

Suboptimal Schemes for Retrospective Data Assimilation Based on the Fixed-Lag Kalman Smoother

R. TODLING*

Universities Space Research Association, NASA/GSFC/DAO, Greenbelt, Maryland

S. E. COHN AND N. S. SIVAKUMARAN*

Data Assimilation Office, NASA/GSFC, Greenbelt, Maryland

(Manuscript received 5 March 1997, in final form 6 November 1997)

ABSTRACT

The fixed-lag Kalman smoother was proposed recently by S. E. Cohn et al. as a framework for providing retrospective data assimilation capability in atmospheric reanalysis projects. Retrospective data assimilation refers to the dynamically consistent incorporation of data observed well past each analysis time into each analysis. Like the Kalman filter, the fixed-lag Kalman smoother requires statistical information that is not available in practice and involves an excessive amount of computation if implemented by brute force, and must therefore be approximated sensibly to become feasible for operational use.

In this article the performance of suboptimal retrospective data assimilation systems (RDASs) based on a variety of approximations to the optimal fixed-lag Kalman smoother is evaluated. Since the fixed-lag Kalman smoother formulation employed in this work separates naturally into a (Kalman) filter portion and an optimal retrospective analysis portion, two suboptimal strategies are considered: (i) viable approximations to the Kalman filter portion coupled with the optimal retrospective analysis portion, and (ii) viable approximations to both portions. These two strategies are studied in the context of a linear dynamical model and observing system, since it is only under these circumstances that performance can be evaluated exactly. A shallow water model, linearized about an unstable basic flow, is used for this purpose.

Results indicate that retrospective data assimilation can be successful even when simple filtering schemes are used, such as one resembling current operational statistical analysis schemes. In this case, however, online adaptive tuning of the forecast error covariance matrix is necessary. The performance of this RDAS is similar to that of the Kalman filter itself. More sophisticated approximate filtering algorithms, such as ones employing singular values/vectors of the propagator or eigenvalues/vectors of the error covariances, as a way to account for error covariance propagation, lead to even better RDAS performance. Approximating both the filter and retrospective analysis portions of the RDAS is also shown to be an acceptable approach in some cases.

1. Introduction

The fixed-lag Kalman smoother (FLKS) has been proposed by Cohn et al. (1994; CST94 hereafter) as an approach to perform retrospective data assimilation. The term *retrospective data assimilation* denotes a procedure to incorporate data observed well *past* each analysis time into each analysis, taking into account error propagation through dynamical effects. Since a goal of *reanalysis* efforts is to produce a long archive of best-possible analyses based on *all* available data, whereas

current reanalysis projects (e.g., Gibson et al. 1997; Kalnay et al. 1996; Schubert and Rood 1995) incorporate only data observed up to and including each analysis time, retrospective data assimilation should be an ultimate goal of reanalysis efforts, as pointed out in CST94. Moreover, although retrospective data assimilation is studied in this article primarily as a means of improving analysis quality, it is foreseeable that such a procedure could also be adopted in numerical weather prediction to improve mid- to long-range forecasts, starting from a given retrospective analysis. The preliminary efforts of Gelaro et al. (1996) can be viewed as an approach to retrospective data assimilation with this purpose.

Cohn et al. (1994) gave a particular derivation of the optimal *linear* FLKS. In that work, it was pointed out that the same algorithm can be derived from the approach of "state enlargement," or "state augmentation" as it is more commonly known, first suggested in the engineering literature by Willman (1969), to reduce the smoothing problem to a filtering problem. In the state

* Current affiliation: General Sciences Corporation (a subsidiary of Science Applications International Corporation), Laurel, Maryland.

Corresponding author address: Dr. Ricardo Todling, Data Assimilation Office, NASA/GSFC, Code 910.3, Greenbelt, MD 20771.
E-mail: todling@dao.gsfc.nasa.gov

augmentation approach, the state vector at each time is appended with the state vector at previous times when the desired smoother estimates are to be calculated. A Kalman filter (KF) problem can then be solved for the augmented system. The first derivation of a smoother algorithm via state augmentation was that of Biswas and Mahalanabis (1972) for the linear fixed-point smoothing problem. Subsequently, Moore (1973) derived a linear fixed-lag smoother via the same approach, which results in the same algorithm as that derived in CST94. Extension of the FLKS formulation to nonlinear systems can be achieved using the same technique of state augmentation, as indicated by Biswas and Mahalanabis (1973), for both the fixed-point and fixed-lag smoothing problems [see also Todling and Cohn (1996a) for an explicit derivation of the extended fixed-lag Kalman smoother]. The utility of state augmentation is that the resulting smoothers are often computationally less demanding than those arising from some other approaches (e.g., Sage and Melsa 1970, section 9.5). For instance, smoothers based on state augmentation avoid inversion of the filter error covariance matrices and of the tangent linear propagator (e.g., Ménard and Daley 1996; see also the appendix of the present article). These inversions are also avoided by an earlier smoother algorithm due to Bryson and Frazier (1963), which can be shown to reduce to the FLKS algorithm of CST94 for the case of linear systems. Algebraic equivalence between smoothers obtained by state augmentation and by methods such as maximum likelihood (Sage and Ewing 1970; Sage 1970) or conditional expectation (Leondes et al. 1970) exists in most cases. The interested reader is referred to Meditch (1973) and Kailath (1975) for detailed reviews of the literature on linear and nonlinear smoothing. The distinction among different types of smoothing problems, and the connection between fixed-interval smoothing and four-dimensional variational (4DVAR) analysis, is drawn in Cohn (1997).

Brute-force implementation of the (extended) FLKS to build an operational retrospective data assimilation system (RDAS) is not possible for the same reasons that a brute-force (extended) KF-based data assimilation system would be impractical: computational requirements are excessive, and knowledge of the requisite error statistics is lacking. Therefore, approximations not only must be employed but cannot be escaped from. Thus, in this article, we develop and evaluate the performance of potentially implementable approximate schemes. To provide an exact evaluation, we choose a barotropically unstable linear shallow water model as a test bed for this investigation. All of the approximate schemes evaluated here have relatively simple nonlinear equivalents.

In the sequel, we first briefly review, in section 2, the linear FLKS of CST94 and outline the performance evaluation technique employed to study the behavior of linear suboptimal filter and smoother algorithms. Section 3 gives a summary of the suboptimal filters and smoothers evaluated subsequently in section 4, in the

context of the linear shallow water model. We draw conclusions in section 5.

2. Review and performance evaluation equations

Before we summarize the linear FLKS algorithm and performance evaluation equations, let us recall that the linear Kalman filter equations, in the notation of CST94, are

$$\mathbf{w}_{k|k-1}^f = \mathbf{A}_{k,k-1} \mathbf{w}_{k-1|k-1}^a, \quad (1a)$$

$$\mathbf{P}_{k|k-1}^f = \mathbf{A}_{k,k-1} \mathbf{P}_{k-1|k-1}^a \mathbf{A}_{k,k-1}^T + \mathbf{Q}_k, \quad (1b)$$

$$\mathbf{K}_{k|k} = \mathbf{P}_{k|k-1}^f \mathbf{H}_k^T \Gamma_k^{-1}, \quad (1c)$$

$$\mathbf{w}_{k|k}^a = \mathbf{w}_{k|k-1}^f + \mathbf{K}_{k|k} (\mathbf{w}_k^o - \mathbf{H}_k \mathbf{w}_{k|k-1}^f), \quad \text{and} \quad (1d)$$

$$\mathbf{P}_{k|k}^a = (\mathbf{I} - \mathbf{K}_{k|k} \mathbf{H}_k) \mathbf{P}_{k|k-1}^f. \quad (1e)$$

Here (1a) is the expression for the forecast n vector $\mathbf{w}_{k|k-1}^f$, obtained through evolution of the analysis n vector $\mathbf{w}_{k-1|k-1}^a$ between two consecutive analysis times t_{k-1} and t_k via the propagator $\mathbf{A}_{k,k-1}$; (1b) is the corresponding expression for the $n \times n$ forecast error covariance matrix $\mathbf{P}_{k|k-1}^f$, where \mathbf{Q}_k is the $n \times n$ model error covariance matrix. The state estimate $\mathbf{w}_{k|k-1}^f$ is updated using (1d), as p_k observations \mathbf{w}_k^o become available at each time t_k : the difference between the observations and their predicted values $\mathbf{H}_k \mathbf{w}_{k|k-1}^f$, expressed via the $p_k \times n$ observation matrix \mathbf{H}_k , is added to the forecast after being weighted by the $n \times p_k$ Kalman gain matrix $\mathbf{K}_{k|k}$. At each observation time, the gain matrix is computed according to (1c), where

$$\Gamma_k \equiv \mathbf{H}_k \mathbf{P}_{k|k-1}^f \mathbf{H}_k^T + \mathbf{R}_k \quad (2)$$

is the $p_k \times p_k$ innovation covariance matrix and \mathbf{R}_k is the $p_k \times p_k$ observation error covariance matrix. The resulting analysis error covariance matrix $\mathbf{P}_{k|k}^a$ is given by (1e), which completes the Kalman filter cycle.

The subscript notation utilized here is common in estimation theory, and is particularly important when considering smoothing problems. Specifically, the forecast vector $\mathbf{w}_{k|k-1}^f$ is the conditional mean of the true state at time t_k , where the conditioning is on all observations up to and including those at time t_{k-1} , hence the double time subscript. Similarly, the analysis vector $\mathbf{w}_{k|k}^a$ is the conditional mean of the true state at time t_k conditioned on data up to and including time t_k . Analogously, the forecast and analysis error covariance matrices carry a second time subscript to indicate the set of observations upon which they are conditioned. A more comprehensive explanation of the Kalman filter equations, including the probabilistic assumptions from which they are derived, can be found elsewhere (e.g., Jazwinski 1970; Gelb 1974; Cohn 1997).

