An Adaptive Ensemble Kalman Filter

HERSCHEL L. MITCHELL AND P. L. HOUTEKAMER

Direction de la Recherche en Météorologie, Atmospheric Environment Service, Dorval, Quebec, Canada

ABSTRACT

To the extent that model error is nonnegligible in numerical models of the atmosphere, it must be accounted for in 4D atmospheric data assimilation systems. In this study, a method of estimating and accounting for model error in the context of an ensemble Kalman filter technique is developed. The method involves parameterizing the model error and using innovations to estimate the model-error parameters. The estimation algorithm is based on a maximum likelihood approach and the study is performed in an idealized environment using a three-level, quasigeostrophic, T21 model and simulated observations and model error.

The use of a limited number of ensemble members gives rise to a rank problem in the estimate of the covariance matrix of the innovations. The effect of this problem on the two terms of the log-likelihood function is that the variance term is underestimated, while the χ^2 term is overestimated. To permit the use of relatively small ensembles, a number of strategies are developed to deal with these systematic estimation problems. These include the imposition of a block structure on the covariance matrix of the innovations and a Richardson extrapolation of the log-likelihood value to infinite ensemble size. It is shown that with the use of these techniques, estimates of the model-error parameters are quite acceptable in a statistical sense, even though estimates based on any single innovation vector can be poor.

It is found that, with temporal smoothing of the model-error parameter estimates, the adaptive ensemble Kalman filter produces fairly good estimates of the parameters and accounts rather well for the model error. In fact, its performance in a data assimilation cycle is almost as good as that of a cycle in which the *correct* model-error parameters are used to increase the spread in the ensemble.

1. Introduction

The Kalman filter includes an explicit description of the evolution of the forecast-error covariances in a data assimilation cycle. Given an exact knowledge of all sources of error, linear dynamics, and a number of other conditions (Maybeck 1979, 204–205), the filter constitutes an optimal data assimilation scheme. Unfortunately, this will not be the case if there are inaccuracies in the statistical description of the error sources. Here, for instance, one may think of inaccurate model-error covariances or observational-error covariances. If the inaccuracies are too large, one may observe filter divergence, that is, the false impression that the filter is performing well, while in fact the analyses are diverging from the true state (Maybeck 1979, p. 338; Maybeck 1982, 23–24). Traditional 3D atmospheric data assimilation schemes exhibit similar behavior, if they are given forecast-error covariances that are substantially too small (Daley 1991, p. 146).

To improve the description of model error and protect the Kalman filter against filter divergence, one may try to estimate the model error and adjust the corresponding term in the forecast-error evolution equation. To estimate the model error, the prediction of forecast-error covariances can be compared with actual differences between forecasts and observations (i.e., innovations), as in, for example, Daley (1992b), Dee (1995), and Blanchet et al. (1997). In fact, the use of innovations for covariance estimation is a well-established practice in 3D atmospheric data assimilation (e.g., Rutherford 1972; Hollingsworth and Lönnberg 1986), where one estimates a few parameters that determine the complete description of the forecast-error covariances. In the Kalman filter context, one already possesses an incomplete estimate of the forecast-error covariances from the evolution of the covariances with the model dynamics. It remains then only to estimate the missing model-error component. The expectation is that the analyses will improve as an ever-increasing portion of the error dynamics is accurately described.

Recently, Evensen (1994) proposed a sequential data assimilation method that is on the one hand an approximation to, and on the other hand a nonlinear extension of, the standard Kalman filter. The method, termed an ensemble Kalman filter, avoids the computationally ex-

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

pensive explicit integration of the error-covariance matrix equation. Instead, the required error statistics are calculated from an ensemble of short-range forecasts (i.e., background or first-guess fields). These forecast fields are obtained by integrating the *nonlinear* forecast model, without any need for linear approximations. The accuracy of the calculation increases as the ensemble size increases, as shown in the recent study by Houtekamer and Mitchell (1998, hereafter HM). In this latter study, like that of Evensen (1994), it was found that ensembles having on the order of 100 members were sufficiently large to give reasonable accuracy.

The present study, a follow-up to HM, aims to develop an adaptive technique for estimating model error that can be used in conjunction with an ensemble Kalman filter. In HM, the atmospheric model was taken to be perfect, that is, the model used in the ensemble Kalman filter to integrate the ensemble of analyses was taken to be identical to the model that defined the evolution of the true atmospheric state. The present study will be performed in the same experimental environment as HM, but the perfect-model assumption will be dropped.

While the development of adaptive filters for use in signal processing has been under way for several decades (Haykin 1996), their use in the context of atmospheric and oceanic data assimilation has been hampered by the high dimension of these problems. One candidate for dealing with the problem of high dimensionality is the reduced-order adaptive filter proposed by Hoang et al. (1997a,b). In this approach a parameterized gain matrix is estimated adaptively using adjoint operators. This would seem to be difficult if one has to deal with nonstationary, nonlinear, unstable dynamics and nonstationary observational networks. Hoang et al. (1997a, appendix 2) and Hoang et al. (1997b, section 2) give a comparison of their adaptive filter with the standard Kalman filter. Other possibilities have been examined by Blanchet et al. (1997) in their study of a reduced-space adaptive filter based on empirical orthogonal functions (EOFs). They compared the ability of three different schemes to estimate a relatively large number (from 16 to 112) of components of the model-error covariance matrix. Since this study was performed using an ocean model formulated in terms of EOFs, it was a natural choice to also parameterize the model error using EOFs. However, EOFs are not so commonly used for atmospheric models, it being, for instance, not obvious which norm should be used to generate them or how to close the system to account for neglected interactions (Selten 1993, 1997). While Blanchet et al. indicate that their model dynamics had no unstable modes, such interactions can be expected to become quite important when unstable modes are indeed present.

In this study, we reduce the complexity of the estimation problem as proposed by Dee (1995). In particular, a homogeneous and isotropic formulation in terms of only a small number of parameters (less than 10) will

be adopted for the model error, the values of the parameters will be estimated at each data assimilation time, and the estimates will be smoothed temporally. In addition, following Dee (1995), our estimation algorithm will be based on the maximum likelihood method. Generalized cross validation (Wahba et al. 1995) might have been tried instead as a basis for the estimation algorithm. In a recent comparison, Dee et al. 1999 (see also Dee and da Silva 1999) found that the differences between the estimates produced by the two methods were generally insignificant.

The ensemble Kalman filter configuration described in HM will be used here. This configuration consists of a *pair* of ensemble Kalman filters, configured so that the assimilation of data into one ensemble of short-range forecasts is done with weights calculated from the other ensemble of short-range forecasts. In a series of 30-day data assimilation cycles in HM, this configuration was found to permit representative ensembles to be maintained, even when the ensemble size was rather small. Also, we will continue to use a cutoff radius, beyond which observations are not used, to avoid having to estimate the small correlations associated with remote observations. This feature greatly alleviates the rank problem associated with the ensemble Kalman filter and, as was shown in HM, leads to analyses with smaller error.

The remainder of this paper is organized as follows: in section 2, the experimental environment will be reviewed briefly and the method for simulating and accounting for model error described. Estimation of the model error and some results relating to the evaluation of the log-likelihood function will be discussed in section 3. In section 4, some 30-day data assimilation experiments will be performed in order to evaluate the performance of the ensemble Kalman filter when supplemented by the model-error estimation algorithm. Section 5 consists of a summary and concluding discussion.

2. The experimental configuration

The experimental environment and basic data assimilation algorithm were described in detail in HM. These are briefly reviewed in this section. In addition, we describe the model-error representation and explain how model error is simulated within the experimental configuration and accounted for by the adaptive algorithm.

a. The experimental environment

The experimental environment is basically defined by a nonlinear global model and observational networks for 0000 and 1200 UTC. The nonlinear model is the three-level quasigeostrophic T21 model of Marshall and Molteni (1993). The model reproduces the main features of the Northern Hemisphere winter circulation and, as shown by Vannitsem and Nicolis (1997, Fig. 2) and

Palmer et al. (1998, Fig. 11), has a large number of unstable modes. The simulation of the true atmospheric state, denoted $\mathbf{\Psi}_t(t)$ where the subscript t stands for true, is the same as was used in HM. It was obtained by integrating the model from an initial time t_0 and interrupting the integration every 12 h.

The observational networks (see Fig. 1 of HM) give the locations where simulated radiosonde and satellite soundings are available each day. The observations are simulated by applying random perturbations to the (known) true state at the locations of the observations [as in Eq. (6) of HM]. Radiosondes observe streamfunction and its horizontal derivatives (u and v) at the three model levels (20, 50, and 80 kPa). This yields a total of nine reported values per radiosonde. Each satellite sounding consists of two thickness observations: one for the difference between the streamfunction at 20 and 50 kPa and another for the 50–80 kPa difference. The observation errors for the various observations are given in Eqs. (1)–(3) of HM, from which it can be seen that the radiosonde observations have much smaller errors than the satellite observations. Note also that the observational errors for different soundings are uncorrelated.

In the present study, model error is an additional aspect of the experimental environment. Its representation will now be described.

b. Model-error parameterization

For a model with m_{deg} degrees of freedom, the model-error covariance matrix $\mathbf{Q}$ is, at any time, a full symmetric matrix with $m_{\mathrm{deg}}(m_{\mathrm{deg}} + 1)/2$ different elements. In general, the number of innovations available at any analysis time is very much smaller than this. Online estimation of the full matrix $\mathbf{Q}$ is thus not possible. Ideally, we would like to use online estimation because it simplifies the algorithms and because it allows for rapid adjustment to changing circumstances. Following Dee (1995), we therefore severely reduce the number of degrees of freedom in $\mathbf{Q}$ by parameterization.

Very little is known about the characteristics of the model error in actual operational forecast models. Therefore, we assume that the model error has a relatively simple form with properties similar to those often assumed in statistical interpolation schemes for short-range forecast error. Thus (e.g., Lorenc 1981; Shaw et al. 1987), we assume horizontal–vertical separability and horizontal isotropy for $\mathbf{Q}$. Daley (1992b) has argued that it may be more justifiable to make simplifying assumptions such as these "about the *model*-error covariance than about the *forecast*-error covariance itself because the model error is not sensitive to the inhomogeneities and the time dependencies of the observation network."

