Estimating Model-Error Covariances for Application to Atmospheric Data Assimilation

ROGER DALEY

Atmospheric Environment Service, Downsview, Ontario, Canada

(Manuscript received 9 September 1991, in final form 15 November 1991)

ABSTRACT

Forecast-error statistics have traditionally been used to investigate model performance and to calculate analysis weights for atmospheric data assimilation. Forecast error has two components: the model error, caused by model imperfections, and the predictability error, which is due to the model generation of instabilities from an imperfectly defined initial state. Traditionally, these two error sources have been difficult to separate.

The Kalman filter theory assumes that the model error is additive white (in time) noise, which permits the separation of the model and predictability error. Progress can be made by assuming that the model-error statistics are homogeneous and stationary, an assumption that is more justifiable for the model-error statistics than for the forecast-error statistics. A methodology for estimating the homogeneous, stationary component of the model-error covariance is discussed and tested in a simple data-assimilation system.

1. Introduction

The accuracy of forecasts from numerical prediction models depends on the intrinsic accuracy of the model and also on the accuracy with which the initial state is known. The growth of forecast error over periods of a few days or more has been the subject of numerous studies, and there is now a wide consensus that the details of the atmospheric flow cannot be deterministically predicted beyond a range of about 10 days.

Forecast error can also be studied over much shorter periods (one day or less). Such forecast-error statistics have great practical value in atmospheric data assimilation. This is because modern data-assimilation techniques attempt to estimate the state of the atmosphere by combining observations with a priori estimates obtained from forecast models. Optimal estimation theory requires some prior knowledge of the forecast-error statistics (in addition to observation error statistics).

Second-moment forecast-error statistics can be estimated from time invariant, homogeneous observation networks with spatially uncorrelated observation errors (i.e., radiosonde networks) by calculating the statistics of innovations (observations minus forecasts). These innovation covariance methods are discussed in Daley (1991, section 4.3) and generally yield homogeneous, stationary (and perhaps, isotropic) estimates of forecast-error covariances or correlations. Although such estimates have considerable validity over the region covered by the radiosonde network, they are of dubious value over the oceans and other data-sparse regions—see Ghil et al. (1981) and Daley (1992c, henceforth referred to as FOR92).

A possible solution to this difficulty is provided by Kalman filter theory in which the forecast-error covariances are predicted from the analysis and model-error covariances using the assimilation model itself. Although a full Kalman filter is not computationally feasible at the present time, simplified or suboptimal implementations are.

However, methods for estimating forecast-error covariances, based on Kalman filter theory, require good estimates of model-error statistics. Such statistics are not available at the present time, and it is the purpose of this work to develop a simple methodology for estimating them. As a first attempt, the methodology focuses on the homogeneous, stationary component of the model-error covariance. While the methodology is intended to be used with existing data-assimilation systems, it is based firmly on Kalman filter theory. Thus, the first task is to examine the Kalman filter equations.

2. Basic equations

The Kalman filter theory has been derived several times in the meteorological literature. See, for example, Ghil et al. (1981), Cohn and Parrish (1991), and Daley (1992c), which will henceforth be referred to as LAG91. The following derivation does not depart radically from these earlier derivations, but the emphasis is different, and a number of the intermediate results will be required in later sections.

A time-independent regular analysis grid \( r_j, 1 \leq j \leq J \) and \( s_n \) is defined to be the column vector of true values of all dependent (state) variables on the grid at
time \( t_n \). The variables \( s^f_n \) and \( s^a_n \) are defined to be the forecast and analyzed values, respectively, of \( s \) on the same grid.

Now assume a fixed observation network with \( I \) stations. Then, an expression for the analysis \( s^a_n \) at time \( t_n \) from the \( I \) observations \( d_n \) can be written

\[
s^a_n = s^f_n + \hat{K}_n (d_n - H s^a_n),
\]

(2.1)

where \( H \) is the (linear, time independent) \( I \times J \) forward interpolation matrix from analysis grid to observation network and \( \hat{K}_n \) is the \( J \times I \) matrix of analysis weights, referred to as the gain matrix. The quantity \( d_n - H s^a_n \) is called the innovation vector. Subtraction of the true values \( s_n \) from both sides of (2.1) gives

\[
e^a_n = (1 - \hat{K}_n H) e^a_n + \hat{K}_n e^p_n,
\]

(2.2)

where \( e^a_n = d_n - H s^a_n \) is the vector of observation errors and \( e^a_n = s_n - s^a_n \) and \( e^f_n = s^f_n - s_n \) are the vectors of analysis and forecast errors, respectively, at the analysis grid points. Taking the outer product of (2.2) and applying the expectation operator yields

\[
P^a_n = (1 - \hat{K}_n H) P_n^f (1 - \hat{K}_n H)^T + \hat{K}_n R \hat{K}_n^T,
\]

(2.3)

where \( R = (e^a_n e^a_n)^T \) is the \( I \times I \) diagonal observation-error covariance matrix. The quantities \( P_n^a = (e^a_n e^p_n)^T \) and \( P_n^f = (e^f_n e^f_n)^T \) are the \( J \times J \) analysis- and forecast-error covariances, respectively. It has been assumed here that the observation error is unbiased and neither serially correlated nor correlated with the model error. Superscript \( T \) indicates matrix transpose.

Minimizing the analysis-error variance at each of the \( J \) analysis grid points yields the optimal Kalman gain matrix (\( \hat{K}_n \), denoted without tilde),

\[
\hat{K}_n = P^a_n H^T (H P^a_n H^T + R)^{-1}.
\]

(2.4)

Substitution of (2.4) into (2.3) yields,

\[
P^a_n = (1 - \hat{K}_n H) P_n^f.
\]

(2.5)

Any data-assimilation system that uses a prior estimate or background field has an equation similar to (2.1). Statistical interpolation procedures calculate the analysis weights using a simplified form of (2.4). In some statistical interpolation procedures the analysis-error variance (diagonal elements of \( P^a_n \)) is estimated. However, the full analysis-error covariance equation (2.5) is rarely used in statistical interpolation schemes.