The linear fixed-lag Kalman smoother algorithm of CST94 consists of the Kalman filter equations (1)–(2) along with a set of equations appended to those of the Kalman filter. We refer to the appended equations as the

retrospective analysis portion of the FLKS. An improved state estimate, referred to as the retrospective analysis, at some past time t_{k-l} , say, can be obtained if we process data beyond time t_{k-l} , $l \geq 1$, up to the current time t_k . This estimate, denoted by $\mathbf{w}_{k-l|k}^a$, is the conditional mean of the true state at time t_{k-l} , where the conditioning is on all observations up to and including time t_k . It can be calculated according to

$$\mathbf{w}_{k-l|k}^a = \mathbf{w}_{k-l|k-1}^a + \mathbf{K}_{k-l|k}(\mathbf{w}_k^o - \mathbf{H}_k \mathbf{w}_{k|k-1}^f), \quad (3)$$

where $\mathbf{K}_{k-l|k}$ is the corresponding retrospective analysis gain matrix. Comparing this expression with the usual filter analysis expression (1d), we see that the retrospective analysis update is based on the same innovation vector ($\mathbf{w}_k^o - \mathbf{H}_k \mathbf{w}_{k|k-1}^f$) at time t_k as that of the filter, and it represents a further correction to a previously computed (retrospective) analysis $\mathbf{w}_{k-l|k-1}^a$; notice the contrast to the filter analysis expression, which represents a correction to the forecast $\mathbf{w}_{k|k-1}^f$.

The FLKS update equation (3) is only applicable for $l \leq k$. If the (fixed) total number of lags is L , meaning that (3) is to be applied in general for $l = 1, 2, \dots, L$, then for $k = 0, 1, \dots, L - 1$, the condition $l \leq k$ is not satisfied for all l . Therefore, the update (3) is applied only for $l = 1, 2, \dots, \min(k, L)$, which is a restricted range of l when $k = 0, 1, \dots, L - 1$. In the language of estimation theory, this restriction corresponds to computing *fixed-point* smoother results for all k up to $k = L - 1$, after which the fixed-lag smoother starts operating. This is an initialization procedure for the fixed-lag smoother (e.g., Gelb 1974, 173–176). In practice, because the FLKS algorithm employed by CST94 already has the structure of a fixed-point algorithm, this procedure simply amounts to controlling the ending points of certain loop statements in a computer code.

In the optimal FLKS algorithm of CST94, the $n \times p_k$ retrospective analysis gain matrix $\mathbf{K}_{k-l|k}$ is given by

$$\mathbf{K}_{k-l|k} = (\mathbf{P}_{k,k-l|k-1}^{fa})^T \mathbf{H}_k^T \mathbf{\Gamma}_k^{-1}, \quad (4)$$

where the innovation covariance matrix $\mathbf{\Gamma}_k$ is the same as that used to calculate the filter gain matrix $\mathbf{K}_{k|k}$ in (1c), since the retrospective analysis update (3) of the FLKS is based on the same innovation vector as KF. The $n \times n$ matrix $\mathbf{P}_{k,k-l|k-1}^{fa}$ is the forecast–retrospective analysis error cross-covariance matrix, and evolves according to the following set of equations:

$$\mathbf{P}_{k,k-l|k-1}^{fa} = \mathbf{A}_{k,k-1} \mathbf{P}_{k-1,k-l|k-1}^{aa}, \quad (5a)$$

$$\mathbf{P}_{k,k-l|k}^{aa} = (\mathbf{I} - \mathbf{K}_{k|k} \mathbf{H}_k) \mathbf{P}_{k,k-l|k-1}^{fa}, \quad \text{and} \quad (5b)$$

$$\mathbf{P}_{k-l|k}^a = \mathbf{P}_{k-l|k-1}^a - \mathbf{K}_{k-l|k} \mathbf{H}_k \mathbf{P}_{k,k-l|k-1}^{fa}. \quad (5c)$$

Here the $n \times n$ matrix $\mathbf{P}_{k-1,k-l|k-1}^{aa}$ is the filter analysis–retrospective analysis error cross-covariance matrix, and the $n \times n$ matrix $\mathbf{P}_{k-l|k}^a$ is the retrospective analysis error covariance matrix. Equations (1)–(5) complete the description of the FLKS algorithm of CST94, with (1)–

(2) giving the filter portion and (3)–(5) giving the retrospective analysis portion. There are a number of advantages to this FLKS algorithm. In particular, model error is incorporated implicitly in the retrospective analysis portion: no model error terms appear in (3)–(5). This point is clarified in the appendix, where the algebraic equivalence of this algorithm with a more well-known alternative formulation is exploited.

As stated in the introduction, the optimal FLKS algorithm described above is not practical for operational implementation of RDASs, due in part to its computational requirements. As a matter of fact, most of the computation arises in the filter portion. Consequently, as only approximate filters are feasible in practice, the resulting RDAS will be suboptimal. In this article, we investigate closely the effects of approximate schemes for either, or both, the filter and the retrospective analysis portions. The schemes we consider approximate *only* the filter and retrospective analysis gains (1c) and (4), respectively, including the innovation covariance (2) on which they depend, by replacing them with gains $\tilde{\mathbf{K}}_{k|k}$ and $\tilde{\mathbf{K}}_{k-l|k}$ identical in form to (1c) and (4) but involving approximate expressions for $\mathbf{P}_{k|k-1}^f$ and $\mathbf{P}_{k,k-l|k-1}^{fa}$. Thus we will be studying approximate expressions for the propagated (predictability) error covariance matrix

$$\mathbf{P}_{k|k-1}^p \equiv \mathbf{A}_{k,k-1} \mathbf{P}_{k-1|k-1}^a \mathbf{A}_{k,k-1}^T, \quad (6)$$

and for the forecast–retrospective analysis error cross-covariance matrix

$$\mathbf{P}_{k,k-l|k-1}^{fa} = \mathbf{A}_{k,k-1} \mathbf{P}_{k-1,k-l|k-1}^{aa}. \quad (7)$$

Calculating the exact expressions (6) and (7) is the most computationally demanding part of the optimal smoother algorithm (cf. Todling 1995). To focus on the issue of performance due to approximating gains by approximating (6) and (7), we make the perfect model assumption, $\mathbf{Q}_k = \mathbf{0}$, in which case the terms predictability error covariance matrix and forecast error covariance matrix are interchangeable: $\mathbf{P}_{k|k-1}^p = \mathbf{P}_{k|k-1}^f$.

For linear systems, filter performance evaluation can be accomplished following the procedure of Todling and Cohn (1994), and here is extended to incorporate the retrospective analysis performance evaluation equations as well. These equations can be obtained from the derivation of the FLKS of CST94 [cf. Eqs. (2.33), (2.39), and (2.45) in CST94], and are valid for general (filter and retrospective analysis) gain matrices $\mathbf{K}_{k|k}$ and $\mathbf{K}_{k-l|k}$:

$$\mathbf{P}_{k|k}^a = (\mathbf{I} - \tilde{\mathbf{K}}_{k|k} \mathbf{H}_k) \mathbf{P}_{k|k-1}^f (\mathbf{I} - \tilde{\mathbf{K}}_{k|k} \mathbf{H}_k)^T + \tilde{\mathbf{K}}_{k|k} \mathbf{R}_k \tilde{\mathbf{K}}_{k|k}^T, \quad (8a)$$

$$\begin{aligned} \mathbf{P}_{k-l|k}^a &= \mathbf{P}_{k-l|k-1}^a + \tilde{\mathbf{K}}_{k-l|k} \mathbf{\Gamma}_k \tilde{\mathbf{K}}_{k-l|k}^T - \tilde{\mathbf{K}}_{k-l|k} \mathbf{H}_k \mathbf{P}_{k,k-l|k-1}^{fa} \\ &\quad - (\mathbf{H}_k \mathbf{P}_{k,k-l|k-1}^{fa})^T \tilde{\mathbf{K}}_{k-l|k}^T, \end{aligned} \quad (8b)$$

and

$$\begin{aligned} \mathbf{P}_{k,k-l|k}^{aa} &= \mathbf{P}_{k,k-l|k-1}^{fa} + \check{\mathbf{K}}_{k|k} \mathbf{\Gamma}_k \check{\mathbf{K}}_{k-l|k}^T - \check{\mathbf{K}}_{k|k} \mathbf{H}_k \mathbf{P}_{k,k-l|k-1}^{fa} \\ &\quad - \mathbf{P}_{k|k-1}^f \mathbf{H}_k^T \check{\mathbf{K}}_{k-l|k}^T. \end{aligned} \quad (8c)$$

Together with (6) and (7) calculated exactly, these equations give the update and evolution of all of the *actual* filter and retrospective analysis error covariances. Expression (8a) is the well-known Joseph formula, and gives the performance of the filter analysis for a general gain matrix $\check{\mathbf{K}}_{k|k}$, and (8b) gives the performance of the retrospective analyses for general gains $\check{\mathbf{K}}_{k-l|k}$. Notice that the performance evaluation equations [(6), (8a)] for the filter are independent of those [(7), (8b), (8c)] for the retrospective analysis, whereas the converse is not true. This is simply a consequence of the fact that the optimal linear filter is independent of the optimal linear retrospective analysis. This independence does not carry over to some nonlinear extensions, for example, to the globally iterated smoother (Jazwinski 1970, 280–281).