Letting i or j indicate horizontal position and k or l indicate level, we postulate that at any time
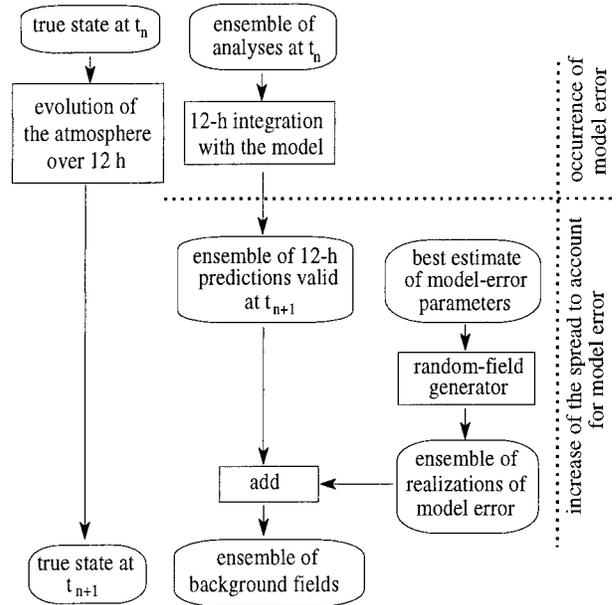

FIG. 1. Setup of an adaptive ensemble Kalman filter in an operational context. The procedure parallels the evolution of the atmospheric state, illustrated in the left column of the figure. The upper part of the figure illustrates the occurrence of model error. The lower part illustrates how the spread in the ensemble is increased to account for model error.

$$\mathbf{Q}(i, j, k, l) = \frac{g^2}{f_0^2} \mathbf{V}_{kl}^q \rho_L(r_{i,j}), \tag{1}$$

where $r_{i,j}$ is great-circle distance. We multiply by the factor g^2/f_0^2, with $f_0 = 2\Omega \sin45°$, in order to obtain model-error covariances for streamfunction. (Note $\Omega = 7.292 \times 10^{-5}$ rad s^{-1} and $g = 9.81$ m s^{-2}). For the three-level model, with levels at 20, 50, and 80 kPa, the model-error vertical covariance matrix (in m^2) is

$$\mathbf{V}^q = \begin{pmatrix} \langle \epsilon_{20}^2 \rangle & \langle \epsilon_{20}\epsilon_{50} \rangle & \langle \epsilon_{20}\epsilon_{80} \rangle \\ \langle \epsilon_{20}\epsilon_{50} \rangle & \langle \epsilon_{50}^2 \rangle & \langle \epsilon_{50}\epsilon_{80} \rangle \\ \langle \epsilon_{20}\epsilon_{80} \rangle & \langle \epsilon_{50}\epsilon_{80} \rangle & \langle \epsilon_{80}^2 \rangle \end{pmatrix}, \tag{2}$$

where the angle brackets, $\langle \cdot \rangle$, denote the expectation operator. We will estimate the six different elements of this symmetric matrix, as described in section 3, subject to the condition that it be positive definite. The isotropic horizontal correlation function, ρ_L, is defined in terms of the length-scale parameter, L, by a second-order autoregressive function (e.g., Daley 1991, p. 117), that is,

$$\rho_L(r) = \left(1 + \frac{r}{L}\right) \exp\left(\frac{-r}{L}\right). \tag{3}$$

The actual estimation of the seven model-error parameters will be performed using a maximum likelihood procedure, to be described in section 3.

c. The basic data assimilation algorithm

The data assimilation algorithm consists of a pair of ensemble Kalman filters. These are configured so that the assimilation of data into one ensemble of 12-h predictions employs the Kalman gain calculated from the other ensemble of 12-h predictions. For an ensemble Kalman filter technique to be functioning well, the spread among the ensemble members must be representative of the difference between the ensemble mean and the true state. The success of an ensemble Kalman filter technique hinges on its ability to maintain such ensembles.

It follows that the ensemble spread must be changed to reflect any and all changes in the ensemble mean that involve uncertainty. To reflect the uncertainty in the simulated observations, we generate randomly perturbed sets of observations every 12 h [as in HM, Eq. (9) and in Burgers et al. 1998, Eq. (12)], for assimilation into the different ensemble background fields. Here also, we use the exact error statistics of the simulated observations when generating these perturbed sets of observations. The use of random (Monte Carlo) perturbations in a high-dimensional phase space is motivated in Stein (1987), Evensen (1994), Houtekamer et al. (1996b), and Anderson (1997).

Model error, like observation error, must also be simulated within the experimental configuration and be accounted for by the ensemble Kalman filter.

d. Model-error simulation

In HM, the ensembles of background fields were obtained by integrating exactly the same model that had been used to produce the simulation of the true atmospheric state. There was, therefore, no model error, and so the forecast error was entirely due to predictability error, that is, errors in the specification of the initial state and their growth due to instabilities in the nonlinear model.

In the more realistic situation considered here, we recognize that atmospheric models are imperfect and we try to account for model error. Assuming that the predictability error and model error are independent, we can write (Daley 1992b)

$$\mathbf{P}_n^f = \mathbf{P}_n^p + \mathbf{Q}_n. \qquad (4)$$

Here $\mathbf{P}_n^f$, $\mathbf{P}_n^p$, and $\mathbf{Q}_n$ are the forecast-, the predictability-, and the model-error covariances and all terms refer to the errors incurred between times $n - 1$ and n.

The way in which an adaptive ensemble Kalman filter might function in an operational environment is illustrated in Fig. 1. The occurrence of model error, caused by the difference between the atmosphere and the model, will act to increase the difference between the ensemble mean 12-h prediction and the true state. The adaptive part of the algorithm, illustrated in the lower part of the figure, would then have the role of increasing
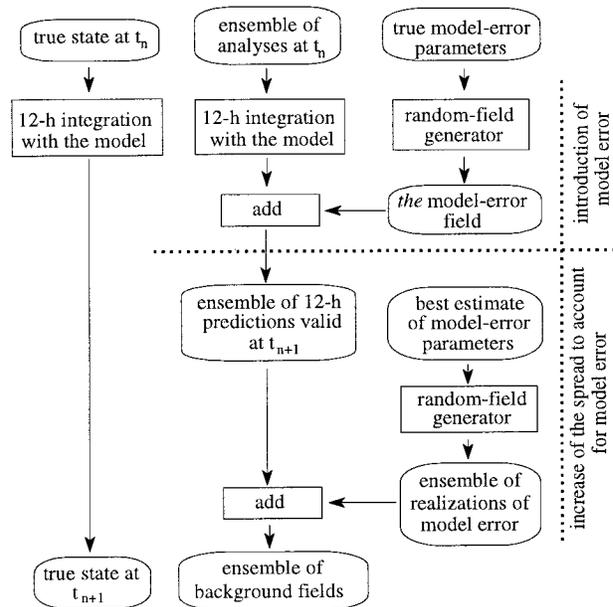

FIG. 2. The upper part of the figure illustrates the procedure used to simulate model error. Given the true model-error parameters, a random field generator is used to produce the model-error field. This field is then added to each member of the ensemble of 12-h predictions. The lower part of the figure illustrates the procedure used to account for model error. As in an operational context, this involves using the best available estimate of the model-error parameters to generate an ensemble of realizations of model error. The entire procedure parallels the evolution of the true state from time t_n to time t_{n+1}, illustrated in the left column.

the ensemble spread to reflect this error. This part of the algorithm uses estimated model-error parameters to generate an *ensemble* of realizations of model error, with imposed zero mean at each grid point, and then adds each member of this ensemble to the corresponding member of the ensemble of 12-h predictions to yield an ensemble of background fields. This new ensemble would have the same mean as the ensemble of 12-h predictions, but an increased spread that reflects the *estimated* model error.

In the present study, the ensembles of background fields will be produced in a way that allows the effect of model error to be simulated in a controlled manner. The procedure for simulating model error is illustrated, for one ensemble of the pair, in the upper portion of Fig. 2. First, an ensemble of 12-h predictions is produced, as in HM. Then the *true* model-error parameters and a random field generator (see appendix A) are used to generate the perturbation field, which constitutes the true model error for this 12-h period. This perturbation field is added to each member of the ensemble. The effect is to change the mean of the ensemble (and therefore also the difference between the mean and the truth) at each model grid point, without changing the spread in the ensemble. The same perturbation field is added

to each member of *both* ensembles, because it represents the true model error.

The adaptive part of the algorithm, illustrated in the lower part of Fig. 2 for one ensemble of the pair, is the same as it would be in an operational context. Here the estimation of model error will be performed in terms of the parameterized representation introduced in section 2b above. In fact, the estimation procedure will be performed separately for each ensemble of the pair, yielding separate estimates of the model-error parameters for each ensemble.

We now turn our attention to the problem of estimating the model-error parameters.

3. Model-error estimation

The method to be used to estimate model error is described in this section and some sensitivity experiments are performed to examine the ability of the algorithm to estimate the model-error parameters from the available data. It appears to be difficult to accurately estimate the log-likelihood function using a small ensemble. For example, the determinant and smallest eigenvalue of the covariance matrix of the innovations tend to be underestimated due to a rank problem even though the estimates of individual elements of the matrix are unbiased.[1] Some techniques, such as imposing a block structure on the matrix and using a Richardson extrapolation to infinite ensemble size, are discussed.

a. Using innovations to estimate model error

As indicated in section 1, model-error estimation is based on the use of innovations. Define $\boldsymbol{\nu}_n$ to be the innovation vector at time t_n, that is, the difference between the observations and the ensemble-mean forecast interpolated to the observations at time t_n. Mathematically

$$\boldsymbol{\nu}_n \equiv \mathbf{d}_n - \mathbf{H}\overline{\boldsymbol{\Psi}_n^f}, \qquad (5)$$

where $\mathbf{H}$ is the (forward) interpolation from a complete model state to the observations and $\mathbf{d}_n$ and $\overline{\boldsymbol{\Psi}_n^f}$ are the observation vector and the ensemble-mean forecast at time t_n. Note that since the ensemble of model-error realizations has zero mean, the ensemble-mean prediction $\overline{\boldsymbol{\Psi}_n^p} = \overline{\boldsymbol{\Psi}_n^f}$ in the present study. Therefore the right-hand side of (5) can be evaluated as soon as the ensemble of predictions is available, as illustrated at the top of Fig. 3. Also since the interpolation operator is perfect,

---

[1] Systematic estimation, or rank, problems may be expected for any quantity that cannot meaningfully be estimated from a two-member ensemble. Note that two is the required minimum in order to have information on both the mean and the covariance. Unbiased estimates of individual matrix elements can already be obtained with just two members.
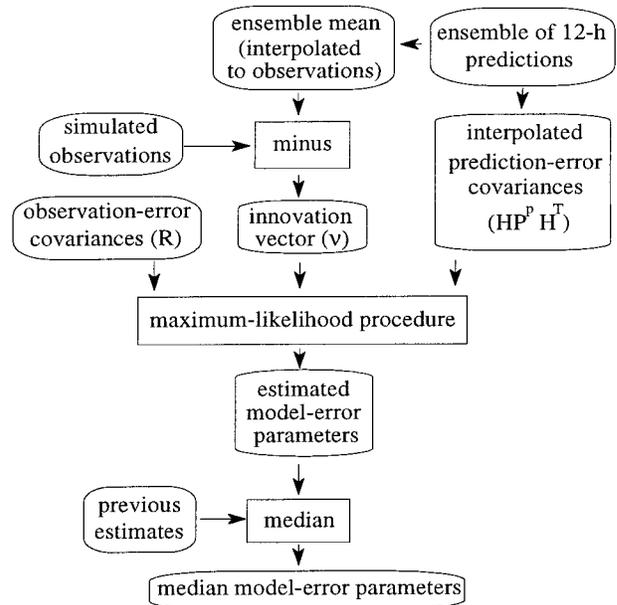

FIG. 3. The procedure used to estimate the model-error parameters. Note the three inputs to the maximum likelihood procedure: $\mathbf{R}$, $\boldsymbol{\nu}$, and $\mathbf{HP}^p\mathbf{H}^T$ [cf. (9) and (10)].

substitution of error-free observations and the true state in (5) gives a zero innovation.