To complete the Kalman filter equations, we require predictive equations for the forecast state variable \( s^f_n \) and the forecast-error covariance \( P^f_n \). Define \( M_n \) to be a model (linear or nonlinear) such that

\[
s^f_{n+1} = M_n (s^a_n).
\]

(2.6)

Assume the model is not perfect and define the model-error growth between times \( t_n \) and \( t_{n+1} \) as \( e^p_n \). Assuming

that this error is purely additive, we can relate the true values at the two adjacent times by

\[
s^a_{n+1} = M_n (s^a_n) + e^p_{n+1}.
\]

(2.7)

Subtract (2.7) from (2.6), giving

\[
e^f_{n+1} = e^p_{n+1} - e^a_n,
\]

(2.8)

where,

\[
e^a_{n+1} = M_n (s^a_n) - M_n (s^a_n) - e^a_n
\]

(2.9)

will be referred to as the predictability error.

There are two terms on the right-hand side of (2.8). If the atmospheric state is known perfectly at time \( t_n \), then the forecast error at time \( t_{n+1} \) is due entirely to imperfections in the model \( e^p_n \). These model errors might be due to inadequacies in the model physical parameterization, boundary conditions, or finite resolution.

On the other hand, if the model is perfect (as is assumed in identical twin experiments), then the forecast error at time \( t_{n+1} \) is due entirely to \( e^a_n \), the error in the specification of the initial state at \( t_n \). Predictability theory is the study of the growth of small errors in nonlinear systems. It has been studied in the atmospheric sciences for many years and can be regarded as a subset of chaos theory. Predictability error \( e^p \) depends both on the initial error \( e^a \) and on the instabilities generated by the model \( M_n \). These instabilities will depend on the physics of the model and the particular initial state \( s^a_n \). The terms model error, predictability error, and forecast error are often used almost interchangeably in literature, but in this paper each type of error has been given a precise definition, which will be followed.

Consider now the expansion of the second right-hand side term of (2.9) in the first two terms of a Taylor series about \( s^a_n \). Then, (2.9) can be written approximately as

\[
e^a_{n+1} \approx \frac{\partial M_n (s^a_n)}{\partial s^a_n} e^a_n = M_n e^a_n,
\]

(2.10)

where \( M_n \) is the Jacobian matrix whose elements are the partial derivatives of the model applied to the state \( s^a_n \) with respect to the elements of \( s^a_n \). The Jacobian \( M_n \) is a \( J \times J \) matrix whose elements depend on \( s^a_n \).

Approximation (2.10) is known as the tangent linear approximation and was studied by Lacarra and Tala-grad (1988) using a nonlinear barotropic \( f \)-plane shallow-water model. They found that the tangent linear approximation was valid for integrations of up to 48 h, particularly for the large-scale Rossby modes. The tangent linear approximation is the basis for the extended Kalman filter and the four-dimensional variational (adjoint) method (Lacarra and Tala-grad 1988).

Now insert (2.10) into (2.8), take the outer product, and apply the expectation operator. The result is

\[
P^a_{n+1} = P^a_n + Q_n,
\]

(2.11)
where \( Q_n = \langle \epsilon_h^2 (\epsilon_q^2)^T \rangle \) is the model-error covariance and
\[
P_{n+1}^* = M_n P_n^{*} M_n^T
\]
(2.12)
is the predictability-error covariance. In deriving (2.11), it has been assumed that the model error is spatially, but not serially, correlated. In other words, the model error is "white" in time. As discussed in Daley (1992a), this assumption implies that the cross term \( M_n (\langle \epsilon_h \epsilon_q^2 \rangle)^T \) and its transpose are zero. It has also been assumed that the model error is unbiased (zero mean) and is uncorrelated with the observation error. Under these conditions the forecast error will be unbiased when the model is linear.

The matrix \( P_{n+1}^* \) depends on the properties of the model, the properties of the atmosphere, and the characteristics of the observing network (because of \( P_n^{*} \)). The model-error covariance \( Q_n \) depends only on the properties of the atmosphere and the model. As discussed in FOR92, \( P_{n+1}^* \) is normally very inhomogeneous, anisotropic, and nonstationary in good part due to the inhomogeneous, time-dependent global observing network. Although there are good reasons to believe that the model-error covariance \( Q_n \) itself is not necessarily homogeneous, isotropic, stationary, etc., the characteristics of the global observing system will clearly have no effect on it.

The Kalman filter formulation of the atmospheric data-assimilation problem consists of (2.1), (2.4), (2.5), (2.6), (2.11), and (2.12). Unfortunately, there are several practical difficulties with the Kalman filter that must be overcome before it becomes operationally feasible. However, improvements to operational systems based on Kalman filter theory are practical now. In particular, simplified forms of (2.11) and (2.12) would be very attractive in operational statistical-interpolation systems.

Statistical interpolation is the most widely used data-assimilation technique at the present time. This technique requires the forecast-error covariance to be specified in some fashion. In operational applications it is frequently necessary to make strong, often unjustifiable, assumptions about the forecast-error covariance. Some common assumptions are homogeneity, stationarity, isotropy, geostrophy, and vertical and horizontal separability (see Daley 1991, section 4.3). The attraction of actually predicting the forecast-error covariance using (2.11) and (2.12) is that it may be more justifiable to make these assumptions 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.

There is, however, a fundamental difficulty: we must have a good estimate of the model-error covariance \( Q_n \) in order to successfully exploit (2.11) and (2.12). This is true both for the Kalman filter itself and any suboptimal system derived from it.

There are two other reasons why good estimates of \( Q_n \) are desirable. First, it has always been difficult to separate model error from forecast error, and (2.11) and (2.12) permit us to do this, in principle. Second, techniques, such as the four-dimensional variational method, assume that the model is perfect. Determination of \( Q_n \) would allow the accuracy of this assumption to be assessed.

Dee et al. (1985) suggested a procedure for estimating the model-error covariance based on the application of an adaptive Kalman filter. It is the main purpose of this paper to derive an alternate procedure for estimating the stationary, homogeneous component of the model error, based on an extension of the zero-lag innovation covariance procedure for estimating forecast-error statistics (see LAG91). This methodology is simple enough to be used with existing data-assimilation systems. The methodology is developed in section 3 and tested with simple models in sections 4 and 5.

3. Estimating \( Q \)—methodology

This technique for estimating model error is related to a recent idea of Bloom and Schubert (1990). As noted earlier, it cannot be assumed that the model-error covariance is stationary and homogeneous. However, it is possible to estimate the homogeneous, stationary component of the model error.