3. Summary of suboptimal filters and smoothers

We now summarize the suboptimal schemes to be evaluated here in the context of the linear shallow water model of the next section. The following are the suboptimal schemes considered in this article for the *filter* portion of the fixed-lag smoother algorithm (see Cohn and Todling 1996, CT96 hereafter; Todling et al. 1996; Todling and Cohn 1996a,b).

a. Constant forecast error covariance filter (CCF)

Here the predictability error covariance matrix $\mathbf{P}_{k|k-1}^p$ given by (6) is replaced in the filter gain expression (1c) by

$$\mathbf{S}_{k|k-1}^p = \alpha_k \bar{\mathbf{S}}, \quad (9)$$

where the parameter α_k is tuned adaptively following the algorithm of Dee (1995), and $\bar{\mathbf{S}}$ is a prescribed time-independent error covariance matrix. This scheme resembles current operational global analysis schemes. In the experiments of the following section the structure of $\bar{\mathbf{S}}$ is given by a weighted outer product of the slow eigenmodes of the governing dynamics over one time step.

b. Partial singular-value decomposition filter (PSF)

In the PSF (see CT96 for details), the dynamical operator $\mathbf{A}_{k,k-1}$ is approximated by the leading part of its singular value decomposition, here abbreviated by $\tilde{\mathbf{A}}_{k,k-1}$, and the predictability error covariance matrix is simplified for use in (1c) as

$$\mathbf{S}_{k|k-1}^p = \tilde{\mathbf{A}}_{k,k-1} \mathbf{S}_{k-1|k-1}^a \tilde{\mathbf{A}}_{k,k-1}^T + \mathbf{T}_{k|k-1}, \quad (10)$$

where the matrix $\mathbf{T}_{k|k-1}$ is an estimate of the trailing error covariance matrix due to the replacement of the dynamics by its leading part.

c. Partial eigendecomposition filter (PEF)

In the PEF (see CT96 for details), the entire predictability error covariance matrix is replaced for use in (1c) by the leading part of its eigendecomposition, which ideally explains most of the predictability error variance; that is,

$$\mathbf{S}_{k|k-1}^p = (\mathbf{W}_N \hat{\mathbf{S}}_N \mathbf{W}_N^T)_{k|k-1} + \mathbf{T}'_{k|k-1}, \quad (11)$$

where $\mathbf{W}_{N;k|k-1}$ is the matrix of the N dominant eigenvectors, with the corresponding N largest eigenvalues arranged along the diagonal of the diagonal matrix $\hat{\mathbf{S}}_{N;k|k-1}$, and $\mathbf{T}'_{k|k-1}$ is an estimate of the trailing error covariance matrix of this approximation, in general distinct from $\mathbf{T}_{k|k-1}$. This approach resembles the reduced-rank square-root filter of Verlaan and Heemink (1995).

d. Reduced resolution filter (RRF)

This approximation follows the approach of Fukumori and Malanotte-Rizzoli (1995; see also Cane et al. 1996; Todling and Cohn 1996b) and involves carrying the error covariances at lower resolution than that of the state estimates. In this case, the predictability error covariance matrix is approximated for use in (1c) by

$$\tilde{\mathbf{S}}_{k|k-1}^p = (\mathbf{B}^+ \mathbf{A}_{k,k-1} \mathbf{B}) \tilde{\mathbf{S}}_{k-1|k-1}^a (\mathbf{B}^+ \mathbf{A}_{k,k-1} \mathbf{B})^T + \mathbf{T}''_{k|k-1}, \quad (12)$$

where $\mathbf{T}''_{k|k-1}$ stands for an estimate of the trailing error covariance matrix accounting for neglected structures due to the approximation; \mathbf{B} is an $n \times m$ matrix representing an interpolation operator that takes vectors from the m -dimensional reduced space where the error covariance matrices $\tilde{\mathbf{S}}_{k-1|k-1}^a$ and $\tilde{\mathbf{S}}_{k|k-1}^p$ are represented to the n -dimensional space of the state estimates; the matrix \mathbf{B}^+ represents an $m \times n$ pseudoinverse of the interpolation operator \mathbf{B} , which in our experiments is taken to be the Moore–Penrose pseudoinverse (e.g., Campbell and Meyer 1991). The matrix \mathbf{B} has columns corresponding to coefficients of a two-dimensional cubic spline interpolation—a spline interpolant with periodic boundary conditions in the zonal direction and an Akima spline in the meridional direction.

It should be pointed out that the approach of reduced resolution filtering is very general, falling in the broad category of order reduction schemes commonly known in estimation theory. In this regard, the PSF scheme described above can be seen as a reduced-order approximation with the matrix \mathbf{B} chosen appropriately. This analogy renders computational costs for the PSF and the RRF comparable. Moreover, the PEF does not represent any more computational savings than either of these two schemes, as it requires an iterative eigensolver just as the PSF does. Finally, the similarity between the CCF scheme and operational schemes, where no dynamics is used in modeling the predictability error covariance, makes the CCF represent at least an order of magnitude savings over the other filtering schemes presented here.

The following are the suboptimal schemes considered here for the *retrospective analysis* portion of the fixed-lag smoother algorithm.

e. Partial singular-value decomposition retrospective analysis (PSRA/PSRA2)

In this category, there are at least two possibilities. The PSRA scheme extends the PSF approximation of the filter gain (1c) to the retrospective analysis gains (4), whereas the PSRA2 scheme extends the PEF approximation to the retrospective analysis gains.

Thus the PSRA scheme approximates the forecast–retrospective analysis error cross-covariance matrix given in (7), for use in (4) by

$$\mathbf{S}_{k,k-l|k-1}^{fa} = \tilde{\mathbf{A}}_{k,k-1} \mathbf{S}_{k-1,k-l|k-1}^{aa} + \mathbf{X}_{k,k-l|k-1}, \quad (13)$$

where $\mathbf{X}_{k,k-l|k-1}$ is a trailing error cross-covariance matrix. Notice that, in principle, the number of singular modes included in $\tilde{\mathbf{A}}_{k,k-1}$ here does not have to be the same as in the PSF. However, in the experiments discussed below the same singular modes are retained in both cases. Also, in the experiments reported here we take $\mathbf{X}_{k,k-l|k-1} = \mathbf{0}$, at all times t_k .

The PSRA2 scheme approximates (7) by performing a partial singular value decomposition of the complete forecast–retrospective analysis error cross-covariance matrix. That is,

$$\mathbf{S}_{k,k-l|k-1}^{fa} = (\mathbf{U}_N \mathbf{D}_N^a \mathbf{V}_N^T)_{k,k-l|k-1} + \mathbf{X}'_{k,k-l|k-1}, \quad (14)$$

where the columns of the $n \times N$ matrix \mathbf{U}_N and the rows of the $N \times n$ matrix \mathbf{V}_N^T contain the N leading left and right singular vectors of the propagated analysis–retrospective analysis error cross-covariance matrix, and the $N \times N$ diagonal matrix \mathbf{D}_N^a contains the N leading singular values. It is important to recognize that the main difference between this scheme and the PSRA scheme in (13) is that in (14) the complete dynamics operator $\mathbf{A}_{k,k-1}$ is used. As an example of how this procedure is implemented, consider the case $l = 1$, and assume that an estimate $\mathbf{S}_{k-1|k-1}^a$ of the analysis error covariance is available. Then from (7), and the ideas of Lanczos-type algorithms, the PSRA2 approximation for $\mathbf{S}_{k,k-1|k-1}^{fa}$ reduces to successive application of the matrices $\mathbf{A}_{k,k-1}$ and $\mathbf{S}_{k-1|k-1}^a$ to a “general” vector \mathbf{u} ; that is,

$$\mathbf{S}_{k,k-1|k-1}^{fa} \mathbf{u} = \mathbf{A}_{k,k-1} \mathbf{S}_{k-1|k-1}^a \mathbf{u}, \quad (15)$$

as many times as required by the convergence criterion and the desired number N of leading singular vectors. As before, the matrix $\mathbf{X}'_{k,k-l|k-1}$ represents the trailing error cross-covariance matrix, which in the experiments discussed in the sequel is neglected.

f. Reduced resolution retrospective analysis (RRRA)

In the RRRA scheme, by analogy with the RRF approximation, we compute the forecast–retrospective

analysis error cross-covariance matrix at reduced resolution:

$$\tilde{\mathbf{S}}_{k,k-l|k-1}^{fa} = (\mathbf{B}^+ \mathbf{A}_{k,k-1} \mathbf{B}) \tilde{\mathbf{S}}_{k-1,k-l|k-1}^{aa} + \mathbf{X}''_{k,k-l|k-1}, \quad (16)$$

where the matrices \mathbf{B} and \mathbf{B}^+ are the interpolation matrices as introduced before, the matrices $\tilde{\mathbf{S}}_{k-1,k-l|k-1}^{aa}$ and $\tilde{\mathbf{S}}_{k,k-l|k-1}^{fa}$ are $m \times m$ error cross-covariance matrices in the reduced space, and the matrix $\mathbf{X}''_{k,k-l|k-1}$ stands for a trailing error cross-covariance matrix. The matrices \mathbf{B} and \mathbf{B}^+ here do not have to be identical to those used in the RRF; however, in the experiments discussed below they are chosen to be so. Also, in the experiments reported here, we take $\mathbf{X}''_{k,k-l|k-1} = \mathbf{0}$, at all times t_k .