Proceeding as in Dee (1995) [see also Daley (1992b)], we obtain the relation

$$\langle \boldsymbol{\nu}_n \boldsymbol{\nu}_n^T \rangle = \mathbf{HP}_n^f \mathbf{H}^T + \mathbf{R}. \qquad (6)$$

Here $\mathbf{P}_n^f$ and $\mathbf{R}$ are the forecast- and observational-error covariances. To derive (6), we have made the simplifying assumption that the forecast and observation errors are not correlated with each other.

Substitution of (4) into (6) yields

$$\langle \boldsymbol{\nu}_n \boldsymbol{\nu}_n^T \rangle = \mathbf{HP}_n^p \mathbf{H}^T + \mathbf{HQ}_n \mathbf{H}^T + \mathbf{R}, \qquad (7)$$

which has also been discussed by Moghaddamjoo and Kirlin (1993) and constitutes the basic equation for model-error estimation. Note that we never actually store a full matrix $\mathbf{P}_n^p$ of prediction-error covariances. Instead, for instance, for the term $\mathbf{HP}_n^p \mathbf{H}^T$, we obtain an approximation from an N-member ensemble [as in HM, Eq. (15)]:

$$\mathbf{HP}_n^p \mathbf{H}^T = \frac{1}{N-1} \sum_{i=1}^{N} \mathbf{H}(\boldsymbol{\Psi}_{n,i}^p - \overline{\boldsymbol{\Psi}_n^p})[\mathbf{H}(\boldsymbol{\Psi}_{n,i}^p - \overline{\boldsymbol{\Psi}_n^p})]^T, \qquad (8)$$

where i is the index of the ensemble member. The rank of $\mathbf{HP}_n^p \mathbf{H}^T$ is less than or equal to the minimum of $N - 1$, the number of model coordinates, and the number of observations. The matrix $\mathbf{HP}_n^p \mathbf{H}^T$ is positive semidefinite, that is, has no negative eigenvalues (Preisendorfer 1988, p. 28).

We will attempt to use (7) to determine the seven parameters in (1) on the basis of the innovations at a

single time t_n. The estimation algorithm is based on the maximum likelihood method.

b. Parameter estimation using the maximum-likelihood method

Following Dee (1995), we assume that the innovation vector, $\boldsymbol{\nu}$ of length n_{innov}, is normally distributed, with zero mean, and covariance matrix $\mathbf{S}(\boldsymbol{\alpha}_*)$. Here $\boldsymbol{\alpha}_*$ is the vector of the seven model-error parameters, that is, $\langle \epsilon_{20}^2 \rangle$, $\langle \epsilon_{50}^2 \rangle$, $\langle \epsilon_{80}^2 \rangle$, $\langle \epsilon_{20}\epsilon_{50} \rangle$, $\langle \epsilon_{50}\epsilon_{80} \rangle$, $\langle \epsilon_{20}\epsilon_{80} \rangle$, and L. Then, dropping the time index n, (7) can be written as

$$\langle \boldsymbol{\nu}\boldsymbol{\nu}^{\text{T}} \rangle = \mathbf{S}(\boldsymbol{\alpha}_*) \equiv \mathbf{H}\mathbf{P}^p\mathbf{H}^{\text{T}} + \mathbf{H}\mathbf{Q}(\boldsymbol{\alpha}_*)\mathbf{H}^{\text{T}} + \mathbf{R}, \quad (9)$$

where the elements of $\mathbf{Q}(\boldsymbol{\alpha}_*)$ are given by (1).

Given the innovation vector $\boldsymbol{\nu}$, maximum likelihood estimation of $\boldsymbol{\alpha}_*$ reduces to finding that value of $\boldsymbol{\alpha}$ that minimizes the log-likelihood function [cf. Dee 1995, Eq. (29)], that is,

$$f(\boldsymbol{\alpha}) = \ln \det\mathbf{S}(\boldsymbol{\alpha}) + \boldsymbol{\nu}^{\text{T}}\mathbf{S}^{-1}(\boldsymbol{\alpha})\boldsymbol{\nu}. \quad (10)$$

With regard to the first term, we note that the determinant of a covariance matrix is sometimes referred to as the generalized variance (Anderson 1984, p. 259). The second term has a χ^2 distribution with n_{innov} degrees of freedom (Lupton 1993, chapter 4; Stuart and Ord 1987, section 15.10), and therefore has expectation value n_{innov} and variance $2n_{\text{innov}}$.

For a given innovation vector $\boldsymbol{\nu}$, the minimizing value of $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}$, is found using a downhill simplex method (Press et al. 1992, section 10.4). This algorithm does not require information about derivatives, which allows for a straightforward imposition of constraints: parameter values that do not obey a desired constraint are simply assigned an almost-zero likelihood.

Thus, the maximum likelihood estimation technique requires that we be able to accurately evaluate the two terms on the right-hand side of (10). Both of these are functions of $\mathbf{S}(\boldsymbol{\alpha})$, which itself is defined, in (9), as the sum of three terms.

c. Finite ensemble size effects

Consider the evaluation of the three terms on the right-hand side of (9). The first term, $\mathbf{H}\mathbf{P}^p\mathbf{H}^{\text{T}}$, can be evaluated from the ensemble, as specified in (8). Since a parameterization for the model error has been specified in section 2b, the next term, $\mathbf{H}\mathbf{Q}(\boldsymbol{\alpha})\mathbf{H}^{\text{T}}$ can be evaluated for any given $\boldsymbol{\alpha}$. As for the final term, the observational-error covariance matrix, $\mathbf{R}$: this is taken to be known, as in the generation of the perturbed observations and the computation of the gain matrices. We are thus in a position to evaluate all of the terms on the right-hand side of (9). However, for small ensembles, we anticipate rank problems relating to the term $\mathbf{H}\mathbf{P}^p\mathbf{H}^{\text{T}}$. These problems and some ways of dealing with them will now be discussed.

For the purpose of this discussion we assume that the model-error covariances and the observational-error covariances are negligibly small. So we use

$$\mathbf{S}(\boldsymbol{\alpha}) = \mathbf{H}\mathbf{P}^p\mathbf{H}^{\text{T}}, \quad (11)$$

and discuss finite ensemble size effects on the two terms of (10). With regard to the first term, some general results relating to the estimation of the generalized variance are given by Anderson (1984, section 7.5.2). In particular his theorem 7.5.3 indicates that the generalized variance will be systematically underestimated. The underestimation may be very serious, in particular, for small ensembles and large matrices. In fact, in the current hypothetical case where $\mathbf{Q} = \mathbf{R} = \mathbf{0}$, the matrix will not be full rank and its determinant will be zero, if the number of ensemble members is smaller than the dimension of the matrix.

For the χ^2 term of (10), we have a similar effect but in the opposite sense. First of all, we note that the eigenvalues of $\mathbf{S}^{-1}$ are the inverses of the eigenvalues of $\mathbf{S}$. As can be seen from Preisendorfer (1988, Fig. 5.9), a random matrix tends to have some very small eigenvalues. These correspond to very large eigenvalues of the inverse matrix, which will cause an overestimation of the second term of (10). In the current hypothetical case, the inverse is not even defined when the number of ensemble members is smaller than the order of the matrix. As was the case for the first term, the estimation problem is systematic and most serious for large matrices and small ensembles.

To reduce the above systematic estimation problems, one might adopt a combination of the following approaches.

1) Use bigger ensembles. We note that $\mathbf{S}$ is of the order of the number of available observations. In the meteorological context, there may easily be more than 10 000 observations. Due to computational constraints, it may be impossible to have significantly more ensemble members than this. In the present experimental environment, we have about 1000 observations every 12 h. We will use ensemble sizes of 16, 32, 64, and 128. We will also perform some experiments with analytical expressions to simulate having an ensemble of infinite size.

2) Use smaller matrices. It is possible to impose a block-diagonal structure on the matrix $\mathbf{S}$. For instance, one may assume that information from different regions of the globe is uncorrelated. The probability of having a certain parameter value, given all the available innovations, then equals the product of the probabilities for each region. This approach has the advantage of reducing computational costs because only relatively small matrices are now involved. However, by assuming that information from different regions is uncorrelated, one modifies the problem. In this study we will subdivide the sphere into congruent regions using three of the regular convex polyhedra of classical geometry (Coxeter 1969,

chapter 10): the cube (6 regions), the octahedron (8 regions), and the icosahedron (20 regions). We will also look at the limiting cases: a single region consisting of the entire sphere and *each* sounding forming its own region.

3) Perform a Richardson extrapolation. It is possible to exploit the fact that the estimation error is partly of a systematic nature, getting smaller as the ensemble gets bigger. In a Richardson extrapolation (e.g., Dahlquist and Björck 1974, 7–8; Press et al. 1992, p. 134), one performs some numerical algorithm for various values of a parameter h, and then extrapolates the result to the continuum limit $h = 0$. The reciprocal of the ensemble size would seem to be a natural choice for the parameter h. We evaluate the log-likelihood function, $f(h = 1/N)$, for the ensemble of size N. We then divide the ensemble into two halves and evaluate the log-likelihood function for each half to obtain $f_1(h = 2/N)$ and $f_2(h = 2/N)$. The mean of the two values gives the value at $h = 2/N$. We will assume that the estimation error is a linear function of h. Note that an ensemble of infinite size would correspond to $h = 0$. An estimate of this limit value, $f(h = 0)$, is obtained using a linear extrapolation of the $h = 1/N$ value and the two $h = 2/N$ values:

$$f(h = 0) - f\left(h = \frac{1}{N}\right)$$
$$= f\left(h = \frac{1}{N}\right) - \frac{f_1\left(h = \dfrac{2}{N}\right) + f_2\left(h = \dfrac{2}{N}\right)}{2}. \quad (12)$$

Having discussed the general effect of the rank problem on the estimation procedure, we now return to our simulated environment in which there are nonzero model and observational errors. The presence of these terms alleviates, but as we shall see does not eliminate, the rank problem [although it does guarantee the positive definiteness of $\mathbf{S}(\boldsymbol{\alpha})$ even for small ensembles].

d. Systematic errors

To examine the behavior of the log-likelihood function (10), we perform some sensitivity experiments for a test case where all relevant covariance matrices are known exactly. We shall generate realizations of the ensemble mean forecast error, of the observational error as well as ensembles of prediction errors. We may then study our suggestions to minimize the rank problem present in the two terms of (10). These experiments will be performed at time t_0; no cycling with the forecast model and the assimilation system will be done.

