

typical length scale, α, and a vertical expansion, a_m, $m = 0, 1, \ldots M$, such that the corresponding model-error covariance matrix is a tunable fraction of the forecast-error covariance matrix used by the operational 3D analysis procedure (as was suggested in MH).

A series of data assimilation cycles was performed to examine the effect of varying the ensemble size and the localization parameter. The results (Table 5) are consistent with those obtained in HM98 and HM01. In particular, they indicate the benefits of increasing the ensemble size and are consistent with previous results that indicated that for a given ensemble size there is an optimal value of the localization parameter.

The experimental setup in this paper is similar to the one used in HM98, a fact that invites a comparison of the results regarding EnKF performance in the two studies. Given that the number of model coordinates and of observations are both significantly larger in the current study, we might expect a corresponding significant increase in the required number of ensemble members. In fact, no such increase was observed. One possible explanation is that the flow evolution in both cases was dominated by the same synoptic-scale dynamics; that is, these dynamics were already resolved by the T21L3 quasigeostrophic model used in HM98.

To view the current (encouraging) results from a broader perspective, the forecast model used in the current study was compared with an operational forecast model with respect to the growth rate of small perturbations in the two models. The growth rate of differences between the forecasts and the verifying analyses was also calculated. The results, presented in Fig. 7, indicate that the perturbation growth rate in the current model is significantly slower than that of the operational model and that the growth rates in both models are slower than the growth rate of actual forecast errors of the operational model. This suggests that the model-error parameterization will be of critical importance for the EnKF that we are developing to assimilate real observations. As discussed in section 5, these results would also seem to indicate that as EnKF applications progress to more complex models than that used in the present study, more ensemble members or a more severe localization may be required.

The formulation of more efficient ensemble-based data assimilation algorithms is an area of active research. Pham (2001) has presented a method of perturbing observations subject to additional constraints, while Anderson (2001) and Whitaker and Hamill (2002) have presented EnKF algorithms that do not require the observations to be perturbed at all. With these methods, some problems that are due to spurious correlations in small ensembles can be avoided. It would appear that these ideas could also be applied in the context of a pair of ensembles and an additive representation of model error such as is used here, but it has yet to be demonstrated that this can yield a computationally competitive algorithm. Hamill and Snyder [2000, Eq. (5)] have employed an algorithm in which the Kalman gain used by each member of an N-member ensemble is computed from the statistics of the other $N - 1$ ensemble members. This latter approach, designed to avoid the inbreeding problem discussed in HM98, may permit more efficient implementations of the EnKF than the current approach using a pair of ensembles. Future implementations of the EnKF will likely involve trade-offs resulting from the desire to have an efficient algorithm that can use small ensembles and a high quality forecast model to yield balanced analyses and representative error statistics.

Acknowledgments. We are grateful to our many colleagues at Direction de la Recherche en Météorologie and the Canadian Meteorological Centre for their help, suggestions, and encouragement: Bernard Dugas, Sylvie Gravel, and Michel Roch furnished a Held–Suarez version of the GEM model and helped us configure a version with the desired resolution; Luc Fillion suggested the use of a 6-h (rather than 12 h) cycle, and provided advice regarding our procedure for measuring imbalance with the digital filter; Bernard Bilodeau, Alain Patoine, and Michel Roch helped us with the model and its various utilities; Christiane Beaudoin, Pierre Gauthier, and Stéphane Laroche graciously furnished various pieces of code that were used in the procedure to generate balanced model perturbations; Pierre Koclas supplied tools to access the operational observation files; Gilbert Brunet, Bertrand Denis, Claude Girard, and Ayrton Zadra provided useful comments and discussion about some of our preliminary results; and Stéphane Laroche and Monique Tanguay provided thoughtful internal reviews of the manuscript. Finally, we thank the official reviewers for their careful reading of the manuscript.

REFERENCES

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

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

Bartello, P., and H. L. Mitchell, 1992: A continuous three-dimensional model of short-range forecast error covariances. *Tellus,* **44A,** 217–235.

Béland, M., and C. Beaudoin, 1985: A global spectral model with a finite element formulation for the vertical discretization: Adiabatic formulation. *Mon. Wea. Rev.,* **113,** 1910–1919.

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., 1982: Methods of sequential estimation for determining initial data in numerical weather prediction. Ph.D. thesis, Courant Institute of Mathematical Sciences, New York University, 183 pp.