Many other suboptimal schemes have been proposed for filtering, particularly in the atmospheric data assimilation literature (Todling and Cohn 1994, and references therein). Since fixed-lag smoothing can always be regarded as filtering for an augmented-state system (e.g., Todling and Cohn 1996a), in principle all of these suboptimal strategies carry over to the fixed-lag smoothing problem. In this article we choose to concentrate only on the approximations presented above.

It is possible to construct approximate RDASs by combining different strategies for approximating the filter and the retrospective analysis portions of the RDAS. For instance, one could choose to approximate both portions equivalently—that is, with two similar schemes like the RRF and RRRA at the same resolution; or one could choose to approximate the filter and calculate the retrospective analysis portion exactly—that is, to approximate (6) and use (7); one could also build hybrid approximations in which the filter and the retrospective analysis employ different strategies. In any case, since our formulation of the fixed-lag smoother is based on the filter, whenever the filter is approximated the smoother becomes suboptimal. The converse is not true, in the sense that if the filter is kept exact and the retrospective analysis equations are approximated—if we use (6) and approximate (7)—only the smoother becomes suboptimal, but not the filter. This, however, may not be a very useful approach, since the major computational requirements are associated with the filter equation (6).

4. Results for a shallow water model

To evaluate the performance of the suboptimal schemes described above, we use the barotropically *unstable* model of CT96, a shallow water model linearized about a meridionally dependent squared-hyperbolic secant jet (Bickley jet; Haltiner and Williams 1980, 175). We refer the reader to Fig. 1 of CT96 for the shape, extent, and strength of the jet. The model domain is shown in Fig. 1 here. The assimilation experiments employ the observing network of CT96: 33 radiosonde stations observing winds and heights every 12 h and distributed outside the strongest part of the jet. The tick marks in the figure indicate the 25×16 model grid. In

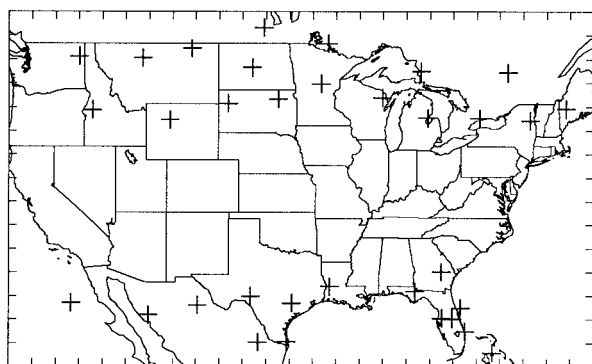


FIG. 1. Model domain and observational network composed of 33 radiosonde stations observing winds and heights every 12 h (same as Fig. 2 of CT96).

the experiments referring to a trailing error covariance matrix we construct it, exactly as in CT96, using the slow eigenmodes of the autonomous unstable dynamics of our shallow water model.

Before evaluating the performance of a few suboptimal RDASs, we discuss results obtained for the *optimal* FLKS. The performance of the optimal filter and fixed-lag smoother can be seen in Fig. 2, which shows the domain-averaged expected root-mean-square (erms) analysis error in the total energy as a function of time. This quantity is calculated as a weighted average of the analysis error variances of the three variables of the model, which are extracted from the main diagonal of the *actual* analysis error covariance matrix. The top curve in the figure corresponds to the filter analysis every 12 h, whereas successive retrospective analysis results are given by successively lower curves, which refer to analyses including data 12, 24, 36, and 48 h ahead in time—that is, lags $l = 1, 2, 3,$ and 4 . The filter curve is the same as that in Fig. 2 of CT96 (shown, here, only up to 5 days). The most relevant results are those for the transient part of the assimilation period, before the filter and smoother begin to approach steady state. Incorporating new data into past analyses reduces the corresponding past analysis errors considerably. The largest impact is on the initial analysis, which would not be the case if a significant amount of model error were present.

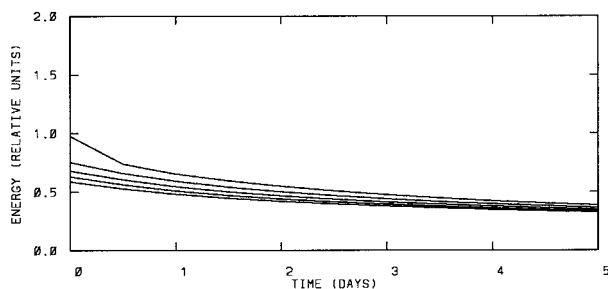


FIG. 2. The erms analysis error in total energy for the Kalman filter (upper curve) and fixed-lag Kalman smoother (lower curves).

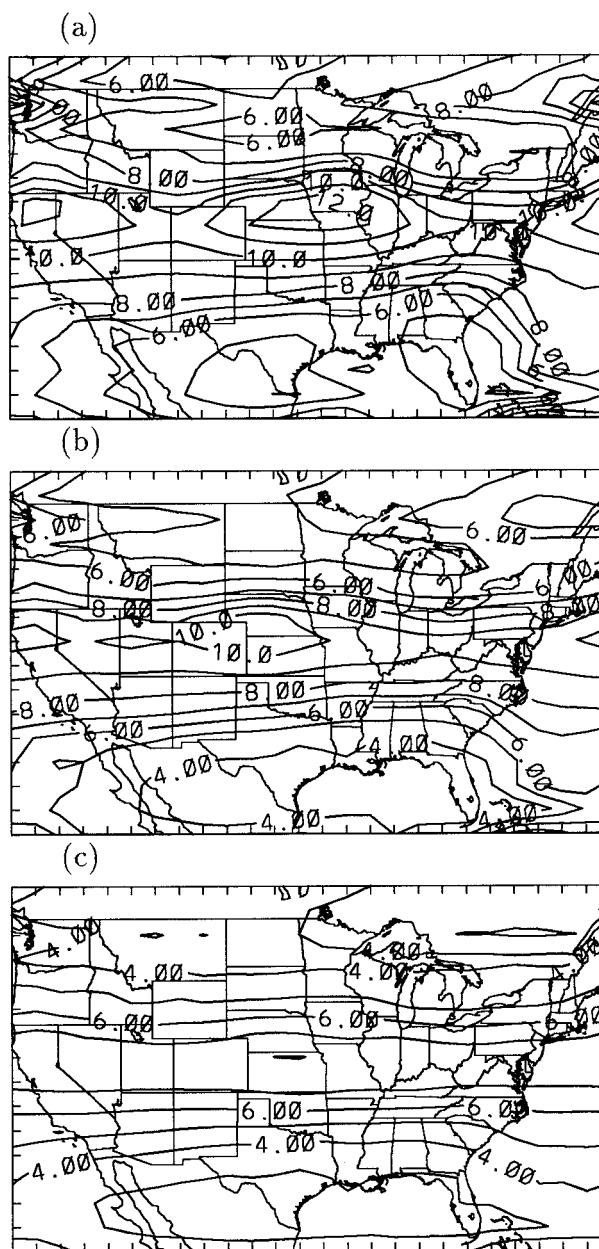


FIG. 3. Analysis error standard deviation in the height field at time $t = 0.5$ days. (a) The filter analysis; (b) and (c) the retrospective analyses with lags $l = 1$ and 4 , respectively. Contour interval is 1 m.

Further illustration of the behavior of the *optimal* FLKS is given in Fig. 3, where we display maps of the analysis error standard deviation in the height field at $t = 0.5$ days. The panels are for the filter analysis errors (Fig. 3a), and for the retrospective analysis errors for lags $l = 1$ (Figs. 3b) and $l = 4$ (Fig. 3c). Thus, in Figs. 3b,c and the analysis errors are reduced by incorporating data 12 and 48 h ahead of the current analysis time ($t = 0.5$ days), respectively. We see not only the overall decrease in error levels from Fig. 3a to Fig. 3c, as expected from Fig. 2, but also that within each panel errors

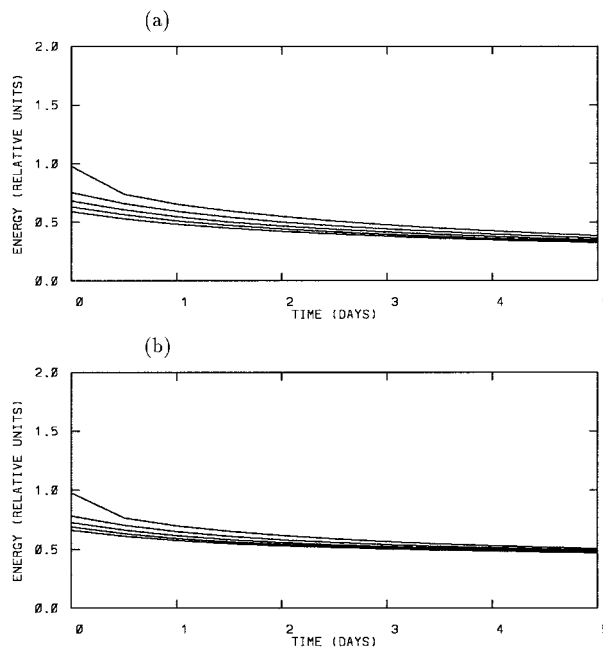


FIG. 4. As in Fig. 2 but for an approximate RDAS using the RRF and RRR schemes for resolutions: (a) 13×16 and (b) 13×12 .

are largest in the central band of the domain, where there are no observations and where the jet is strongest. Furthermore, the error maximum in the center of the domain moves westward and diminishes as more data are incorporated into the analysis through the smoothing procedure (from Figs. 3a to 3c). This property of the FLKS of propagating and reducing errors in the direction opposite of the flow has already been observed in the experiments of CST94 and Ménard and Daley (1996).