To estimate model error, we use the radiosonde observations of $\boldsymbol{\Psi}$ and the satellite thicknesses. Here, the

0000 UTC observational network will be used. This network yields 171 radiosonde observations of $\boldsymbol{\Psi}$ and 612 satellite thicknesses. The forecast- and predictability-error covariances are assumed to be of the same form as the model-error covariance [given by (1)], with the same horizontal correlation [given by (3)]. We set $L = 0.2$ rad.

We specify the following height-error covariance matrix [cf. HM, Eq. (4)]

$$\mathbf{V}^f(t_0) = \begin{pmatrix} 390 & 80 & 20 \\ 80 & 120 & 20 \\ 20 & 20 & 50 \end{pmatrix}. \quad (13)$$

(The units for these covariances are m².) This matrix is partitioned, according to (4), by assuming that predictability error and model error account for 70% and 30%, respectively, of these covariances. This allows us to specify the two vertical covariance matrices, $\mathbf{V}^p(t_0)$ and $\mathbf{V}^q(t_0)$, which implicitly define the covariance matrices $\mathbf{P}^p$ and $\mathbf{Q}$:

$$\mathbf{V}^f(t_0) = \mathbf{V}^p(t_0) + \mathbf{V}^q(t_0), \quad (14)$$
$$\mathbf{V}^p(t_0) = 0.7\mathbf{V}^f(t_0), \quad (15)$$
$$\mathbf{V}^q(t) = 0.3\mathbf{V}^f(t_0). \quad (16)$$

Note that the data assimilation experiments of section 4 will be initialized using $\mathbf{V}^p(t_0)$, $\mathbf{V}^f(t_0)$, and $L = 0.2$ rad. The model-error statistics will be taken to be independent of time.

Using the specified parameter values for L and $\mathbf{V}^f(t_0)$, we use a random-field generator (appendix A) to generate a forecast-error realization. A vector of observational errors is generated using the error covariances specified in section 2a, that is, as implied by Eqs. (2) and (3) of HM. Since the interpolation operator is linear and perfect, these realizations of the forecast and observational error allow us to calculate the innovation vector, $\boldsymbol{\nu}$. Using the specified values for L and $\mathbf{V}^p(t_0)$, we also generate an ensemble of prediction-error fields. Because the observational-error covariance, $\mathbf{R}$, is assumed to be known, $\mathbf{S}(\boldsymbol{\alpha})$ can now be calculated, from (9), as the sum of its three constituent terms for any trial vector $\boldsymbol{\alpha}$. Using $\boldsymbol{\nu}$ and $\mathbf{S}(\boldsymbol{\alpha})$, the two terms of (10) can now be evaluated.

1) CONVERGENCE TESTS

Our first objective is to determine whether there is any (systematic) variation in the estimated values of the two terms on the right-hand side of (10) as the ensemble size changes. To do this, we simply evaluate the two terms of the log-likelihood function at the *correct* vector, $\boldsymbol{\alpha}_*$, of the parameter values. Repeating the experiment 100 times, with different realizations of the forecast and observational errors and of the ensemble of prediction-error fields, allows us to obtain statistically significant conclusions.
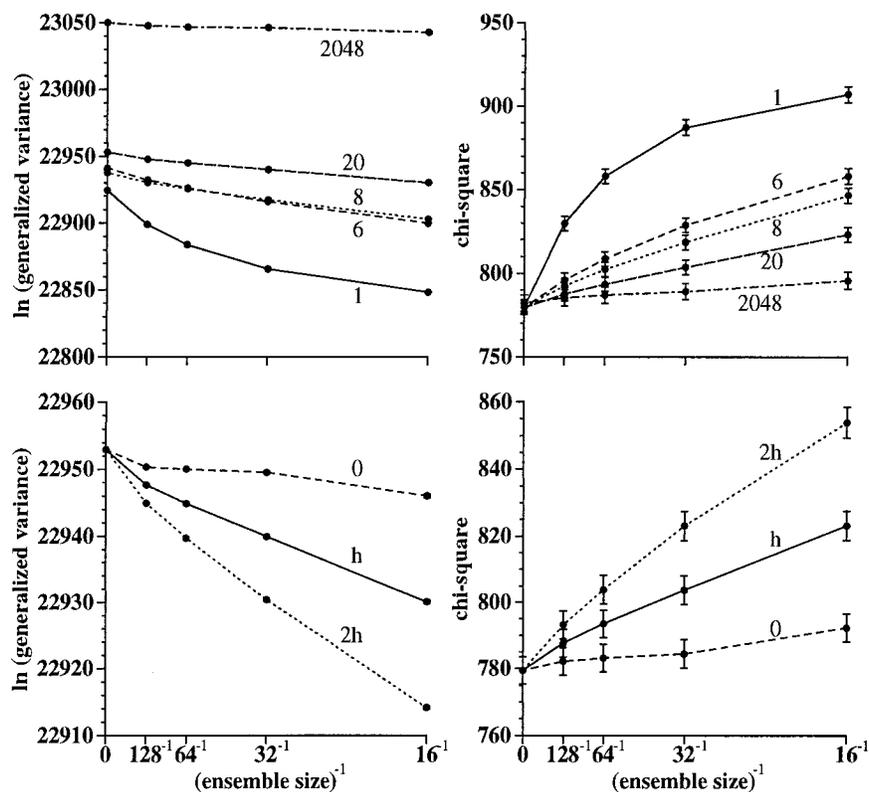
FIG. 4. The upper panels show the dependence of the two terms of the log-likelihood function on the reciprocal of the ensemble size and on the number of regions. The left-hand panel is for the term $\ln \det \mathbf{S}(\boldsymbol{\alpha}_*)$, while the right-hand panel is for the term $\boldsymbol{\nu}^{\mathrm{T}} \mathbf{S}^{-1}(\boldsymbol{\alpha}_*) \boldsymbol{\nu}$. For each ensemble size, the mean value (and its uncertainty) of 100 evaluations of these two terms are shown. In each panel, the values along the ordinate were obtained using analytical expressions. Each curve is labeled to indicate the number of regions. The curve labeled 1 is for a single region consisting of the entire sphere and the curve labeled 2048 is for the case where each point of the horizontal grid is located in its own region. The lower panels are similar, but show the effect of the Richardson extrapolation for the case with 20 regions. In these panels, the solid lines (labeled h) are for the mean values of 100 evaluations with ensembles of size N; the dotted lines (labeled $2h$) are for ensembles of size $N/2$; and the dashed lines (labeled 0) are for the extrapolated values.

In the upper left- and right-hand panels of Fig. 4, we show the values of these two terms as a function of the reciprocal of ensemble size for different subdivisions of the sphere. In each panel, the discs show the mean value of the corresponding 100 realizations. Also plotted in each panel (but not visible in the left-hand panel) are error bars, which indicate the standard deviation of the mean values. In the present test, since the prediction-error term in (9) is prescribed, it can be obtained from its analytical specification without using ensembles. The results obtained using such an approach correspond to the results that would be obtained with an infinite ensemble and are plotted accordingly along the ordinate in each of the two upper panels. It can be seen that these values are indeed limiting values for the ensemble results.

Consideration of the upper left-hand panel of Fig. 4 shows that the first term is severely underestimated for small ensembles, as could be expected from the discussion in the previous subsection. In fact, in the case

of a single (global) region and $N = 16$, the underestimate as compared to the limiting value is about 76 units. This is a very large number, equal to almost twice the standard deviation of the second term of (10), which is $\sqrt{2n_{\mathrm{innov}}} = \sqrt{2 \times 783} = 39.6$ in the present case. Even with $N = 128$, there is still a systematic error of 25 units. Subdividing the sphere into regions, which are assumed to be independent, increases the limiting values but reduces the sensitivity to the ensemble size. Using 20 regions reduces the difference between the $N = 16$ and limiting value to 23 units. In the extreme case, where each column of grid points forms its own region, the corresponding difference is only seven units. The effect of subdividing the sphere into regions is to impose a block structure on $\mathbf{S}$. This reduces the rank problem of $\mathbf{S}$, while increasing its determinant.

The corresponding results for the second term of (10) are shown in the upper right-hand panel of Fig. 4. The sensitivity to the ensemble size is now of the opposite sign, as expected. Since the second term depends on the

realizations of the innovation vector, significant error bars are now visible even after 100 experiments. In this panel, no significant change of the limiting values is observed. In fact, within the error bars, the five limiting values are all equal to 783 (the value of n_{innov} in this case). Finally we note that, with six or more regions, the systematic error in both panels seems to depend linearly on the reciprocal of the ensemble size. This strongly suggests that a Richardson extrapolation to infinite ensemble size may be useful.

The lower panels of Fig. 4 show the effect of a Richardson extrapolation for the case where the sphere is subdivided into 20 regions. In each panel the solid line is copied from the corresponding upper panel. The dotted line shows the mean of two evaluations with half the ensemble size, averaged as before over 100 different realizations. The dashed line shows the extrapolated value. The difference between the extrapolated value and the value for the complete ensemble is, from (12), identical to the difference between the value for the complete ensemble and the value for the mean of the two small ensembles. A comparison of the complete-ensemble values for ensemble sizes of 16, 32, and 64 members (joined by the solid curve) with the corresponding half-ensemble mean values (located on the dotted curve) shows that the value for the complete ensemble of size N is almost identical to the mean value for two ensembles of size N, indicating that a sufficient number of realizations has been used. It can be seen in the lower left-hand panel that the Richardson extrapolation reduces the systematic error at $N = 16$ from 23 to 7 units. For the χ^2 term of (10), it can be seen in the lower right-hand panel, that there is a corresponding reduction of the systematic error from 44 to 13 units. In fact for ensembles of size 32 or larger, the random error with 100 realizations is of the same order as the systematic error.

We conclude from this first experiment that the use of larger ensembles, of block-diagonal matrices, and of a Richardson extrapolation leads to a significant reduction of the rank problem with the log-likelihood function. However, subdividing the sphere, as has been done here in order to produce block-diagonal matrices, modifies the original problem and in particular increases the absolute value of the generalized variance. Since minimizing the log-likelihood function involves the evaluation of this function for different parameter values and differencing the results, this *may* not be a serious problem. To further examine this issue, we will perform some further experiments to evaluate how useful the proposed measures actually are for estimating model-error parameters from the available data.

2) ESTIMATION OF THE VERTICAL COVARIANCE MATRIX

Here, we evaluate the log-likelihood function for different trial values of the parameters, starting with the true vector, $\boldsymbol{\alpha}_*$. After averaging over many realizations, this vector should minimize the log-likelihood function. We also evaluate the log-likelihood function with vertical covariances that have been reduced by a factor of 1.5. We now expect higher values of the log-likelihood function on average, indicating that the modified parameter values lead to an inferior match with the innovations. We have the same expectation for the case where the vertical covariances are increased (again by a factor of 1.5) with respect to their correct values.