More concretely, suppose \( r \) is the spatial coordinate and \( \tilde{C}_n(r, r + \Delta r) \) is the covariance of some variable between points \( r \) and \( r + \Delta r \) at time \( t_n \). Then, we can extract the homogeneous component of this covariance by integration over the domain
\[
\tilde{C}_n(\Delta r) = \frac{\int r \tilde{C}_n(r, r + \Delta r) dr}{\int dr}.
\] (3.1)

The stationary component \( \tilde{C}(\Delta r) \) is obtained by averaging over time. In practice, covariances are known at discrete spatial locations, and the integral of (3.1) would be replaced by an appropriate numerical quadrature.

The methodology for estimating \( \tilde{Q} \) is an extension of the (lag-zero) innovation covariance method that has been frequently used in the past to estimate forecast-error statistics. This technique is traditionally applied over radiosonde networks, particularly the North American radiosonde network. Such networks have three properties that make the innovation technique viable. These properties are fixed observation locations, spatially uncorrelated observation errors, and homogeneity of the network. (In a homogeneous observation network, the observations stations are irregularly located, but the data density is approximately uniform—see FOR92).
The estimated forecast-error statistics, obtained by this technique, are only strictly valid over the domain sampled by the radiosonde network. Traditionally, these statistical results are then extrapolated to the remainder of the global domain by simple (ad hoc) extensions. Thus, although the estimated forecast-error statistics over the North American radiosonde network may not be perfect, they are considerably better than the corresponding statistics for most of the rest of the globe. In the same way, although the estimation of $\hat{Q}$ obtained over the radiosonde network using the proposed methodology would be only strictly valid in the domain sampled by the network, it would be used over the whole global domain.

Consider the homogeneous, stationary form of (2.11),

$$\hat{Q} = \hat{P}^f - \hat{P}^p. \tag{3.2}$$

The methodology is based on the simple idea that separate estimates of $\hat{P}^f$ and $\hat{P}^p$ are subtracted to obtain an estimate of $\hat{Q}$. We assume that there exists a data-assimilation procedure where the analyzed values are produced by (2.1) from observations obtained from a radiosonde network and forecasts obtained from (2.6) using a model $M_n$. We further assume that the gain matrix depends on specified forecast- and observation-error covariances and is either of the form (2.4) or is some rational approximation to (2.4), such as those used in statistical interpolation procedures. The additional Kalman filter equations (2.5), (2.11), and (2.12) may also be used, but are not necessary.

a. **Step 1—estimating $\hat{P}^f$**

The first step is a straightforward application of the lag-zero innovation covariance technique to the estimation of the homogeneous, stationary component of the forecast-error covariance over the radiosonde network. A long time series of innovations $d_n - Hs_n^f$ is collected at the observation stations; the biases are removed and then correlated, producing the innovation covariance.

When the observation error is not correlated with the forecast error, the lag-zero innovation covariance can be written (in the notation of LAG91) as

$$C_n^e = \langle (d_n - Hs_n^f)(d_n - Hs_n^f)^T \rangle = \text{HP}_n^f H^T + R, \tag{3.3}$$

where $\text{HP}_n^f H^T$ is the forecast-error covariance between the observation stations. After time averaging, the homogeneous component of the innovation covariance is then determined by plotting all its elements as a function of displacement between observation locations and objectively or subjectively analyzing them. This stationary, homogeneous innovation covariance may be anisotropic, but its isotropic component can be determined by averaging azimuthally (see Daley 1991, section 4.3). The fact that the observation errors are neither spatially correlated nor correlated with the forecast errors can be exploited to separate the homogeneous, stationary component of the forecast $\text{HP}_n^f H^T$ and observation $R$ error covariances. We assume, following normal practice, that the homogeneous, stationary components of $\text{P}^f$ and $\text{HP}_n^f H^T$ are the same. This practice is acceptable if the network is sufficiently dense. Note that the forecast-error covariance, estimated in this way, can be obtained for any data-assimilation scheme; it requires neither the complete Kalman filter equations nor an optimal gain matrix.

We must now estimate the second right-side term of (3.2). From (2.11) and (2.12) it is clear that there must be two steps; estimating the analysis-error covariance itself and then estimating the predictability-error covariance from it.

b. **Step 2—estimating $\hat{P}^p$**

We consider first estimating the homogeneous, stationary component of the analysis-error covariance. There are several ways of doing this, but we always assume that we already have available estimates of the stationary, homogeneous components of the forecast-error covariance $\hat{P}^f$ and the observation-error covariance $\hat{R}$ from step 1.

The simplest way of estimating $\hat{P}^p$ is by application of (2.5) to calculate $\hat{P}^p_n$ and then averaging spatially and temporally. Unfortunately, most data-assimilation schemes do not include (2.5).

However, there is another method of obtaining $\hat{P}^p$ for use in data-assimilation procedures that does not require (2.5). The method is based on the observed-minus-analysis procedure developed by Hollingsworth and Lönnberg (1989) for assessing the performance (closeness to optimality) of statistical interpolation systems. We will use this procedure for a quite different purpose. The idea is very similar to the lag-zero innovation covariance technique, except this time the residuals $d_n - Hs_n^p$ are required. Following LAG91, we will refer to this as a (zero-lag) residual covariance technique. The lag-zero residual covariance is denoted as

$$D_n^e = \langle (d_n - Hs_n^p)(d_n - Hs_n^p)^T \rangle = \langle (e_n^p - Hc_n^p)(e_n^p - Hc_n^p)^T \rangle = Z_n + H K_n Z_n H^T = H K_n Z_n H^T, \tag{3.4}$$

where $Z_n = R + \text{HP}_n^f H^T$, using (2.2). Assuming optimality, $K_n$ is given by (2.4), and (3.4) becomes

$$D_n^e = R - H(1 - K_n H) P_n^f H^T = R - \text{HP}_n^p H^T, \tag{3.5}$$

using (2.5). This equation is identical to (6b) of Hollingsworth and Lönnberg (1989) when $H = 1$. Assuming
\( \hat{\mathbf{R}} \) has been estimated (from step 1), then the analysis error at observation stations is given by
\[
\mathbf{H} \mathbf{P}^\text{a}\mathbf{H}^T = \hat{\mathbf{R}} - \mathbf{D}^n_0.
\] (3.6)

Spatial and temporal averaging of \( \mathbf{D}^n_0 \) leads to an estimate of the homogeneous, stationary component of the analysis-error covariance at observation stations.