We now study the behavior of *suboptimal* RDASs. We start with schemes that approximate the filter and retrospective analysis portions similarly. In this category, we investigate the behavior of the RRF and RRR corresponding to expressions (12) and (16), respectively, as well as the behavior of the PSF and PSRA corresponding to expressions (10) and (13), respectively.

The results of Todling and Cohn (1996b) showed that the RRF described above, with resolutions 13×16 and 13×12 , provides good filter performance in our shallow water model context. This was attributed mainly to the fact that at these resolutions the barotropically unstable jet is fairly well resolved. As a matter of fact, the meridional jet is fully resolved at resolution 13×16 . In Fig. 4 we show results of the performance evaluation for the RRF and RRR algorithms at these resolutions (Fig. 4a for 13×16 ; Fig. 4b for 13×12). No trailing error covariance models are taken into account here, that is, $\mathbf{T}_{k|k-1}'' = \mathbf{0}$ and $\mathbf{X}_{k,k-l|k-1}'' = \mathbf{0}$ for the RRF and RRR, respectively. As in Fig. 2, the upper curve in each panel is for the performance of the filter analysis, and the lower curves in each panel are for the perfor-

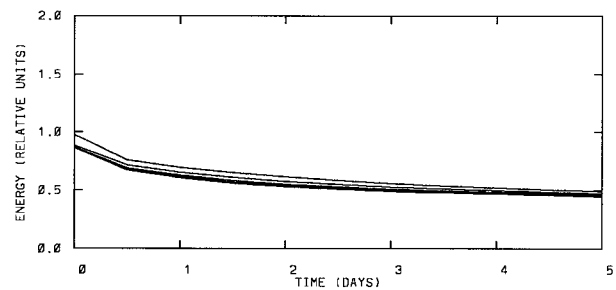


FIG. 5. As in Fig. 2 but for an approximate RDAS using the PSF and PSRA schemes simultaneously, both with 54 modes.

mance of the successive retrospective analyses. Comparison of Fig. 4a with the optimal FLKS results of Fig. 2 shows remarkable agreement between the filter and retrospective analysis results when the jet is fully resolved. The agreement with the coarse meridional resolution result in Fig. 4b is still quite good, especially during the transient part of the assimilation. Asymptotically, the analysis error levels for the case of 13×12 resolution are somewhat higher than those at 13×16 resolution.

The results of the RRF can be improved by modeling the trailing error covariance matrix $\mathbf{T}_{k|k-1}''$ to account for scales that are neglected in the RRF approximation. The impact of modeling this covariance matrix is greatest in experiments where the jet is least well resolved by the RRF scheme (results not shown). In these cases, the retrospective analysis results also improve due to improvement of the filter results. We have not investigated the impact of modeling the RRR trailing error cross-covariance.

Along similar lines, we investigate the performance of an RDAS using the PSF algorithm for the filter portion, and the PSRA algorithm for the retrospective analysis portion. From the experiments of CT96, we know that using the first 54 singular modes of the 12-h propagator of the linear shallow water model—those with singular values greater than or equal to one—is enough to produce a stable suboptimal filter. Moreover, we learned in CT96 that adaptively tuning a modeled trailing error covariance matrix $\mathbf{T}_{k|k-1}$ improves the filter results; we use the same procedure here. However, we do not model the trailing error cross-covariance matrix for the retrospective analysis portion, that is, we take $\mathbf{X}_{k,k-l|k-1} = \mathbf{0}$ at all times.

Figure 5 shows performance results for the PSF–PSRA suboptimal RDAS when the first 54 modes are used for both approximations (out of a total of 325 slow modes). The filter results, when compared to the optimal results of Fig. 2, are once again quite good—the reader is encouraged to compare the top curve of Fig. 5 with the curve labeled S54 in Fig. 11 of CT96; results now are better due to the adaptively tuned trailing error covariance matrix. The PSRA results, on the other hand, are not nearly as good as those for the optimal smoother

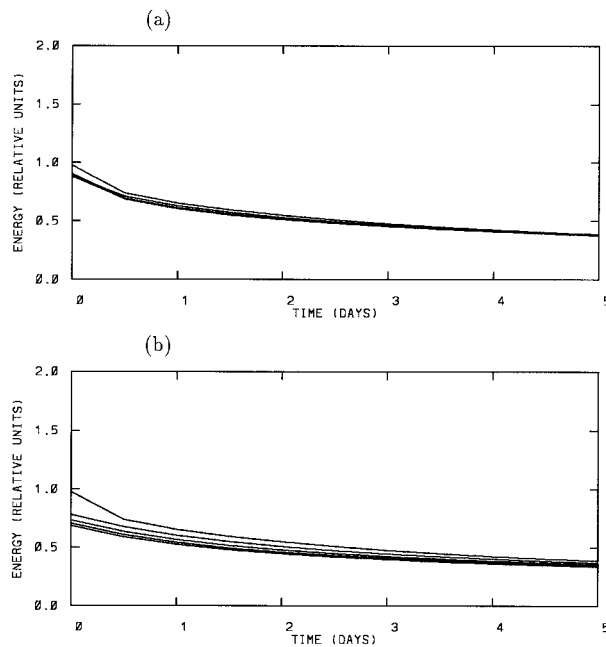


FIG. 6. As in Fig. 2 but for an RDAS using the KF and approximate retrospective analysis schemes: (a) PSRA and (b) PSRA2, with singular 54 modes.

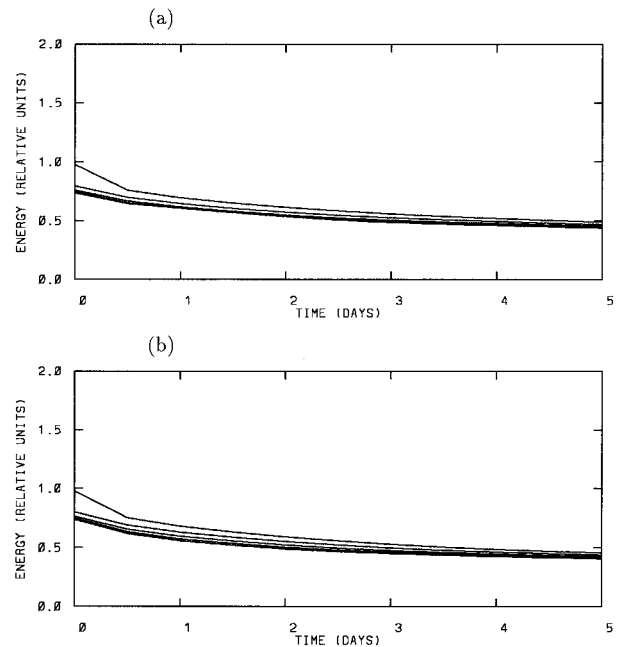


FIG. 7. As in Fig. 2 but for an RDAS using the PSRA2 scheme for the retrospective analysis portion, and either the (a) PSF or (b) PEF for the filter portion, all with 54 modes.

(Fig. 2), with little difference among results for lag $l = 1$ and those for higher lags $l = 2, 3,$ and 4 in Fig. 5. The next two experiments demonstrate that this poor smoother performance can be attributed mostly to neglecting the trailing forecast–retrospective analysis error cross-covariance matrix $\mathbf{X}_{k,k-l|k-1}$ in the PSRA algorithm. A further experiment later in this section, where the PSF scheme is combined with the exact retrospective analysis algorithm, also shows much better smoother performance than that seen in Fig. 5.

To investigate the PSRA scheme further, we compare performance results between two RDASs using the KF for the filter portion, with the retrospective analysis portion given by either the PSRA scheme or the PSRA2 scheme, both with 54 singular modes retained. Thus, the suboptimality in these two RDASs is solely in the retrospective analysis portion. Figure 6 shows the erms errors in the total energy for these two cases: Fig. 6a corresponds to the RDAS using the PSRA scheme and Fig. 6b corresponds to the RDAS using the PSRA2 algorithm. The filter curves in both panels (topmost curves) are identical to one another, as well as to the filter curve in Fig. 2 for the optimal FLKS case. Comparison between the lower curves in the two panels of Fig. 6 shows the superiority of the PSRA2 scheme: beyond lag $l = 1$ little is gained in the PSRA scheme, even when based on the KF (cf. Fig. 6a with Fig. 5), whereas successively higher lags do have a significant impact in the PSRA2 scheme (Fig. 6b). The poor performance of the PSRA scheme indicates that its neglected trailing part contains a large amount of cross-

(co)variance when retaining just the 54 singular modes of the propagator with singular values larger than or equal to one. The PSRA2 scheme with 54 modes, on the other hand, captures most of the cross-(co)variance, as comparison of Fig. 6b with the optimal result in Fig. 2 indicates. We conclude from these experiments that the trailing cross-covariance matrix is more significant in some approximate retrospective analysis schemes than in others. Moreover, the singular modes of the propagated filter analysis–retrospective analysis error cross-covariance matrix (employed in the PSRA2 scheme), rather than the singular modes of the propagator itself (employed in the PSRA scheme), contain most of the information relevant for retrospective analysis. This distinction between the PSRA and PSRA2 schemes is completely analogous to the distinction between the PSF and PEF schemes drawn in CT96, where somewhat better performance of the PEF scheme over the PSF scheme was demonstrated. This distinction is even more pronounced in the retrospective analysis context, as seen in Fig. 6.