In Fig. 5, we show the values of the log-likelihood function evaluated with perturbed parameters minus the values obtained with $\boldsymbol{\alpha}_*$. Four-hundred realizations are used for each plotted value. The upper panels are for the case with 20 regions and the lower panels for the case where each sounding is considered to be in its own region. The left-hand panels are for the comparison between correct and reduced covariances and the right-hand panels are for the comparison between correct and amplified covariances. In each panel a positive value indicates that $\boldsymbol{\alpha}_*$ is observed to be more likely than the perturbed vector. Looking first at the upper left-hand panel, we observe positive values in all cases. Oddly enough, the results seem to be most significant when small ensembles are used without extrapolation. An explanation can be obtained by comparing with the upper right-hand panel. Here the results are worst for the cases with the smallest ensembles and without extrapolation. The negative values seem to indicate that variances larger than the true ones provide a more likely match to the available data. In other words, using small ensembles, no Richardson extrapolation and 20 regions, the vertical covariances of the model error are likely to be severely overestimated. As can be seen from this panel, the use of larger ensembles and a Richardson extrapolation act to correct the overestimation. The lower panels of Fig. 5 show that the Richardson extrapolation used with the maximal subdivision of the sphere eliminates the problem for all ensemble sizes studied here. The significantly positive values (but compare the scale with that of Fig. 4) suggest that 400 realizations are sufficient to enable us to correctly distinguish between true and perturbed values, and thus to determine the vertical covariances to within 50%. The results of Fig. 5 suggest that we use this maximal subdivision of the sphere together with a Richardson extrapolation for estimating the vertical covariances of the model error in our data assimilation experiments in section 4.

3) ESTIMATION OF THE HORIZONTAL LENGTH SCALE

Using the same 400 realizations as for the previous experiment, we now turn to the estimation of the horizontal length scale. For these experiments the vertical covariances are given by the true values of the previous section, that is, as specified in (16). The true length scale, L_*, is taken to be 0.2 rad. A ''positive'' change
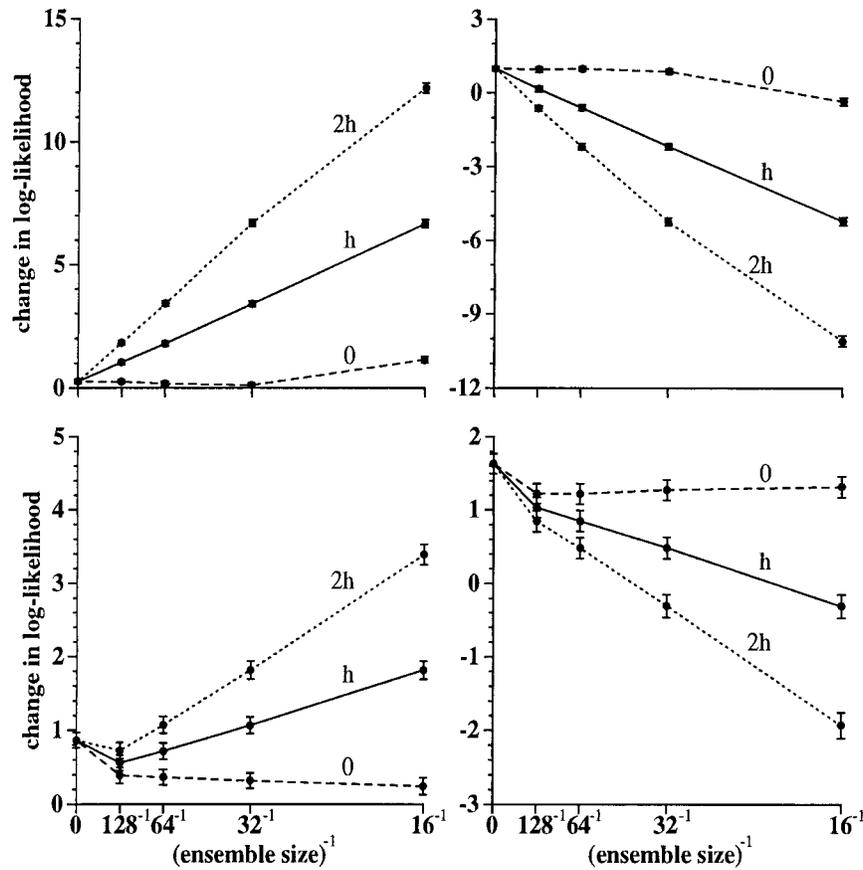
FIG. 5. Sensitivity of the log-likelihood function to changes in the vertical covariances for different ensemble sizes and numbers of regions, both with and without the Richardson extrapolation. All panels show the log-likelihood function evaluated with perturbed values of the vertical covariances minus the log-likelihood function evaluated with the true vertical covariance values. In the left-hand panels, the perturbed values are set to two-thirds of the true vertical covariance values. In the right-hand panels, the perturbed values are set to 1.5 times the true vertical covariance values. The upper panels are for 20 regions and the lower panels for the case where each sounding is located in its own region. The solid lines (labeled *h*) are for ensembles of size *N*, the dotted lines (labeled 2*h*) are for ensembles of size *N*/2 and the dashed lines (labeled 0) are for the extrapolated results.

to the true length scale consists of increasing its value by 50%, that is, to 0.3 rad. For a negative perturbation, we divide the true value by 1.5. As the number of regions increases and their size decreases, the sensitivity of the log-likelihood function to the length scale will decrease. In fact, the sensitivity is zero in the extreme case where all grid columns are considered to be in different regions.

The upper and lower panels of Fig. 6 show the sensitivity to ensemble size when the sphere is subdivided into 8 and 20 regions, respectively. Similar to the layout used for Fig. 5, the left-hand panels are for the comparison between true and reduced length scales and the right-hand panels are for the comparison between true and increased length scales. In all panels, positive values indicate that L_* is observed to be more likely than the perturbed length scale. We first compare the solid curves in the upper panels with those in the lower panels. These

curves indicate that slightly more significant results (i.e., slightly higher values) are obtained with eight regions than with 20 for the largest ensemble sizes. However, due to the rank problem, the results with eight regions change more rapidly with decreasing ensemble size. (Note the difference in scale between the upper and lower panels). The solid (and dotted) curves indicate a dependence of the systematic error on the reciprocal of the ensemble size that is more quadratic than linear. One result is that in the case of the left-hand panels the effect of the Richardson extrapolation can at best be categorized as neutral. On the other hand in the right-hand panels, where the curvature of the systematic error curve is somewhat different, the Richardson extrapolation clearly leads to much flatter curves. With respect to the sensitivity of the log-likelihood function to the length scale, we would expect better and better results as the ensemble size increases, allowing the use of larger and
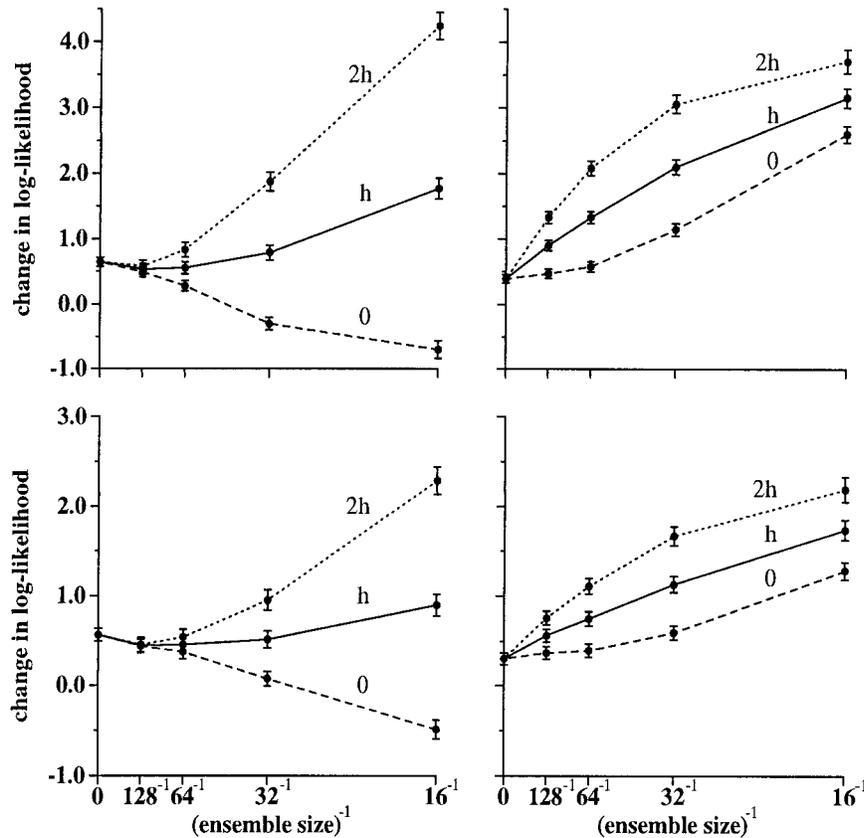
FIG. 6. As in Fig. 5 but for the horizontal length-scale parameter. The upper panels are for 8 regions and the lower panels are for 20 regions.

larger regions. From Fig. 6, we may conclude that the configuration with eight regions is to be preferred over the configuration with 20 regions for ensemble sizes of 128 or larger.

4) SUMMARY

We have found that for ensembles of any realistic size (say less than 1000 members), the log-likelihood function changes systematically as a function of ensemble size (see solid curves in upper panels of Fig. 4). This undesirable behavior will cause a strong bias in any estimate of model-error parameters. To deal with this problem, we have proposed subdividing the sphere into a number of congruent regions and using a Richardson extrapolation to infinite ensemble size. With these two measures, the bias in the estimates is, in general, very significantly reduced.

For the remaining experiments (including the data assimilation experiments in section 4), we will fix the ensemble size at 32 members, that is, we will use a configuration with two ensembles of 32 members each, configured as proposed in HM. Since we will be estimating all model-error parameters simultaneously, we will employ the 20-region subdivision of the sphere (to-gether with a Richardson extrapolation to infinite ensemble size) for the model-error estimation algorithm.

e. Performance of the algorithm

In the previous sections we have seen how one might deal with the systematic errors that occur when relatively small ensembles are used. We would now like to obtain quantitative information about the performance of the proposed configuration. In particular, we wish to see if all parameters are separately identifiable and what accuracy might be anticipated.