One significant difficulty with the application of (3.6) is that, in general, \( \mathbf{P}^\text{a}\mathbf{H}^T \) < \( \mathbf{H} \mathbf{P}^\text{a}\mathbf{H}^T \). This occurs because at observation stations there are colocated observations and the analysis error has a local minimum. In other words, we cannot simply ignore the distinction between calculating covariances at observation stations and calculating them at grid points, as we did in step 1.

However, although application of (3.6) to estimate analysis-error variances at grid points may not be a good idea, this equation may be used to estimate the homogeneous, stationary analysis-error correlations at grid points. We would obtain the analysis-error variances at grid points from the diagonal elements of (2.5). As noted earlier, the diagonal elements of (2.5) are estimated in some statistical interpolation schemes, which makes the methodology practical with existing systems.

Now (2.5) or (3.6) only produce accurate estimates of the analysis error covariance if the gain matrix is optimal or near-optimal, which, in turn, requires that \( \mathbf{P}^\text{a} \) and \( \hat{\mathbf{R}} \) be specified accurately. There was no requirement for optimality in step 1, and thus, estimates of analysis-error covariances are likely to be less satisfying than those of forecast-error covariances. Attempting to achieve near-optimality in step 2 requires an iteration in step 1. Thus, after estimates \( \mathbf{P}^\text{a} \) and \( \hat{\mathbf{R}} \) are obtained in step 1, they should be inserted into (2.4) or whatever approximation is used to calculate the gain matrix. This new gain matrix is then used in a repeat of step 1 to calculate a new estimate \( \mathbf{P}^\text{a} \). This is repeated until the forecast-error variances averaged over the domain of the radiosonde network, which have been estimated by the innovation covariance technique, are as close as possible to the values specified in (2.4). At this point, the system is as close to optimal as is possible with a homogeneous, stationary estimate of the forecast-error covariance.

c. **Step 3—estimating \( \mathbf{P}^\text{a} \)**

The last step is the estimation of the stationary, homogeneous component of the predictability error covariance from the estimated stationary, homogeneous analysis-error covariance. If we had a complete Kalman filter system including an equation for predicting the forecast-error covariance, we could estimate the predictability-error covariance using (2.12).

However, we do not require (2.12) to estimate the homogeneous, stationary component of the predictability-error covariance. At any time \( t_n \),
\[
\mathbf{P}_n^p = \left( \mathbf{M}_n \mathbf{e}_n^a (\epsilon_n^a)^T \right) \mathbf{M}_n^T
\]
\[
\approx \left[ \mathbf{M}_n (s_n^a) - \mathbf{M}_n (s_n^a - \epsilon_n^a) \right] \times \left[ \mathbf{M}_n (s_n^a) - \mathbf{M}_n (s_n^a - \epsilon_n^a) \right]^T,
\] (3.7)

where \( \mathbf{M}_n \) is the original (possibly nonlinear) model of (2.6). All we wish to determine is \( \mathbf{P}_n^p \), the stationary, homogeneous component of \( \mathbf{P}_n^p \), and we do this using (3.7) in a standard predictability experiment in which we not only determine the predictability-error-growth variance but also the covariance. To do this, we select an ensemble of \( M \) initial conditions from the past, \( s_m^a, 1 \leq m \leq M \) (Note that the choice of index \( M \) is not to be confused with \( M \)—the model.) We wish to perturb each of these initial states with an ensemble of random perturbations \( \epsilon_{m,k}^a \), \( 1 \leq k \leq K \) that are each chosen from a population that has the homogeneous covariance \( \mathbf{P}^p \). Thus, these perturbations will be spatially correlated, which means their spectrum is not, in general, white.

We then calculate the difference fields \( \mathbf{M}_n (s_m^a) - \mathbf{M}_n (s_m^a - \epsilon_{m,k}^a) \) and calculate the variances and covariances of this difference field. We then average these covariances over all the \( M \times K \) cases. The final step is to average over the domain to produce the homogeneous, stationary estimate \( \mathbf{P}_n^p \).

Finally, subtraction of \( \mathbf{P}_n^p \) from \( \mathbf{P}_n^\text{a} \), following (3.2), yields the homogeneous, stationary estimate \( \mathbf{Q} \). In the next two sections we will examine the three steps using two simple models.

4. **Estimating \( \mathbf{P}^\text{a} \) and \( \mathbf{P}^p \)—experiment**

The methodology proposed in section 3 is complicated and relies on several assumptions. It is not clear, a priori, how successful it is likely to be in practice. Consequently, two simple experiments have been performed that will help to clarify some of the ideas of section 3 and illuminate some of the potential problems with the methodology. In this section, we will perform steps 1 and 2 (estimation of \( \mathbf{P}^\text{a} \) and \( \mathbf{P}^p \)) using the one-dimensional quasigeostrophic model described in sections 4 and 5 of LAG91.

The model \( \mathbf{M} \) is linear and time independent. We assume a periodic domain \( 0 \leq x \leq 2\pi a \) and specify \( a = 2500 \) km. The Coriolis parameter \( f_0 = 10^{-4} \text{ s}^{-1} \), the number of grid points \( J = 49 \), and the grid length \( \Delta x = 320 \) km. We set the Courant number \( U \Delta t / \Delta x = 1 \), where \( U \) is the constant advection velocity and \( \Delta t \) is the time between data insertions. We will specify model- and observation-error statistics as follows.

The model-error covariance \( \mathbf{Q} \) is defined following sections 4 and 5 of LAG91 and section 2 of FOR92. Denote \( \mathbf{F}_x \) as the transform between Fourier and real space and define
\[
\mathbf{Q} = \mathbf{F}_x \mathbf{Q} \mathbf{F}_x^T, \quad (4.1)
\]
where \( \mathbf{Q} \) is a diagonal \( J \times J \) matrix of spectral variances.
\((E_h^0)^2 \tilde{q}(p)\) obtained from (2.2) of FOR92, with specified length scale \(l_q = a/6\) and specified variance \((E_h^0)^2\). The model-error covariance \(Q\) defined in (4.1) is actually the height--height covariance and is stationary, homogeneous, and isotropic. Wind--wind and height--wind covariances can be constructed from (4.1) geostrophically following FOR92.