The good performance of the PSRA2 scheme when combined with the KF suggests evaluating the performance of two hybrid RDASs. Figure 7a shows results for the combined PSF–PSRA2 algorithm, with 54 modes, and Fig. 7b shows results for the combined PEF–PSRA2 algorithm, with 54 modes as well. Both the PSF and PEF filtering strategies include an adaptive tuning procedure for the modeled trailing error covariance matrices $\mathbf{T}_{k|k-1}$ and $\mathbf{T}'_{k|k-1}$, respectively, following CT96. Because the performance of the PEF is only slightly

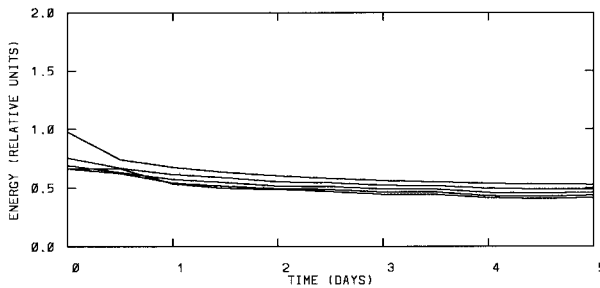


FIG. 8. As in Fig. 2 but for the adaptive CCF scheme and exact retrospective analysis equations.

better than that of the PSF (top curve in each panel), both with 54 modes, the RDASs using these suboptimal filters yield retrospective analyses differing only slightly from each other, as a comparison of the two panels in Fig. 7 indicates. Comparing Fig. 7a (PSF-PSRA2) and Fig. 7b (PEF-PSRA2) with Fig. 6b (KF-PSRA2) shows that employing a suboptimal filter degrades the performance of the retrospective analyses, especially beyond lag two. Comparing Fig. 7a (PSF-PSRA2) to Fig. 5 (PSF-PSRA), however, demonstrates again the superior performance of the PSRA2 scheme, especially for the first two lags.

We evaluate next the performance of schemes that approximate only the filter portion and carry out the retrospective analysis calculations exactly. We start with an RDAS in which the adaptive CCF scheme is used for the filter part. Figure 8 shows the evolution of the actual errors up to day 5 (same as Fig. 3 of Todling et al. 1996). While the performance of the CCF scheme (top curve) is worse than that seen in Fig. 2 for the optimal KF, it is significantly worse only beyond day one; adaptive tuning of more than a single parameter would likely improve this filter result. As a consequence of suboptimality of the CCF scheme, the performance of the CCF-based retrospective analyses shown in Fig. 8 is also suboptimal. However, a comparison between Figs. 2 and 8 indicates that retrospective analysis based on a suboptimal filter can be viewed as a way of improving suboptimal filter performance toward optimal filter performance. For instance, notice that by day 2.5, the lag-2 suboptimal retrospective analysis of Fig. 8 has about the same error level as that of the optimal filter analysis of Fig. 2.

We also see in Fig. 8 that the retrospective analysis at 12 h for lag-4 is worse than the retrospective analyses for smaller lags. This is an indication that we can expect to extract only so much information from the data when using the rather crude approximate forecast error covariance matrix of the CCF. Even though the correct dynamics is still used to propagate filter information from 60 h back to 12 h, the approximate forecast error covariance used by the CCF is largely different from the true forecast error covariance at 60 h, and we expect that the retrospective analysis results can actually de-

grade during the transient period, the more lags we include in the retrospective scheme.

When comparing the RDAS using the CCF scheme (Fig. 8) with the RDASs using the RRF-RRRA of Fig. 4 and the PSF-PSRA of Fig. 5, we see that the performance of the CCF scheme itself is not much different than that of the RRF with 13×12 resolution and that of the PSF with 54 modes (top curve in each figure). The performance of the CCF-based retrospective analysis, however, exceeds that of the 13×12 RRF-based RRA scheme and the 54-mode PSF-based PSRA scheme, for every lag, beyond the initial transient assimilation period. During the transient period, the RRF-RRRA algorithm shows better performance, for high lags, than either the CCF-based retrospective analysis algorithm or the PSF-based PSRA algorithm.

Analogously to Fig. 3, we show in Fig. 9 maps of the actual height analysis error standard deviation at day 0.5, for the experiment of Fig. 8. The panels are arranged as before: (a) filter analysis; (b) lag $l = 1$ retrospective analysis; and (c) lag $l = 4$ retrospective analysis. Comparing panels (a) and (b) with the corresponding panels in Fig. 3, we see that the CCF scheme and the resulting lag $l = 1$ retrospective analysis perform remarkably well. However, the retrospective analysis for lag $l = 4$ (Fig. 9c) is not significantly better than for lag $l = 1$ (Fig. 9b), as one might expect from Fig. 8 at day 0.5, and in fact compares poorly with the optimal case (Fig. 3c), particularly over the data-void central band.

Finally, we examine the performance of the more sophisticated PSF and PEF suboptimal filters and the corresponding suboptimal RDASs, using the exact retrospective analysis formulas. In both cases we retain only 54 leading modes and we adaptively tune the trailing error covariance matrices as before. In Fig. 10a the top curve refers to the performance of the PSF, whereas that in Fig. 10b refers to the performance of the PEF. The PSF result is identical to that displayed in Fig. 5 since the filter here retains the same number of modes as before. A comparison of the PSF-based retrospective analyses of Fig. 10a, which use the exact retrospective analysis formulation, and the PSF-based retrospective analyses of Fig. 5, where this formulation was approximated by the PSRA algorithm, shows clearly the superior performance of the exact formulation. The PSF-based RDAS (Fig. 10a) performance is similar to, and the PEF-based RDAS (Fig. 10b) performance is superior to, that of the CCF-based RDAS of Fig. 8. The RDAS using the PEF (Fig. 10b) presents very good long-term performance, with its results being fairly close to those of the optimal FLKS in Fig. 2, and only slightly inferior to those of the 13×16 RRF-RRRA scheme of Fig. 4a.

In Fig. 11 we show maps of the actual height analysis error standard deviations for the experiment using the PSF of Fig. 10a. Performance relative to the optimal case (Fig. 3) tends to worsen with increasing lag number, particularly over the data-void central band. Compared

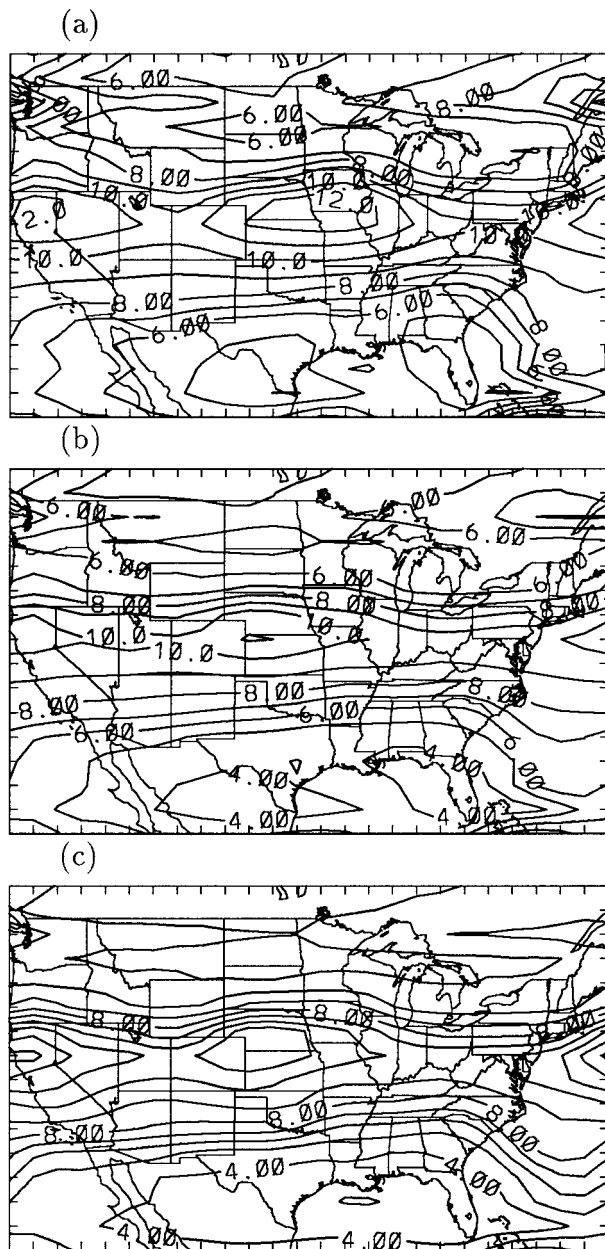


FIG. 9. As in Fig. 3 but using the CCF-based RDAS of Fig. 8.

to the maps of Fig. 9, however, there is improvement in the analyses over specific regions of the domain. In particular, the lag $l = 4$ retrospective analysis in Fig. 11c shows a considerable error reduction beyond that of Fig. 9c over the central part of the domain and the Atlantic Ocean.