For this purpose, we use the same 400 realizations as above, but now allow the minimization procedure to converge to the most likely value. From the resulting 400 realizations $\hat{\alpha}$ of α, we can obtain the median, α_{median}, the mean, α_{mean}, and the rms error of $\hat{\alpha}$. These are presented in Table 1. It can be seen that, in agreement with our expectations from appendix B, the median tends to underestimate the variances, while the mean leads to overestimates. The mean length-scale seriously overestimates the true value due to the presence of some very large estimates $\hat{L}$. For all parameters, the uncertainty of single estimates is of the same order as the

TABLE 1. Accuracy of the parameter estimates as calculated from 400 realizations $\hat{\alpha}$ of α. For each of the seven parameters, the true value, α_*, the median, α_{median}, the mean, α_{mean}, as well as the rms error in the individual estimates of α are given.

Parameter	α_*	α_{median}	α_{mean}	$\sigma(\hat{\alpha})$
$\langle \epsilon_{20}^2 \rangle$	117	102.5	131.5	106.3
$\langle \epsilon_{50}^2 \rangle$	36	30.5	38.0	30.1
$\langle \epsilon_{80}^2 \rangle$	15	10.7	15.2	15.3
$\langle \epsilon_{20}\epsilon_{50} \rangle$	24	27.9	26.6	32.6
$\langle \epsilon_{50}\epsilon_{80} \rangle$	6	6.6	5.8	11.3
$\langle \epsilon_{20}\epsilon_{80} \rangle$	6	6.8	4.9	12.5
L	0.2	0.19	0.26	0.27

estimates themselves. Nevertheless, the results indicate that, after smoothing, useful estimates can be obtained.

From the 400 realizations, we also computed the covariance matrix $\overline{(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_{mean})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_{mean})^T}$ and the corresponding correlation matrix. This was recommended to us by P. Gauthier (1998, personal communication) and can be used to investigate whether the chosen parameterization contains dependent parameters that cannot be separately identified. We found nonzero correlations between most of the vertical covariances. However, the highest correlation (0.30) was still fairly small. The correlations between the length-scale parameter and the variances were insignificant. An eigenvalue analysis of the correlation matrix yielded eigenvalues between 0.43 and 1.64. This suggests that all parameter values can be identified from the observations. Note that a zero eigenvalue would correspond to dependent (redundant) parameters, while completely independent parameters would yield unit eigenvalues. Such independent parameters might be estimated independently, if it were convenient; perhaps using different subsets of observations.

To see if the above results were very dependent on the distribution of $\boldsymbol{\alpha}$, we also computed Spearman rank-order correlation coefficients (Lupton 1993, chapter 13; Press et al. 1992, section 14.6). These correlations were very similar to the original ones, which suggests that the above conclusions are robust. This encourages us to believe that the parameter estimation algorithm will perform properly in a data assimilation cycle.

4. Data assimilation cycle results

The adaptive ensemble Kalman filter will be evaluated by examining its performance in a 30-day data assimilation cycle. As in HM, the initial time (t_0) is denoted 0000 UTC of day 1 and the final time is 0000 UTC of day 31. As in our previous study, we use the domain-averaged streamfunction error squared north of 20°N at 50 kPa, as a measure of the error variance. We use a pair of 32-member ensembles, configured as proposed in HM, and set r_{max}, the cutoff radius for the analysis algorithm, equal to 20°. The latter choice is indicated by Fig. 5 of HM.

In the experiments, we take the true model error to be of the form (1) with $\mathbf{V}^q$ given by (16) and ρ_L defined

by (3) with $L = 0.2$ rad. (Note that the parameters defining the true model error are taken to be time invariant.) The ensemble mean will be perturbed every 12 h with model-error realizations having these statistics, as illustrated in the upper portion of Fig. 2. The initial ensemble of 12-h predictions is generated as described in HM [Eqs. (7) and (8)] with random perturbation fields also taken to be of the form (1) with $\mathbf{V}^p$ given by (15) and ρ_L defined by (3) with $L = 0.2$ rad.

The purpose of the first two data assimilation experiments is to establish standards of comparison for the adaptive ensemble Kalman filter. In the first experiment, no attempt is made to estimate the model error or increase the spread between the ensemble members to account for it. Thus the data assimilation algorithm proceeds subject to the assumption that the model error can be neglected. The results of this experiment are shown for both ensembles of the pair in the two upper panels of Fig. 7. As in HM, two measures of error are shown in each panel: the rms difference between the ensemble mean and the true state and the rms spread in the ensemble. The results in the two panels are seen to be rather similar and indicate that the spread in the ensemble severely underestimates the error in the ensemble mean. These results indicate the type of error that can be incurred by assuming that the model is perfect when, in fact, model error is significant.

The second data assimilation experiment is similar to the first, but at the opposite extreme. Again no attempt is made to estimate the model error, but in this case, rather than completely neglect the model error as before, representative ensembles of realizations of model error are produced using the *true* model-error parameters. (More specifically, the true parameters are used as the best ''estimate'' of the parameters in the lower part of Fig. 2.) The results of this experiment, shown for both ensembles of the pair in the two lower panels of Fig. 7, indicate the dramatic improvement that this produces. It can be seen that the ensemble spread is now a very good estimate of the error in the ensemble mean and that there has been a substantial decrease in the ensemble-mean error as compared with the upper panels.

The results in the lower panels of Fig. 7 indicate that our procedure for increasing the spread in the ensemble (lower part of Fig. 2) works well. As in HM (Fig. 3), given appropriate statistical information about the observational- and model-error covariances, the configuration with a pair of ensemble Kalman filters is able to maintain ensembles having a spread that accurately reflects the error variance of the ensemble mean. If there were other sources of model error in addition to the imposed model error of section 2b, these would also have to be accounted for so that representative ensembles could be maintained. For example, as shown in HM, the use of small ensembles in a configuration with a single Kalman filter can lead to a deficient spread due to an inbreeding problem. Using (9) with such a configuration could result in larger model-error estimates
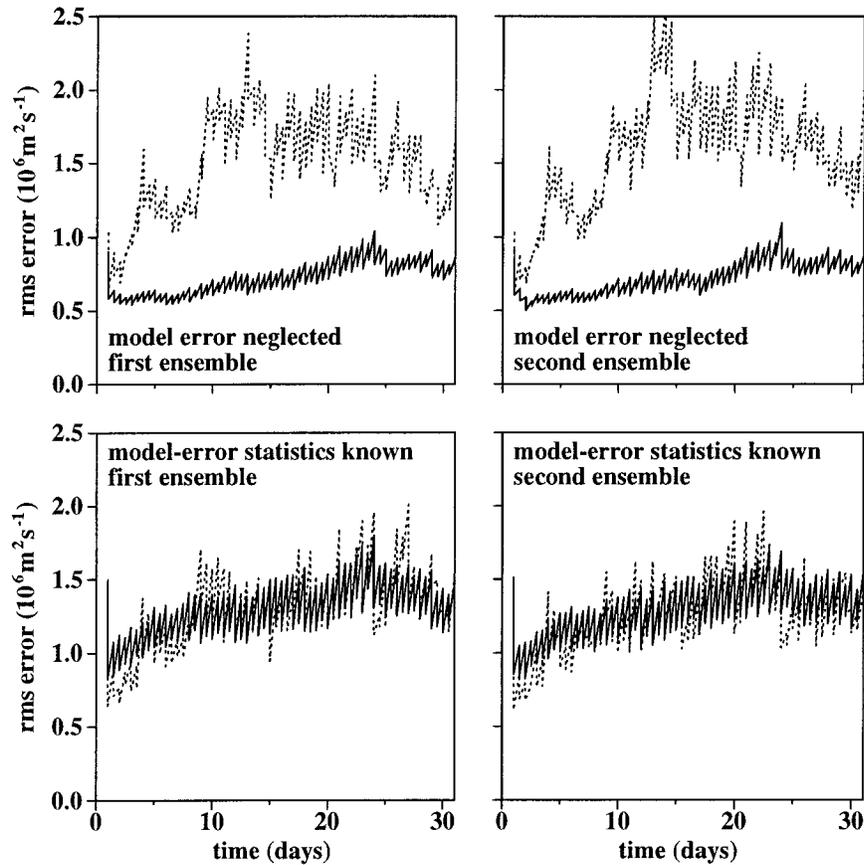
Fig. 7. Forecast and analysis error at 50 kPa every 12 h during two 30-day assimilation experiments. In each panel, the solid line shows the rms spread in the ensemble, while the dashed line indicates the rms error of the ensemble mean. The upper panels show the results, for both ensembles of a pair, of an experiment where no attempt is made to estimate the model-error parameters or to account for the model error. In the case of the lower panels, the model error is accounted for using the true model-error parameters.

because the model error would then also include a component due to the data assimilation system. This error would depend on the observational network and therefore be nonhomogeneous and anisotropic. It was not investigated in the context of this study.

Given that here the only model-error component is that described in section 2b, the performance of our adaptive filter will depend crucially on the ability of the parameter-estimation procedure to provide accurate estimates of the parameters of the imposed model error. We now proceed to examine the performance of the parameter-estimation procedure and the adaptive ensemble Kalman filter.

Since single-sample estimates tend to be inaccurate, it is desirable to smooth them in time (e.g., Dee 1995; Blanchet et al. 1997). Therefore, as the best estimate of the model-error parameters in Fig. 2, we will use a smoothed estimate. Since the horizontal correlation function given by (3) could just as well have been expressed in terms of a parameter $c = L^{-1}$ and since, in general, taking mean values of L would lead to different

results than taking mean values of c, and to protect the smoothed estimate against the effect of some very large single estimates of L (which would have a large impact on the mean), we decided to smooth using the median, which is more robust. In fact, we smooth L and each of the six parameters that define the vertical covariance matrix using the median, as illustrated in the bottom portion of Fig. 3. (As indicated in appendix B, the median will tend to underestimate the variances.) To ensure that the resulting vertical covariance matrix is positive definite, we perform an eigenanalysis after the medians are calculated and adjust any problematic eigenvalues to a small positive value.