The observation network consists of \(I\) observation stations with spatially uncorrelated observation errors. There are height observations (no wind observations) at each station with rms height error \(E_h^0 = 10\) m.

Before commencing the main experiment, we will first perform a preliminary experiment that will help to clarify the relationship between the forecast- and model-error covariances. This was explored previously for the asymptotic, scalar case by Ghil et al. (1981, pp. 206–211). For this preliminary experiment, we assume that there are \(I = 49\) observation stations that coincide with the analysis grid (\(I = 1\)). We integrate the second moment Kalman filter equations (2.4), (2.5), (2.11), and (2.12) to equilibrium, which produces forecast- and analysis-error covariances \(\mathbf{P}^f\) and \(\mathbf{P}^a\) that are stationary and homogeneous. As noted in LAG91 (for this simple model), if the analysis-error covariance is homogeneous, then \(\mathbf{P}^a = \mathbf{M}^a \mathbf{M}^a \mathbf{T} = \mathbf{P}^a\).

The results are plotted in Fig. 1. Two cases are shown: an inaccurate model \((E_h^0)^2 = 1000\) m\(^2\) (Fig. 1a) and an accurate model \((E_h^0)^2 = 10\) m\(^2\) (Fig. 1b). In each panel, the forecast-error covariance \(P_{hh}^f\), the analysis-error covariance \(P_{hh}^a = P_{hh}^a\), and the model-error covariance \(Q_{hh}\) of the height field are plotted as a function of displacement.

Now suppose that over this network we already have obtained (by innovation covariance techniques—step 1) some estimate of the homogeneous component of the forecast-error covariance \(\mathbf{P}^f\). Then, Fig. 1 shows that we already possess a good (poor) estimate of \(Q\) if the model is accurate (inaccurate). As \(Q\) decreases, \(\tilde{q}\) and \(\mathbf{P}^f\) become increasingly different at small displacements, but remain essentially equal at large displacements. Thus, simply assuming that the model-error covariance is the same as the forecast-error covariance is likely to be a good approximation at large displacements, but it is likely to be substantially in error at small displacements if the model is accurate. Multivariate calculations with the same model (not shown) indicate that this result remains true even when there are both wind and height observations.

Now the main experiment of the this section is described. The model is as described earlier, with the same values of \(a, f_0, J,\) and \(U\Delta t/\Delta x\). We assume that the state variables are forecast using (2.6), except that the model \(\mathbf{M}\) is linear and time independent in this case. The true states satisfy (2.7), and the error satisfies (2.8). Since the model is linear, it is not necessary to construct a “truth” in this experiment, and consequently, we will only consider the integration of (2.8).

We will specify the model error \(\epsilon_h\) in the following way:

\[
\epsilon_h^g = \mathbf{F} \xi_h^g, \tag{4.2}
\]

where \(\xi_h^g\) is a column vector of length \(J\) of spectral coefficients chosen randomly from an unbiased normal distribution with expected variance \((E_h^0)^2 \tilde{q}(p)\) for each wavenumber. We specify \(l_q\) as before, and set \((E_h^0)^2 = 100\) m\(^2\).

We assume that the analyzed values are related to the observed and forecast values by (2.1). The gain matrix \(\mathbf{K}\) is assumed to be time independent and to have the form (2.4); that is,

\[
\mathbf{K} = \mathbf{P}^f/\mathbf{H}^T(\mathbf{H} \mathbf{P}^f/\mathbf{H}^T + \mathbf{R})^{-1}, \tag{4.3}
\]

in which \(\mathbf{P}^f\) and \(\mathbf{R}\) are estimates that are specified in some way. Since \(\mathbf{H}\) is linear, we again subtract off the “truth” and deal with (2.2) instead of (2.1).

We assume that there are \(J = 32\) observation stations at which there are height observations (no wind ob-

---

**Fig. 1.** Height-error covariances (m\(^2\)) as a function of displacement \(x\) (km) for two cases: (a) \((E_h^0)^2 = 1000\) m\(^2\) and (b) \((E_h^0)^2 = 10\) m\(^2\).
servations) with expected rms observation error $E^o_d$.
The observation network is fixed and homogeneous
with spatially and serially uncorrelated and unbiased
observation errors. On this network, the observation
density is approximately uniform. In this way, we
mimic a radiosonde network (apart from the lack of
wind observations). The forward-interpolation oper-
ator $H$ is exactly the same as described in section 4 of
FOR92. The observation errors $e^o_d$ are chosen from an
unbiased normal distribution with expected variance
$(E^o_d)^2 = 100 \text{m}^2$.

In summary, the main experiment consists of inte-
grating (2.8), with $e^o_d$ defined in (2.10), to produce the
forecast error from the analysis and model errors and
(2.2) to calculate the analysis error from the forecast
and observation errors. We assume initially that the
observation and model errors (and their statistics) are
unknown and that the estimates $\hat{P}$ and $\hat{R}$ required
to determine the gain matrix [(4.3)] have been specified
in an ad hoc fashion.

Step 1 is to calculate the (lag-zero) innovation co-
variance. We assume that in an operational system this
particular calculation would be done in real time and
that the sample size (i.e., the number of data assimila-
tion cycles) would be quite large. Thus, the experi-
ment was run for $N = 360$ time steps.

The innovations $d_n - Hs^a_n = e^o_n - Hs^a_n$ were calculated
at each observation station, the sample mean was re-
moved, and then the innovations were correlated. The
values were plotted on a scatter diagram as a function
of $|x|$, where $x$ is the displacement between observa-
tion stations. The use of $|x|$, rather than $x$ itself, is the
one-dimensional analog of finding the isotropic com-
ponent of the covariance.

The height innovation covariance ($\text{m}^2$) is plotted as
a function of $|x|$ in kilometers (abscissa) in Fig. 2a.
The covariance between each station pair is plotted as a
dot. The dashed line at $269 \text{m}^2$ indicates the inno-
vation variance averaged over the I observation sta-
tions. It was noted that the scatter increased (not
shown) as $N$ decreased due to increased sampling error,
but there was very little reduction in scatter when $N$
exceeded 200. This is because the actual forecast-error
statistics are not homogeneous in this system, as the
observation network is not uniform.