5. Conclusions

In this article we evaluated the performance of approximate (suboptimal) RDASs based on the FLKS formulation of Cohn et al. (1994). This formulation has

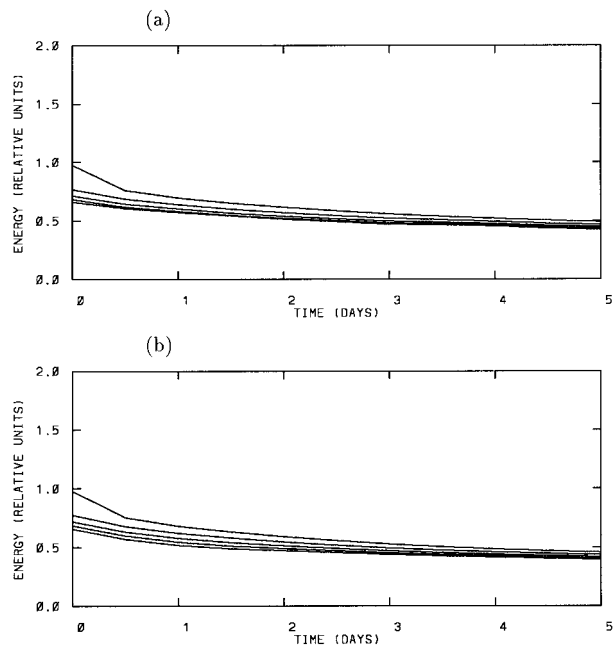


FIG. 10. As in Fig. 2 but using (a) PSF and (b) PEF, both with 54 modes.

several practical advantages over more commonly known smoother formulations. In particular, it avoids a number of large matrix inversions. This formulation also separates naturally into a filter portion and a retrospective analysis portion, enabling a variety of suboptimal implementations. Model error is incorporated implicitly in the retrospective analysis portion, because the filter portion is based directly on the Kalman filter, which already takes model error explicitly into account. Thus, a version of the retrospective analysis portion could be implemented operationally and remain unchanged while improvements in the filter portion, such as accounting for model error, take place.

For linear dynamics and observing systems, performance evaluation equations for approximate RDASs based on the FLKS formulation follow directly from the approach of state augmentation and the usual performance evaluation equations for linear filters utilizing general gain matrices. In this way, we examined the performance of a variety of suboptimal RDASs for a barotropically unstable shallow water model. We concentrated on evaluating the performance obtained when using approximate expressions for the error covariance propagation in the filtering portion of the RDAS, as well as for the error cross-covariance propagation in the retrospective analysis portion of the RDAS. Our experiments indicate that successful retrospective data assimilation schemes can be designed by approximating either the filter portion alone or by approximating both the filter and retrospective analysis portions simultaneously. An important conclusion from these experiments is that a few lags of suboptimal retrospective analysis may ac-

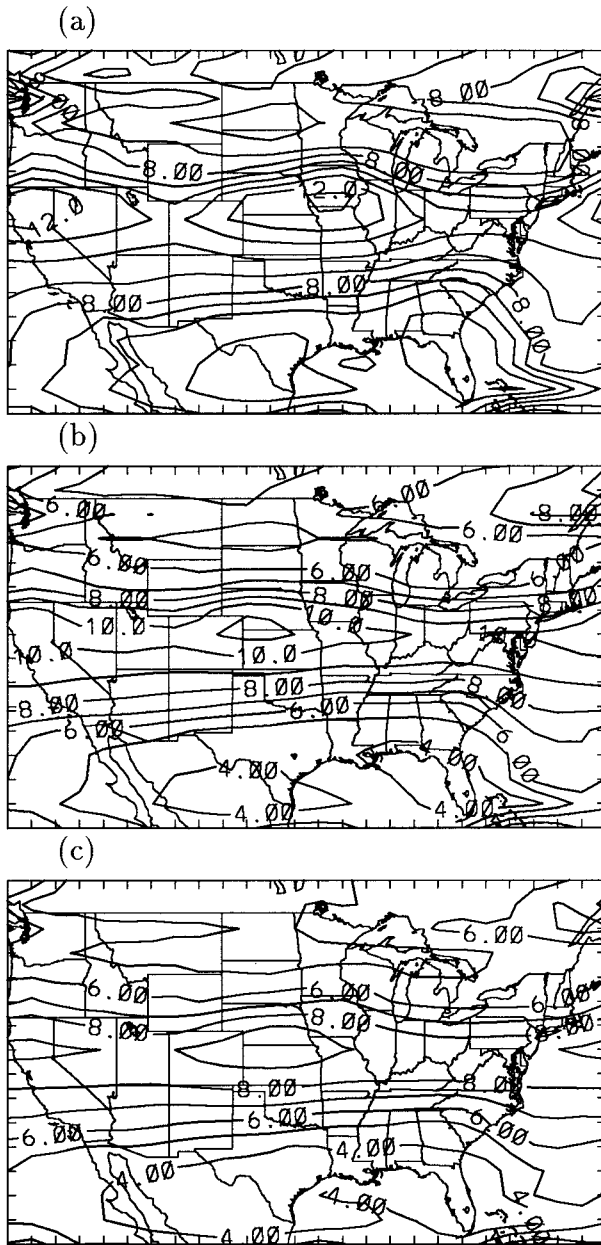


FIG. 11. As in Fig. 3 but using the PSF-based RDAS of Fig. 10a.

comply the performance of an optimal filter analysis. Sophisticated approximate filters that take dynamics of error covariances into account present the best suboptimal retrospective analysis performance. It would be interesting to verify how well these conclusions hold in more realistic data assimilation environments.

Acknowledgments. It is a pleasure to thank Dick P. Dee for many discussions throughout the course of this work. The numerical results were obtained on the Cray C90 through cooperation of the NASA Center for Computational Sciences at the Goddard Space Flight Center.

This research was supported by a fellowship from the Universities Space Research Association (RT) and by the NASA EOS Interdisciplinary Project on Data Assimilation (SEC and NSS).

APPENDIX

Implicit Account of Model Error in Retrospective Analysis

In this appendix we wish to clarify how the retrospective analysis formulation, Eqs. (3)–(5), takes model error into account implicitly. For the sake of argument, we assume here the existence of the inverses of the forecast and analysis error covariance matrices $\mathbf{P}_{k|k-1}^f$ and $\mathbf{P}_{k|k}^a$, respectively, as well as that of the propagator $\mathbf{A}_{k,k-1}$. Using the inverse of the forecast error covariance matrix, we can relate the retrospective analysis gain matrix $\mathbf{K}_{k-l|k}$ to the filter gain matrix $\mathbf{K}_{k|k}$ by

$$\begin{aligned} \mathbf{K}_{k-l|k} &= (\mathbf{P}_{k,k-l|k-1}^{fa})^T (\mathbf{P}_{k|k-1}^f)^{-1} \mathbf{K}_{k|k} \\ &= (\mathbf{P}_{k,k-l|k-1}^{fa})^T (\mathbf{A}_{k,k-1} \mathbf{P}_{k-1|k-1}^a \mathbf{A}_{k,k-1}^T + \mathbf{Q}_k)^{-1} \mathbf{K}_{k|k}, \end{aligned} \tag{A1}$$

where we used (4), (1c), and (1b). This equation already shows one way in which the model error covariance matrix \mathbf{Q}_k is embedded in the retrospective analysis gain matrices.

A way to make the model error contribution explicit in the expression for the retrospective analysis $\mathbf{w}_{k-l|k}^a$ is to convert (3) into a more well-known expression found in the literature (e.g., Gelb 1974, 175). After a tedious manipulation of the formula for the retrospective analysis gain matrix (A1), using both the filter and retrospective analysis update expressions (1d) and (3), it can be shown that

$$\begin{aligned} \mathbf{w}_{k-l|k}^a &= \mathbf{A}_{k-l,k-l-1} \mathbf{w}_{k-l-1|k-1}^a \\ &\quad + \mathbf{U}_{k-l,k-l-1} (\mathbf{w}_{k-l-1|k-1}^a - \mathbf{w}_{k-l-1|k-l-1}^a) \\ &\quad + \mathbf{B}_{k,l} (\mathbf{w}_{k|k}^a - \mathbf{w}_{k|k-1}^f), \end{aligned} \tag{A2}$$

where the $n \times n$ matrices $\mathbf{U}_{k-l,k-l-1}$ and $\mathbf{B}_{k,l}$ are given by

$$\mathbf{U}_{k-l,k-l-1} = \mathbf{Q}_{k-l} (\mathbf{A}_{k-l,k-l-1}^T)^{-1} (\mathbf{P}_{k-l-1|k-l-1}^a)^{-1} \tag{A3a}$$

and

$$\mathbf{B}_{k,l} = \prod_{j=k-l}^{k-1} \mathbf{P}_{j|j}^a \mathbf{A}_{j+1,j}^T (\mathbf{P}_{j+1|j}^f)^{-1}, \tag{A3b}$$

and correspond to the gain matrices in formulation (A2). The retrospective analysis equation (A2) and the gains (A3) are the well-known forms found in Gelb (1974), in a different notation.

For retrospective data assimilation purposes, expression (A2) presents no advantage over (3), particularly due to the appearance of the inverses of the adjoint propagator and of the forecast and analysis error covariance matrices in the gain matrices $\mathbf{B}_{k,l}$ and $\mathbf{U}_{k-l,k-l-1}$.

However, (A2) provides a useful bridge to clarify further the way model error is implicit in the FLKS formulation employed in CST94. Mere algebraic manipulation converts (3) into the more commonly known equation (A2). In this latter expression, the model error covariance matrix \mathbf{Q}_{k-l} appears explicitly through the definition of the gain matrix $\mathbf{U}_{k-l,k-l-1}$. This matrix represents the weights given to the difference between the retrospective analysis $\mathbf{w}_{k-l-1|k-1}^a$ at time t_{k-l-1} , including data up to time t_{k-1} , and the filter analysis $\mathbf{w}_{k-l-1|k-l-1}^a$ at the same time t_{k-l-1} , but using data up to time t_{k-l-1} [see also the discussion in Meditch (1969, 239–240)].