We now repeat the 30-day data assimilation cycle, estimating the model-error parameters every 12 h and attempting to use the temporally smoothed estimates to account for the model error. In Fig. 8, we show the individual estimates and the evolution of the median for the 50-kPa rms model error and for $\sqrt{L}$ for each ensemble of the pair. (Note that these quantities, rather than the 50-kPa model-error variance and L itself, are
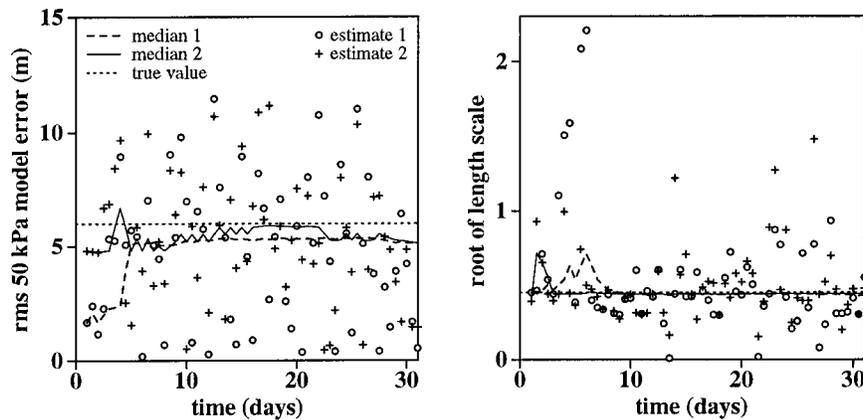
FIG. 8. The left-hand panel shows individual estimates and the evolving median estimate of 50-kPa rms model error every 12 h for each ensemble of the pair during a 30-day assimilation cycle. The right-hand panel is similar but for the square root of the horizontal length scale, L. In each panel, the dotted line indicates the corresponding true value.

shown to minimize the apparent dispersion.) The true values are given by the dotted lines in each panel. It can be seen that, while many of the individual estimates are quite inaccurate, within 10 days the temporally smoothed estimates have essentially converged to their asymptotic values. While the latter value is virtually indistinguishable from the true value in the case of the length scale, the estimation procedure seems to yield an underestimate of the true value in the case of the 50-kPa variance. Partly, this could be expected due to the use of the median for the temporal smoothing, and furthermore, the very small positive values for 32 ensemble members in the upper-left panel of Fig. 5 indicate that with 20 regions the estimation procedure has great difficulty in distinguishing between the correct parameter values and parameter values that are too small.

The performance of the adaptive ensemble Kalman filter in terms of forecast and analysis error over the 30-day assimilation period is shown in Fig. 9. These results can be directly compared to the results of Fig. 7, both

in terms of the magnitude of the ensemble mean error and in terms of the degree of agreement between the ensemble spread and the error in the ensemble mean. A comparison shows that, in terms of both criteria, the performance of the adaptive ensemble Kalman filter is almost as good as that of the cycle where the correct model-error parameters were used to account for model error. Thus within the context of the present experimental setup, the adaptive ensemble Kalman filter performs very well.

5. Summary and concluding discussion

The objective of this study was to extend the ensemble Kalman filter technique developed earlier by Evensen (1994) and HM. With a view to its eventual operational implementation, the most important limitation of the technique proposed in the latter study was the assumption that a perfect forecast model was available. Our goal here has been to drop this assumption and to
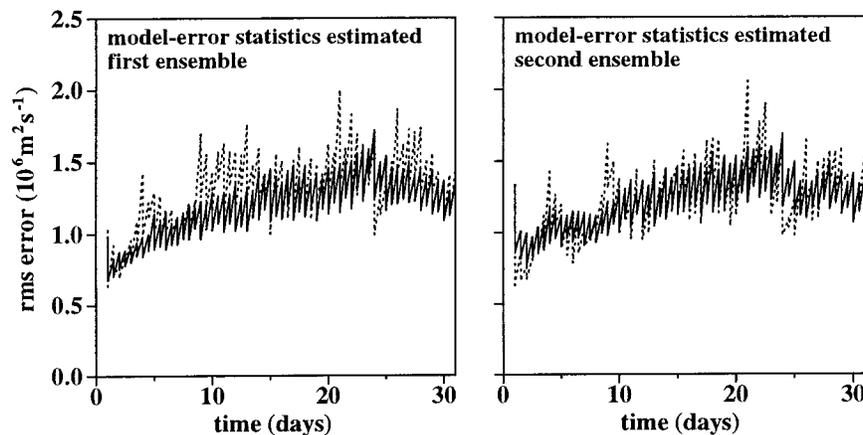

FIG. 9. As in Fig. 7 but the estimated model-error parameters are used in an attempt to account for the model error.

develop an adaptive algorithm capable of estimating some of the statistics of the model error and using these estimates to account for the model error. Like HM, this study has been performed in the context of the three-level, quasigeostrophic, T21 model of Marshall and Molteni (1993).

To develop an adaptive algorithm, we have basically followed the approach proposed by Dee (1995), that is, the model error has been parameterized in terms of a small number of parameters and we have used the innovations at each analysis time to estimate the parameters. The model mean errors have been taken to be zero in this study, as is often done [but see Dee and da Silva (1998), where a sequential bias estimation and correction algorithm is developed]. In fact, we have expressed the model error as a product of a vertical covariance matrix and an isotropic, horizontal correlation function.

As in Dee (1995), the parameters were estimated using the maximum likelihood method. In practice, this method reduces to the minimization of the log-likelihood function given by (10). To estimate the model-error parameters at any given analysis time, we have used all of the streamfunction observations from the radiosonde network and all the available satellite thickness observations. As in Dee (1995), the observational-error statistics were taken to be known. In an operational environment, they would have to be estimated.

The use of a limited number of ensemble members gives rise to a rank problem in the estimation of **S**, the covariance matrix of the innovations. The effect of this rank problem on the terms of the log-likelihood function is that one term (the natural logarithm of the generalized variance) is underestimated, while the other term (the χ^2 term) is overestimated.

In view of our desire to use relatively small ensembles, a number of strategies have been developed to deal with these systematic estimation problems. Subdividing the globe into a number of supposedly independent regions has the effect of imposing a block-diagonal structure on the matrix **S**, which greatly improves its conditioning. Another strategy consists of dividing a given ensemble into two and evaluating the log-likelihood function for both the original ensemble and its two halves. These values can then be combined to perform a Richardson extrapolation to infinite ensemble size.

For the data assimilation cycles, it was decided to use a pair of ensembles having 32 members each, configured as proposed in HM. It seemed prudent (see Table 1) to temporally smooth the model-error parameter estimates and the median estimate was chosen for this purpose, as discussed in appendix B. (In an operational context, one might use the median over a moving time window.) It was found that the adaptive ensemble Kalman filter produced fairly useful smoothed estimates of the parameters and accounted rather well for the model error. In fact, its performance was much better than that of a cycle in which the effect of the model error was neglected and was almost as good as that of a cycle in which the *correct* model-error parameters were available.

In an operational environment, the number of available observations might be larger than was the case here by at least an order of magnitude, but one might still be constrained to limit the number of ensemble members to about 100. The result would be an increase in the severity of the rank problem, which could compromise our ability to estimate moderately large length scales. A modification to our strategy for partitioning the sphere might allow this problem to be overcome. The idea is to prespecify the desired number of "regions" (*M,* say) and then randomly assign to each sounding a number from 1 to *M*. All soundings that had been assigned the same number would then constitute a single "region." The resulting regions would impose a block-diagonal structure on the covariance matrix of the innovations, which would still allow for the estimation of correlations of global extent. Another possible way of dealing with the rank problem would be by accumulating the innovations for a fixed observational network over a period spanning several data assimilation times before evaluating the log-likelihood function.

Unlike the present experimental environment, the model error is no longer of a known form in an operational context. [This situation has been simulated by Dee (1995) with encouraging results.] For an operational implementation, one might want to develop a more "appropriate" parameterized form for the model error. One might, for example, want to account for possible latitudinal structure of the model-error standard deviation and length scale or possible serial correlation of model error (Daley 1992a; Zupanski 1997). Serial correlation will occur if the model error is correlated with the atmospheric state (Mitchell and Daley 1997a, b). In the context of a primitive equation model, the model-error perturbations will have to be balanced in some sense, since adding unbalanced fields to 6-h forecasts is detrimental (Daley 1991, section 6.3).

If model error is significant, it is likely easier to estimate and one may be able to use more sophisticated formulations, such as the global descriptions that are utilized for forecast error in current 3D analysis algorithms (e.g., Rabier et al. 1998). One would then take $\mathbf{Q}(\alpha)$ to be of the same functional form as the covariance matrix $\mathbf{P}^f(\boldsymbol{\alpha}_{3D})$, which is used for the forecast error by the 3D assimilation procedure. Note that such a formulation is likely to produce balanced model-error perturbations (Derber et al. 1991). If the model error is not so large, it will be more difficult to estimate all the parameters of such a complex description. However, as the model error is smaller, a simpler formulation, based on fewer parameters, will likely be adequate. In the extreme case, it might be assumed that the model-error covariances are simply proportional to the forecast-error covariances used by the 3D system, that is, $\mathbf{Q}(\alpha) = \alpha\mathbf{P}^f(\boldsymbol{\alpha}_{3D})$, for $0 \leq \alpha \leq 1$. {Note that this is similar to a prescription by Dee and da Silva [1998, Eq. (122)]

for the bias prediction-error covariance.} With this formulation, there is only a single parameter, α, to be estimated from the innovations. We note that this estimation could be performed assuming that all elements of the innovation vector are independent. This would make **S** diagonal and eliminate the rank problem.

Model error would be reduced if error sources in the forecast model could be identified and eliminated. As part of this effort, different versions of the model could be used for different ensemble members (Houtekamer et al. 1996a; Houtekamer and Lefaivre 1997). In this way, some of the flow-dependent intermittent sources of model error could be simulated and the unexplained model error could be reduced. We could then redefine **P**^{p} to also include the *explained* part of the model error and redefine **Q** in (1) to just consist of the remaining *unexplained* part of the model error. It would seem that we could then still use (7) to estimate **Q**.

Further experiments in a more realistic environment will be required to investigate to what extent the most important aspects of model-error description and estimation have now been dealt with. However, the encouraging results of the present study strengthen our belief that an (adaptive) ensemble Kalman filter using the Canadian Meteorological Centre (CMC) global forecast model is feasible on the current CMC computers.

APPENDIX A

Generation of 3D Fields Having a Prescribed Covariance Structure

An (approximate) method for generating realizations of random fields, having on average a prescribed covariance structure and zero mean, was given in HM. An important component of that method was an algorithm, operating in grid space, for generating horizontal (2D) random fields. In the current study relatively long correlation lengths may occur, for which that algorithm is not efficient. However, due to the fact that only homogeneous and isotropic correlation functions are prescribed, it is possible to use an efficient algorithm in spectral space.

Following Weber and Talkner [1993, Eq. (9)] we first obtain a spectral expansion f_n, truncated at wavenumber 21, of the second-order autoregressive function (3) using

$$\rho(r) \approx \sum_{n=0}^{21} f_n P_n(\cos r), \tag{A1}$$

$$f_n = \frac{2n + 1}{2} \int_0^\pi P_n(\cos r)\rho(r) \sin r \, dr, \tag{A2}$$

where P_n are the Legendre polynomials. The f_n are obtained numerically from (A2). Any negative coefficients are set to zero. (Such coefficients sometimes occurred at large wavenumbers when length scales were large.)

For $\mu = \sin(\text{latitude})$ and $\lambda = \text{longitude}$, a random field $X(\mu, \lambda)$, truncated at T21, on the sphere can now be obtained from [see Yaglom 1987, Eq. (4.191)–(4.195)]:

$$X(\mu, \lambda) = \sum_{n=0}^{21} \sum_{m=-n}^{n} P_{m,n}(\mu)e^{im\lambda}Z_{nm}, \tag{A3}$$

$$\langle Z_{nm}Z_{lk}^* \rangle = \frac{4\pi}{2n + 1} \delta_{nl}\delta_{mk}f_n. \tag{A4}$$

Here $P_{m,n}$ are associated Legendre polynomials, Z_{nm} are random (Gaussian) variables with zero mean and variance given by (A4), * denotes complex conjugation, and the angle brackets, $\langle \cdot \rangle$, denote the expectation operator.