The innovation covariance procedure used in this
experiment is somewhat simpler than those used op-
erationally. Thus, covariances rather than correlations
have been used. The covariances have been grouped
into bins of approximately equal $|x|$; and the Fisher
Z transform (see Hollingsworth and Lönnberg 1989)
has not been used to reduce the sampling error in cor-
relations that are significantly different from zero.
However, the sample size, $N = 360$, is probably suffi-
ciently large that these refinements are not necessary.

By averaging over $N = 360$ assimilation cycles, a sta-
tionary, homogeneous (and isotropic) estimate of the
innovation covariance $\hat{P}^{-1}/H^T + \hat{R}$ is produced at
observation stations [(3.3)]. Here $\hat{P}$ and $\hat{R}$ are the
estimated forecast- and observation-error covariances
that are to be determined. Following section 3, the ho-
mogeneous, stationary components $\hat{P}$ are equated with
$\hat{P}^{-1}/H^T$. Since the observation error is not spatially cor-
related, the innovation covariance at nonzero dis-

![Fig. 2. Height-error covariances (\text{m}^2) as a function of $|x|$ for the experiment of section 4. The solid curves show the fits to the scattered points: (a) forecast-error covariance and (b) analysis-error covariance.](image-url)
placements must represent the forecast-error covariance. Thus, the homogeneous, stationary, isotropic component of the forecast-error covariance is estimated by fitting a smooth curve through the scattered points of Fig. 2a and extending it to $|x| = 0$. In this case, the solid curve of Fig. 2a was fitted manually. From the intercept on the ordinate we can read off the forecast-error variance as $(E_x^o)^2 = 170$ m$^2$ and the observation-error variance as $(E_x^o)^2 = 269$ m$^2 - 170$ m$^2 = 99$ m$^2$.

Figure 2a actually represents the final iteration of the iteration procedure described under step 2 of section 3. Thus, in the first step of the iteration, ad hoc values of $\hat{R}$ and $\hat{P}^f$ were specified in order to calculate the weight matrix from (2.4). At the next iteration step, these ad hoc values were replaced by estimates of $\hat{R}$ and $\hat{P}^f$ obtained from the innovation covariance procedure. The iteration continued until the specified values of the observation- and forecast-error covariances agreed approximately with those obtained from the innovation covariance procedure. This took about three iterations in this case. In operational practice, this sort of iterative procedure is done naturally, as the forecast- and observation-error covariances that are estimated by the innovation procedure are used to update the corresponding specified quantities that are used in the assimilation system. From Fig. 2a it can be seen that the estimated observation-error variance (99 m$^2$) is actually quite close to the specified value (100 m$^2$).

After completing the iteration of step 1, the residual covariance (3.4) was calculated in exactly the same way as the innovation covariance using $N = 360$ samples. Presumably, in operational practice, the residual covariance could be calculated in real time as is the innovation covariance. Following (3.6), the residual covariance is subtracted from the observation-error covariance $\hat{R}$ estimated in step 1. The resulting analysis-error covariances between the observation station pairs $\mathbf{H} \mathbf{P}^a \mathbf{H}^T$ are plotted as dots in Fig. 2b. The value at $|x| = 0$ is obtained by averaging the variance over the $l$ observation stations. This value is 51 m$^2$. Following step 1, we fit a solid curve through the scattered points and the value at $|x| = 0$ to estimate $\hat{P}^a$.

By way of comparison, we also estimated $\hat{P}^a$ using (2.5) directly. The result is plotted as the dashed curve in Fig. 2b. The changes in slope of the dashed line occur at the grid points, which are 320 km apart. The analysis-error variance determined by (2.5) is 66 m$^2$. The residual covariance method (3.6) and the analysis-error covariance (2.5) both produce covariances with a smaller characteristic scale (Fig. 2b) than the forecast-error covariance (Fig. 2a), which is to be expected for the univariate height case (FOR92). However, we note, as suggested in section 3, that the residual covariance method considerably underestimates the analysis-error variance (at the analysis grid points).

As noted earlier, the simple one-dimensional linear advection model has the property that for homogeneous analysis-error covariance $\hat{P}^a = \mathbf{M} \hat{P}^f \mathbf{M}^T = \hat{P}^f$. Performing step 3 (estimation of $\hat{P}^a$) with this model is therefore trivial. Provided enough cases are run to minimize the sampling error, the result will always be $\hat{P}^a = \hat{P}^f$.

The final step was performed by calculating $\mathbf{Q}$ from $\hat{P}^f$ and $\hat{P}^a = \hat{P}^f$, using (3.2). In this case, $\hat{P}^f$ was determined from Fig. 2a. The analysis-error correlation was determined by the residual covariance method (solid curve of Fig. 2b) and the analysis-error variance from the zero intercept of the dashed curve of Fig. 2b [(2.5)]. The results are plotted in Fig. 3. The actual height–height model-error covariance $\mathbf{Q}$ (solid curve) is plotted as a function of displacement $|x|$ (abscissa). The estimate $\hat{Q}$ is plotted as the dashed curve. In the next section, an attempt is made to perform step 3 of section 3 using a more realistic model.

5. Estimating $\hat{P}^a$ from $\hat{P}^f$—experiment

Step 3, the estimation of the homogeneous, stationary component of the predictability-error covariance $\hat{P}^a$ from the homogeneous, stationary component of the analysis-error covariance $\hat{P}_a$, could not be adequately tested with the simple linear model of section 4. In that model, although the predictability error could grow locally for a short time due to advection from upstream, there could be no domain-averaged error growth because the model was neutral with constant coefficients. In more realistic atmospheric models the situation is very different. These models are nonlinear, which means that there is invariably domain-averaged predictability-error growth. If looked at from the tangent linear perspective, the action of the linearized

![Fig. 3. Model height-error covariance (m$^2$) as a function of displacement $|x|$. Solid line is the true model-error covariance; dashed line is the estimate.](image)

Unauthenticated | Downloaded 10/21/23 04:00 PM UTC
model on a given initial state generates local instabilities, and thus, the predictability error can grow locally through instability as well as advective effects. Since these important effects could not be captured with the model of section 4, step 3 is examined with a nonlinear model.