We emphasize that in the formulation of the FLKS employed in CST94, the retrospective analysis (3) is a system driven exclusively by appropriately weighted innovations (cf. Moore 1973): $\mathbf{w}_{k-1|k}^a$ updates the filter analysis $\mathbf{w}_{k-1|k-1}^a$ by the weighted innovation $\mathbf{w}_k^o - \mathbf{H}_k \mathbf{w}_{k-1|k-1}^a$; $\mathbf{w}_{k-2|k}^a$ depends on the same innovation and on $\mathbf{w}_{k-2|k-1}^a$, which in turn updates the filter analysis $\mathbf{w}_{k-2|k-2}^a$ by the weighted innovation $\mathbf{w}_{k-1}^o - \mathbf{H}_{k-1} \mathbf{w}_{k-1|k-2}^a$; and so on. Therefore the retrospective analysis *updates* are obtained by adding weighted innovations to filter analyses, each of which already have incorporated the contribution from model error. On the other hand, the retrospective analyses computed from (A2) are not retrospective analysis *updates*. The retrospective analysis (A2) is a system driven not only by filter analysis increments [last term in (A2)], which are weighted innovations, but also by a weighted difference between the retrospective analysis at the *previous* time and the filter analysis at that time [second term in (A2)].

An illustration of the fact that model error is accounted for implicitly in the FLKS formulation of CST94 can be given for the case of a perfect model—that is, when $\mathbf{Q}_k = \mathbf{0}$ for all t_k . In this case, it follows immediately from (A1) and (5a) that

$$\begin{aligned} \mathbf{K}_{k-l|k} &= (\mathbf{P}_{k-1,k-l|k-1}^{aa})^T \mathbf{A}_{k,k-1}^T (\mathbf{A}_{k,k-1} \mathbf{P}_{k-1|k-1}^a \mathbf{A}_{k,k-1}^T)^{-1} \mathbf{K}_{k|k} \\ &= (\mathbf{P}_{k-1,k-l|k-1}^{aa})^T (\mathbf{P}_{k-1|k-1}^a)^{-1} \mathbf{A}_{k,k-1}^{-1} \mathbf{K}_{k|k}, \end{aligned} \quad (\text{A4})$$

which for $l = 1$ reduces to

$$\mathbf{K}_{k-1|k} = \mathbf{A}_{k,k-1}^{-1} \mathbf{K}_{k|k}, \quad (\text{A5})$$

since $\mathbf{P}_{k-1,k-1|k-1}^{aa} = \mathbf{P}_{k-1|k-1}^a$. For $l = 2$, we have

$$\mathbf{K}_{k-2|k} = (\mathbf{P}_{k-1,k-2|k-1}^{aa})^T (\mathbf{P}_{k-1|k-1}^a)^{-1} \mathbf{K}_{k-1|k}, \quad (\text{A6})$$

where we used result (A5). Taking $l = 1$ and replacing $k \rightarrow k - 1$ in (5b) and (5a), and substituting the result in the expression above, we have

$$\begin{aligned} \mathbf{K}_{k-2|k} &= \mathbf{P}_{k-2|k-2}^a \mathbf{A}_{k-1,k-2}^T (\mathbf{I} - \mathbf{K}_{k-1|k-1} \mathbf{H}_{k-1})^T \\ &\quad \times (\mathbf{P}_{k-1|k-1}^a)^{-1} \mathbf{K}_{k-1|k}, \\ &= \mathbf{P}_{k-2|k-2}^a \mathbf{A}_{k-1,k-2}^T (\mathbf{P}_{k-1|k-2}^a)^{-1} \mathbf{K}_{k-1|k}, \\ &= \mathbf{A}_{k-1,k-2}^{-1} \mathbf{K}_{k-1|k}, \end{aligned} \quad (\text{A7})$$

where the second equality follows from (1e) and the last equality is obtained from (1b) with null model error. Continuing inductively, we can write

$$\mathbf{K}_{k-l|k} = \mathbf{A}_{k-l+1,k-l}^{-1} \mathbf{K}_{k-l+1|k}, \quad (\text{A8})$$

for $l = 1, 2, \dots, L$, which is a simple recursion for the retrospective analysis gain at lag l in terms of the gain at lag $l - 1$, beginning with the filter gain $\mathbf{K}_{k|k}$. An equivalent expression, also for the case of no model error, can be found in Wunsch (1996, 355) for the *fixed-interval* smoother gain.

One might consider using (A8) as an approximation for the retrospective analysis gains (A1) in the case when model error is present. Although the assumption of invertibility of the propagator is extremely stringent for atmospheric data assimilation, a quasi-inverse approximation similar to that of Pu et al. (1997) and Wang et al. (1997) could be employed. Performance evaluation experiments, like those in section 4 of the present article, have been conducted using both of these approximations. Results indicate that the quasi-inverse approximation performs well when (A8) is used, for the perfect model case; the use of (A8) together with the quasi-inverse propagator approximation in the presence of model error, however, does not perform very well in general (results not shown here).

REFERENCES

- Biswas, K. K., and A. K. Mahalanabis, 1972: An approach to fixed-point smoothing problems. *IEEE Trans. Aerosp. Electron. Syst.*, **8**, 676–682.
- , and —, 1973: Suboptimal algorithms for nonlinear smoothing. *IEEE Trans. Aerosp. Electron. Syst.*, **9**, 529–534.
- Bryson, A. E., and M. Frazier, 1963: Smoothing for linear and nonlinear dynamic systems. TDR 63–119, 12 pp. [Available from Wright–Patterson Air Force Base, OH 45433.]
- Campbell, S. L., and C. D. Meyer Jr., 1991: *Generalized Inverses of Linear Transformations*. Dover Publications, 272 pp.
- Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi, 1996: Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. *J. Geophys. Res.*, **101**, 22 599–22 617.
- Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75**, 257–288.
- , and R. Todling, 1996: Approximate Kalman filters for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **74**, 63–75.
- , N. S. Sivakumaran, and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.*, **122**, 2838–2867.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1196.
- Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model. *J. Geophys. Res. Oceans*, **100**, 6777–6793.
- Gelaro, R., E. Klinker, and F. Rabier, 1996: Real and near-real time corrections to forecast initial conditions using adjoint methods: A feasibility study. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., J58–J60.
- Gelb, A., Ed., 1974: *Applied Optimal Estimation*. The MIT Press, 374 pp.
- Gibson, J. K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA description. ECMWF Re-Analysis Project Rep., Ser. 1, ECMWF, Reading, United Kingdom, 72 pp.
- Haltiner, G. J., and R. T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology*. John Wiley and Sons, 477 pp.

- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Kailath, T., 1975: Supplement to "A survey of data smoothing for linear and nonlinear dynamic systems." *Automatica*, **11**, 109–111.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Leondes, C. T., J. B. Peller, and E. B. Stear, 1970: Nonlinear smoothing theory. *IEEE Trans. Syst. Sci. Cybernetics*, **6**, 63–71.
- Meditch, J. S., 1969: *Stochastic Linear Estimation and Control*. McGraw-Hill, 394 pp.
- , 1973: A survey of data smoothing for linear and nonlinear dynamic systems. *Automatica*, **9**, 151–162.
- Ménard, R., and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*, **48A**, 221–237.
- Moore, J. B., 1973: Discrete-time fixed-lag smoothing algorithms. *Automatica*, **9**, 163–173.
- Pu, Z., E. Kalnay, and J. Sela, 1997: Sensitivity of forecast error to initial conditions with a quasi-inverse linear model. *Mon. Wea. Rev.*, **125**, 2479–2503.
- Sage, A. P., 1970: Maximum *a posteriori* filtering and smoothing algorithms. *Int. J. Control*, **11**, 641–658.
- , and W. S. Ewing, 1970: On filtering and smoothing for nonlinear state estimation. *Int. J. Control*, **11**, 1–18.
- , and J. L. Melsa, 1970: *Estimation Theory with Applications to Communications and Control*. McGraw-Hill, 529 pp.
- Schubert, S. D., and R. B. Rood, 1995: Proceedings of the workshop on the GEOS-1 five-year assimilation. NASA Tech. Memo. 104606, Vol. 7, 201 pp. [Available online from <http://dao.gsfc.nasa.gov/subpages/tech-reports.html>.]
- Todling, R., 1995: Computational aspects of Kalman filtering and smoothing for atmospheric data assimilation. *Numerical Simulations in the Environmental and Earth Sciences*, Cambridge University Press, 191–202.
- , and S. E. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.*, **122**, 2530–2557.
- , and —, 1996a: Some strategies for Kalman filtering and smoothing. *Proc. ECMWF Seminar on Data Assimilation*, Reading, United Kingdom, ECMWF, 91–111. [Available online from http://dao.gsfc.nasa.gov/DAO_people/todling.]
- , and —, 1996b: Two reduced resolution filter approaches to data assimilation. *Proc. Ninth Brazilian Meteorological Congress*, São José dos Campos, Brazil, Braz. Meteor. Soc., 1069–1072. [Available online from http://dao.gsfc.nasa.gov/DAO_people/todling.]
- , N. S. Sivakumaran, and S. E. Cohn, 1996: Some strategies for retrospective data assimilation: Approximate fixed-lag Kalman smoothers. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 238–240.
- Verlaan, M., and A. W. Heemink, 1995: Reduced rank square root filters for large scale data assimilation problems. *Proc. Int. Symp. on Assimilation of Observations in Meteorology and Oceanography*, Vol. 1, Tokyo, Japan, WMO, 247–252.
- Wang, Z., K. K. Droegemeier, L. White, and I. M. Navon, 1997: Application of a Newton adjoint algorithm to the 3-D ARPS storm scale model using simulated data. *Mon. Wea. Rev.*, **125**, 2460–2478.
- Willman, W. W., 1969: On the linear smoothing problem. *IEEE Trans. Autom. Control*, **14**, 116–117.
- Wunsch, C., 1996: *The Ocean Circulation Inverse Problems*. Cambridge University Press, 442 pp.