APPENDIX B

Smoothing of Maximum Likelihood Estimates

Two possible strategies for smoothing a set of single-sample estimates are taking the mean or the median of the set of estimates. We use a simple example, modeled on (9), to show how these two strategies may lead to rather different results, especially when the number of available observations is small.

Suppose we have n_{innov} observations, where n_{innov} will vary from 1 to 20. We take the following error covariance matrices:

$$\mathbf{HP}^p\mathbf{H}^T = \mathbf{I}, \quad \mathbf{R} = \mathbf{I}, \quad \mathbf{HQ}(\alpha)\mathbf{H}^T = \alpha\mathbf{I}, \quad \alpha_* = 0.3, \tag{B1}$$

where **I** is the identity matrix and the other matrices and symbols may be interpreted as in sections 2 and 3.

Using a random number generator, we obtain 10 000 different single-sample estimates of α_*. Each estimate is obtained using the following procedure.

1) An innovation vector is generated as a random draw from a normal distribution with mean zero and covariance 2.3**I** [cf. Eq. (9)].
2) A single estimate $\hat{\alpha}$ is obtained by substituting the identity matrix for $\mathbf{HP}^p\mathbf{H}^T$ and **R** and using the maximum likelihood method. Negative estimates are set equal to zero.

From the 10 000 values of $\hat{\alpha}$, we compute both the mean, α_{mean}, and the median, α_{median}. We observe in Fig. B1a that $\alpha_{\text{median}} < \alpha_*$ and $\alpha_{\text{mean}} > \alpha_*$. In fact, the distribution
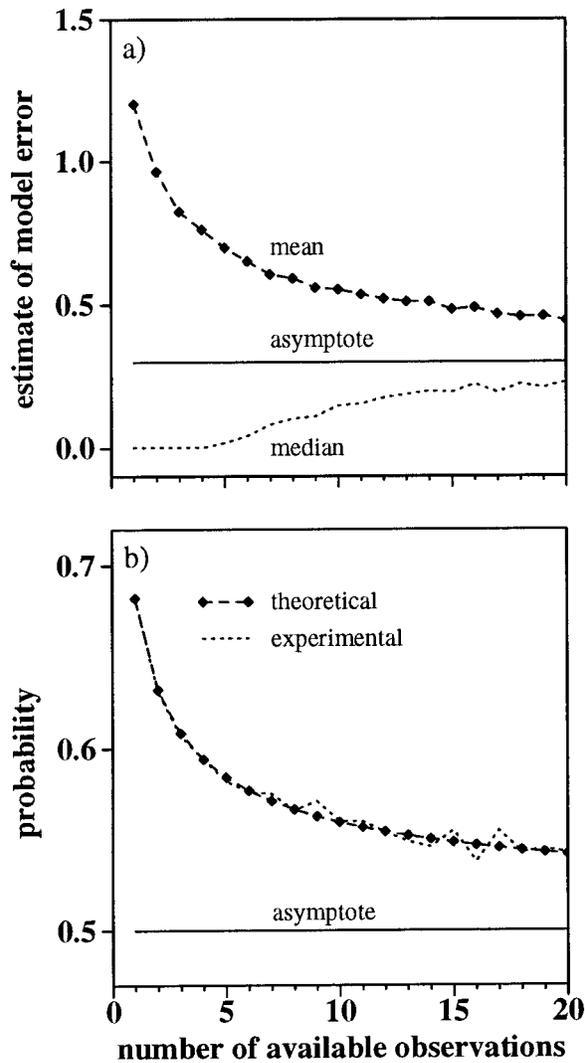
side can also be estimated experimentally from the above 10 000 realizations, since α_* is known. From Fig. B1b, we see that the theoretical and experimental probabilities agree with each other. As n_{innov} increases, the probabilities asymptotically approach 0.5 from above, which explains why the median asymptotically approaches α_* from below in Fig. B1a.

In the current example, a better estimate of α_* could be obtained by using a more appropriate percentile, that is, $P(\chi^2_{n_{innov}} < n_{innov}) \times 100\%$ rather than 50% as for the median. For example, for $n_{innov} = 1$, one might estimate α_* as the value of α below which 68.3% of all individual estimates occur. In practice, the dependence of **S** and **Q** on α will likely be much more complicated than was the case here. Also, due to spatial correlations, it may not be so obvious how many independent pieces of information there are.

Fig. B1. (a) The mean and median of 10 000 single-sample estimates, $\hat{\alpha}$, as a function of the number n_{innov} of available observations. (b) The probability that a χ^2 variate is less than the mean of the distribution (labeled as theoretical) and the probability that $\hat{\alpha} < \alpha_*$ as determined experimentally (also as a function of n_{innov}).

is highly skewed; for $n_{innov} < 5$ more than 50% of the single estimates are equal to zero. The distribution has a very long tail to positive values. For larger values of n_{innov}, α_* is approached asymptotically by both α_{median} and α_{mean}.

For the extreme case where $n_{innov} = 1$, we expect the innovation to be smaller than $\sqrt{2.3}$ with probability $P(\chi^2_1 < 1)$ (≈ 0.683). In general, since the minimization of the log-likelihood function $\ln|2 + \alpha|^{n_{innov}} + (2 + \alpha)^{-1}|\nu|^2$ leads to $2 + \hat{\alpha} = |\nu|^2/n_{innov}$, we have

$$P(\hat{\alpha} < \alpha_*) = P(\chi^2_{n_{innov}} < n_{innov}). \tag{B2}$$

The probability on the right-hand side is the distribution function of a χ^2 distribution with n_{innov} degrees of freedom evaluated at n_{innov}. The probability on the left-hand

REFERENCES

Anderson, J. L., 1997: The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Mon. Wea. Rev.,* **125,** 2969–2983.

Anderson, T. W., 1984: *An Introduction to Multivariate Statistical Analysis.* 2d ed. Wiley and Sons, 675 pp.

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.

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

Coxeter, H. S. M., 1969: *Introduction to Geometry.* 2d ed. Wiley and Sons, 469 pp.

Dahlquist, G., and Å. Björck, 1974: *Numerical Methods.* Prentice-Hall, 573 pp.

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

——, 1992a: The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Wea. Rev.,* **120,** 164–177.

——, 1992b: Estimating model-error covariances for application to atmospheric data assimilation. *Mon. Wea. Rev.,* **120,** 1735–1746.

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

——, and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.,* **124,** 269–295.

——, and ——, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.,* **127,** 1822–1834.

——, G. Gaspari, C. Redder, L. Rukhovets, and A. M. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: Applications. *Mon. Wea. Rev.,* **127,** 1835–1849.

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.

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.

Haykin, S., 1996: *Adaptive Filter Theory.* 3d ed. Prentice-Hall, 989 pp.

Hoang, S., P. De Mey, O. Talagrand, and R. Baraille, 1997a: A new reduced-order adaptive filter for state estimation in high-dimensional systems. *Automatica,* **33,** 1475–1498.

——, R. Baraille, O. Talagrand, X. Carton, and P. De Mey, 1997b:

Adaptive filtering: Application to satellite data assimilation in oceanography. *Dyn. Atmos. Oceans,* **27,** 257–281.

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 L. Lefaivre, 1997: Using ensemble forecasts for model validation. *Mon. Wea. Rev.,* **125,** 2416–2426.

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

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

——, ——, and ——, 1996b: The RPN ensemble prediction system. *Proc. ECMWF Seminar on Predictability,* Vol. 2, Reading, Berkshire, United Kingdom, ECMWF, 121–146.

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.,* **109,** 701–721.

Lupton, R., 1993: *Statistics in Theory and Practice.* Princeton University Press, 188 pp.

Marshall, J., and F. Molteni, 1993: Toward a dynamical understanding of planetary-scale flow regimes. *J. Atmos. Sci.,* **50,** 1792–1818.

Maybeck, P. S., 1979: *Stochastic Models, Estimation and Control.* Vol. 1. Academic Press, 423 pp.

——, 1982: *Stochastic Models, Estimation and Control.* Vol. 2. Academic Press, 289 pp.

Mitchell, H. L., and R. Daley, 1997a: Discretization error and signal/error correlation in atmospheric data assimilation. Part I: All scales resolved. *Tellus,* **49A,** 32–53.

——, and ——, 1997b: Discretization error and signal/error correlation in atmospheric data assimilation. Part II: The effect of unresolved scales. *Tellus,* **49A,** 54–73.

Moghaddamjoo, A. R., and R. Kirlin, 1993: Robust adaptive Kalman filtering. *Approximate Kalman Filtering,* G. Chen, Ed., World Scientific, 65–85.

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

Preisendorfer, R. W., 1988: *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, 425 pp.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in FORTRAN. The Art of Scientific Computing.* 2d ed. Cambridge University Press, 963 pp.

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). Part II: Structure functions. *Quart. J. Roy. Meteor. Soc.,* **124,** 1809–1829.

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.,* **29,** 809–815.

Selten, F. M., 1993: Toward an optimal description of atmospheric flow. *J. Atmos. Sci.,* **50,** 861–877.

——, 1997: Baroclinic empirical orthogonal functions as basis functions in an atmospheric model. *J. Atmos. Sci.,* **54,** 2099–2114.

Shaw, D. B., P. Lönnberg, A. Hollingsworth, and P. Undén, 1987: Data assimilation: The 1984/85 revisions of the ECMWF mass and wind analysis. *Quart. J. Roy. Meteor. Soc.,* **113,** 533–566.

Stein, M., 1987: Large sample properties of simulations using Latin hypercube sampling. *Technometrics,* **29,** 143–151.

Stuart, A., and J. K. Ord, 1987: *Kendall's Advanced Theory of Statistics.* 5th ed. Vol. 1, *Distribution Theory,* Charles Griffin and Co., 604 pp.

Vannitsem, S., and C. Nicolis, 1997: Lyapunov vectors and error growth patterns in a T21L3 quasigeostrophic model. *J. Atmos. Sci.,* **54,** 347–361.

Wahba, G., D. R. Johnson, F. Gao, and J. Gong, 1995: Adaptive tuning of numerical weather prediction models: Randomized GCV in three- and four-dimensional data assimilation. *Mon. Wea. Rev.,* **123,** 3358–3369.

Weber, R. O., and P. Talkner, 1993: Some remarks on spatial correlation function models. *Mon. Wea. Rev.,* **121,** 2611–2617.

Yaglom, A. M., 1987: *Correlation Theory of Stationary and Related Random Functions.* Vol. 1, *Basic Results,* Springer-Verlag, 526 pp.

Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.,* **125,** 2274–2292.