Step 3 essentially consists of a standard predictability experiment and is closely related to the experiment of Bloom and Schubert (1990). This sort of experiment has been done many times in the past. The particular experiment described here is oriented toward data assimilation, and thus, the emphasis is on the short-term (24 h or less) predictability-error growth. Most predictability experiments choose perturbations from a white-noise spectrum, whereas in this experiment, the perturbations will be chosen from several red-noise spectra. Since analysis error normally has a red-noise spectrum, this choice is more realistic. Finally, although most predictability experiments are concerned with the estimation of predictability-error variances, this one will examine covariances as well.

The present experiment was performed with a nonlinear, spherical, constant-absolute-vorticity model (Daley 1982). The governing equation is

$$\frac{\partial \nabla^2 \psi}{\partial t} = -\nabla \cdot (\nabla^2 \psi + f) \psi, \quad (5.1)$$

where $\psi$ is the streamfunction, $\nabla^2$ is the spherical Laplacian operator, $\nabla \cdot$ is the spherical divergence operator, $f$ is the (latitudinally variable) Coriolis parameter, and $\psi = k \times \nabla \psi$. Here $\psi$ is the nondivergent wind, $k$ is the vertically pointing unit vector, and $\nabla$ is the spherical gradient operator.

In order to make the results more compatible with the results of the previous sections, a pseudoheight is introduced:

$$h = f_0 \psi / g, \quad (5.2)$$

where $g$ is the gravitational constant and $f_0 = 10^{-4}$ s$^{-1}$ is the same fixed value of the Coriolis parameter used in section 4. From now on we will refer to the results in terms of heights and height error statistics, but it must be remembered that we are really talking about streamfunctions.

The model is integrated using a leapfrog algorithm in time and a spectral transform technique to calculate the nonlinear terms. The application of the spectral transform method requires that the model variables be expanded in a finite series of spherical harmonic functions,

$$h(\lambda, \phi) = \sum_{m=0}^{N} \sum_{m=-n}^{n} h_m^n P_m^n (\sin \phi) \exp (im\lambda), \quad (5.3)$$

where $\lambda$, $\phi$ are longitude and latitude, $m$ is the zonal wavenumber, $n$ is the two-dimensional wavenumber, $h_m^n$ is the spectral expansion coefficient, and $P_m^n$ is the associated Legendre polynomial of degree $m$ and order $n$. The parameter $N$ defines the limit of the expansion and implies that (5.3) is a triangular truncation denoted $T_N$.

The model was integrated from three initial global height (really, streamfunction) fields. These were the 0000 UTC 1 January European Centre for Medium-Range Weather Forecasts (ECMWF) 500-mb analyses for 1986, 1987, and 1988.

There were nine initial perturbation fields chosen; three for each initial time. These initial perturbations were chosen in the following way. The perturbations were supposed to be representative of the analysis error that should be specified to be homogeneous and isotropic according to the methodology of section 3. To specify such perturbations properly requires reference to the theory of homogeneous turbulence on the sphere, which has been described by Boer (1983) and references therein. All of the relevant mathematical formulas and derivations are in Boer (1983); hence, we will only discuss them briefly here.

Homogeneous, isotropic turbulence on the sphere is a function only of $S$, the great-circle distance between two points, defined by

$$\cos(S/r_e) = \sin \phi_1 \sin \phi_2 + \cos \phi_1 \cos \phi_2 \cos(\lambda_2 - \lambda_1), \quad (5.4)$$

where $\phi_1$, $\phi_2$ and $\lambda_1$, $\lambda_2$ are the latitudes and longitudes, respectively, of the two points, $S$ is the great-circle distance between them, and $r_e$ is the radius of the earth.

Boer (1983) shows that random fields expanded in the form (5.3) are isotropic and homogeneous on the sphere if (a) there is no correlation between different spectral coefficients and (b) the expected variance of each spectral coefficient with the same value of $n$ is the same. Condition (b) implies that the real and imaginary (or sine and cosine) components for the same wavenumber have the same expected variance.

Thus, the expected variance for a homogeneous, isotropic field can be written simply as a function of $n$. As shown by Boer (1983), this spectral variance can be transformed into an isotropic, homogeneous covariance matrix (function of $S$ only) by a transform involving only the ordinary Legendre polynomials $P_n^m$. Particular care must be taken with the normalization factor $(2n + 1)/2$.

The isotropic, homogeneous initial perturbation covariance was chosen to have the form

$$C_{hh}(S) = (E_h)^2 (1 + S/l_h) \exp(-S/l_h), \quad (5.5)$$

where $E_h$ is the rms analysis error and $l_h$ is a specified analysis-error scale, which will be varied to control the redness of the spectrum. The corresponding spectral variance $\hat{c}_h(n)$ is calculated from (5.5) using Gaussian quadrature. Finally, initial perturbations $\Delta h_n^n$ are chosen randomly from an unbiased normal distribution with expected variance $\hat{c}_h(n)$. 

Unauthenticated | Downloaded 10/21/23 04:00 PM UTC
For each of the nine cases, we integrate the model forward 24 h from one of the three initial states. We then add one of the perturbation fields to one of the initial states and perform the integration again. This makes a total of 12 model integrations. We then calculate the difference between each of the perturbed and unperturbed model integrations [(2.9)] at the end of each 6-h period and store the difference fields in the form of spectral coefficients. The global mean coefficient ($n = 0$) is time independent in (5.1), and consequently, the $n = 0$ perturbation coefficient is also time independent.

We calculate the spectral variance of the difference fields for each spectral coefficient, for each of the nine
cases every 6 h. We then average over the nine cases. The final step is to average over all spectral coefficients with the same value of \( n \), thus producing the homogeneous, isotropic component of the difference fields (in spectral form). These spectral variances can then be transformed into real space covariances that are functions only of \( S \).

Figure 4 shows three such experiments performed with a model in which \( N = 30 \) (i.e., \( T_{30} \) truncation) and where the rms analysis height error at \( t = 0 \) is \( E_h = 10 \) m. All panels show the homogeneous, isotropic component of the difference field at 0 and 24 h. The 24-h period, rather than the more usual 6-h period, was chosen in order to enhance the difference. Figures 4a, 4c and 4e show the difference spectra \( \hat{h}_l(n, t) \) on a log–log plot as a function of \( n \), while Figs. 4b, 4d, and 4f show the corresponding difference covariances \( C_{hh}(S, t) \) as a function of \( S \).