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

——, and D. P. Dee, 1988: Observability of discretized partial differential equations. *SIAM J. Numer. Anal.,* **25,** 586–617.

——, and D. F. Parrish, 1991: The behavior of forecast error co-

variances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119,** 1757–1785.

——, 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.

Côté, J., J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998a: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part II: Results. *Mon. Wea. Rev.,* **126,** 1397–1418.

——, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998b: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.,* **126,** 1373–1395.

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

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

Derber, J. C., D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Wea. Forecasting,* **6,** 538–547.

Ehrendorfer, M., and R. M. Errico, 1995: Mesoscale predictability and the spectrum of optimal perturbations. *J. Atmos. Sci.,* **52,** 3475–3500.

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.

Fillion, L., H. L. Mitchell, H. Ritchie, and A. Staniforth, 1995: The impact of a digital filter finalization technique in a global data assimilation system. *Tellus,* **47A,** 304–323.

Fox-Rabinovitz, M. S., G. L. Stenchikov, M. J. Suarez, L. L. Takacs, and R. C. Govindaraju, 2000: A uniform- and variable-resolution stretched-grid GCM dynamical core with realistic orography. *Mon. Wea. Rev.,* **128,** 1883–1898.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723–757.

Gauthier, P., M. Buehner, and L. Fillion, 1999a: Background-error statistics modelling in a 3D variational data assimilation scheme: Estimation and impact on the analyses. *Proc. ECMWF Workshop on Diagnosis of Data Assimilation Systems,* Reading, United Kingdom, ECMWF, 131–145. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom.]

——, C. Charette, L. Fillion, P. Koclas, and S. Laroche, 1999b: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part 1: The global analysis. *Atmos.–Ocean,* **37,** 103–156.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Applications of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods,* L. Bengtsson, M. Ghil, and E. Källén, Eds., Springer-Verlag, 139–224.

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

Held, I. M., and M. J. Suarez, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Amer. Meteor. Soc.,* **75,** 1825–1830.

Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus,* **38A,** 111–136.

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

——, and ——, 1999: Reply. *Mon. Wea. Rev.,* **127,** 1378–1379.

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

Kelly, G., and J. Pailleux, 1988: Use of satellite vertical sounder data in the ECMWF analysis system. ECMWF Tech. Memo. 143, 46 pp. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom.]

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

Lönnberg, P., and A. Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus,* **38A,** 137–161.

——, and D. Shaw, Eds., 1987: ECMWF data assimilation scientific documentation. ECMWF, 96 pp. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom.]

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

——, C. Charette, C. Chouinard, and B. Brasnett, 1990: Revised interpolation statistics for the Canadian data assimilation procedure: Their derivation and application. *Mon. Wea. Rev.,* **118,** 1591–1614.

——, ——, S. J. Lambert, J. Hallé, and C. Chouinard, 1993: The Canadian global data assimilation system: Description and evaluation. *Mon. Wea. Rev.,* **121,** 1467–1492.

——, C. Chouinard, C. Charette, R. Hogue, and S. J. Lambert, 1996: Impact of a revised analysis algorithm on an operational data assimilation system. *Mon. Wea. Rev.,* **124,** 1243–1255.

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

Petersen, D. P., 1973: Transient suppression in optimal sequential analysis. *J. Appl. Meteor.,* **12,** 437–440.

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

Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Undén, J. Eyre, A. Hollingsworth, and F. Bouttier, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Quart. J. Roy. Meteor. Soc.,* **124,** 1809–1829.

Rukhovets, L., J. Tenenbaum, and M. Geller, 1998: The impact of additional aircraft data on the Goddard Earth Observing System analyses. *Mon. Wea. Rev.,* **126,** 2927–2941.

Simmons, A. J., 1982: The forcing of stationary wave motion by tropical diabatic heating. *Quart. J. Roy. Meteor. Soc.,* **108,** 503–534.

——, 1996: The skill of 500 hPa height forecasts. *Proc. ECMWF Seminar on Predictability,* Reading, United Kingdom, ECMWF, Vol. I, 19–68. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom.]

——, R. Mureau, and T. Petroliagis, 1995: Error growth and estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.,* **121,** 1739–1771.

Tippett, M. K., S. E. Cohn, R. Todling, and D. Marchesin, 2000: Low-dimensional representation of error covariance. *Tellus,* **52A,** 533–553.

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

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