Figures 4a–b show the case of an initial white-noise spectra; that is, \( h_i \rightarrow 0 \) in (5.5). For the spectral plots, the total error variance is shown summed over all spectral coefficients with the same value of \( n \), which means summing over \( 2n + 1 \) coefficients. This implies that for a white-noise spectrum \( \hat{C}_{hh}(n, 0) \) is proportional to \( n \), which is evident from Fig. 4a. The corresponding covariance is shown as the \( t = 0 \) curve in Fig. 4b. This curve represents the covariance of a white-noise spectrum truncated at \( N = 30 \). A higher (lower) resolution model would have a narrower (wider) covariance in this case. Figures 4c–d and Figs. 4e–f are in the same format as Figs. 4a–b and represent redder (larger values of \( h_i \)) initial perturbation spectra.

It is quite evident from Fig. 4 that the perturbation growth rate is strongly affected by the spectrum of the initial analysis error. Clearly, a white-noise perturbation (i.e., analysis error) grows more rapidly than a red-noise perturbation.

We will now examine this phenomena more systematically. First define \( \tilde{E}_h^2(t) = C_{hh}(0, t) \). Also, a turbulent microscale \( L_h(t) \) is defined following Daley (1991a) and FOR92 as

\[
L_h(t) = \left( \frac{C_{hh}(S, t)}{C_{hh}(S, t)} \right)_{S=0}.
\]

The \( t = 0 \) covariances in Figs. 4b, 4d, and 4f have turbulent microscales \( L_h(0) = 293 \), \( L_h(0) = 462 \), and \( L_h(0) = 891 \) km, respectively.

In Fig. 5, the ratios \( \tilde{E}_h^2(t)/E_h^2(0) \) (Fig. 5a) and \( L_h(t)/L_h(0) \) (Fig. 5b) are plotted as a function of time (abscissa) and initial perturbation turbulent microscale \( L_h(0) \) (ordinate). The value of \( L_h(0) \) at the bottom of each panel corresponds to the initial white-noise spectrum of Figs. 4a–b. For the white-noise initial difference spectrum, the difference variance doubles in about 18 h, while for a very red-noise spectrum it hardly increases at all. However, for a very red-noise spectrum, the characteristic scale actually decreases with time. Presumably, this is because the smallest scales have not yet saturated at 24 h and the difference variance is still growing there. For characteristic error scales of about 500 km, there is very little change in scale with time.

These experiments were repeated with different-amplitude initial perturbations \( E_h(0) = 1 \) m and \( E_h(0) = 30 \) m. There was essentially no change in the result, as would be expected. Also, the effect of model resolution was tested by integrating a lower-resolution model \( N = 20(T_{30}) \). The results, in the same format as Fig. 5, are shown in Fig. 6. In this case a white-noise initial spectra corresponds to \( L_h(0) = 430 \) km. It can be seen that there is essentially no change in error growth rates for any perturbations that can be adequately resolved by both models. The temporal change in scale, however, is somewhat more sensitive to the model resolution, with the time invariant characteristic scale shifting from 500 km in Fig. 5 to about 660 km in Fig. 6. This effect is obviously a function of the model resolution. The implications of these results for atmospheric data assimilation are discussed in the next section.

6. Summary and conclusions

Forecast-error statistics are absolutely vital for improving model performance and calculating analysis weights for atmospheric data assimilation. Forecast error is due to imperfections in the model and incomplete knowledge of the initial state. Traditionally, it has been difficult to disentangle these two sources of error. Kalman filter theory provides a mechanism for separating these error sources and examining them individually. It relies on the assumption that model error is additive noise that is white in time. This assumption has been used extensively in other disciplines, but its validity is unknown in atmospheric modeling. However, once this assumption has been made, it is possible to separate forecast error into a component due to imperfections in the model (model error) and a component that is due to the action of the model on the analysis error (predictability error).

The principal advantage of this formulation lies in the fact that the forecast-error statistics, which are usu-
ally inhomogeneous and nonstationary, are actually predicted. The principal drawback is that the model-error statistics are usually not known.

While, model-error statistics have not yet been estimated for any operational forecast model, forecast-error statistics, have been studied for some time. From forecast-error statistics alone, it is possible to approximately estimate the model-error covariance at large displacements (section 4). At small displacements, the model-error covariance can be easily estimated when the model is not very accurate.

Obtaining more satisfactory estimates of model-error statistics is not difficult in principle, although there are a number of practical details that must be considered. The idea is to separately estimate the homogeneous, stationary components of the forecast and predictability error and difference them to obtain the homogeneous, stationary component of the model error. The first step is to obtain forecast-error statistics over the domain sampled by a reliable radiosonde network using traditional innovation covariance techniques. The next step is to estimate the stationary, homogeneous component of the analysis-error covariance by using residual covariance techniques. Estimates of the analysis-error covariance require near-optimality for a really good estimate. The final step is to estimate the homogeneous, stationary component of the predictability-error covariance by a standard predictability experiment.

The first two steps of the method (estimation of the homogeneous, stationary components of the forecast- and analysis-error covariances) were satisfactorily tested with a simple linear advection model (section 4). The final step (determination of the predictability-error covariance) could only be examined meaningfully with a nonlinear model. Consequently, in section 5, a predictability experiment was performed with a nonlinear spherical conservation-of-absolute-vorticity model.

In this experiment, it was found that the redder the analysis-error spectrum, the more slowly the predictability error grew with time. The characteristic scale of the analysis error tended to increase over time for a very white initial error spectrum and decrease for a very red initial spectrum. For analysis-error spectra with characteristic scales of 400–600 km (which is probably close to what it is in operational practice), the growth was small and there was little change in characteristic scale. This result was consistent with the findings of Bloom and Schubert (1990).

The above results suggest that, as a first approximation, the homogeneous, stationary component of the 12-h predictability-error covariance is not markedly different from the homogeneous, stationary component of the analysis-error covariance.

It should be emphasized that the technique, while more complicated than existing schemes for estimating forecast-error covariances, could be implemented within existing data-assimilation schemes. The potential payoff is that the model-error covariance, determined in this way, could be used to more accurately determine forecast-error covariances in data-sparse regions. This would, in turn, lead to more optimal analysis weights and reduced analysis errors.

Acknowledgments. The author would like to thank Stephen Bloom, Herschel Mitchell, Richard Ménard, and Siegfried Schubert for their careful review of the manuscript. The diagrams were drawn by Tom Chivers of the Atmospheric Environment Service.

REFERENCES