# SOME STRATEGIES FOR KALMAN FILTERING AND SMOOTHING

R TODLING* AND S E COHN**
* USRA/NASA Data Assimilation Office
** NASA/GSFC Data Assimilation Office
Greenbelt, U.S.A.

Summary: In this lecture we use the approach of state augmentation to derive a fixed–lag smoothing algorithm for nonlinear dynamics and observation processes. This extended fixed–lag Kalman smoother involves the commonly–known extended Kalman filter and a corresponding nonlinear extension of the smoother counterpart. For many reasons, this algorithm is impractical for applications to atmospheric data assimilation, which motivates the investigation of approximate schemes. In this regard, we evaluate the performance of approximations to the Kalman filter and the fixed–lag Kalman smoother applied to a linear shallow–water model, for which there is an exact performance evaluation procedure.

## 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.* In that work, the optimal *linear* FLKS was derived and studied in the context of a stable linear shallow–water model. We use this lecture as an opportunity to derive an extension to the FLKS for *nonlinear* dynamics and observing processes. The resulting algorithm is referred to as the *extended* FLKS since its derivation is based on that of the extended Kalman filter (EKF; e.g., Jazwinski 1970, p. 278). As a consequence, the filter portion of the extended FLKS is just the EKF. Brute–force implementation of the extended FLKS to create an operational retrospective data assimilation system (RDAS) is not possible for the same reasons that a brute–force EKF–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 lecture, we also develop and evaluate the performance of potentially useful approximate schemes. To provide an exact evaluation we choose a 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 derivation of the extended FLKS in this lecture, we use 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. This approach could have been adopted to derive the linear FLKS of CST94, as pointed out in that work and also in Anderson and Moore (1979). In the state 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 filter 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 the linear fixed–lag smoother via the same approach. 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. The utility of state augmentation is that the resulting smoothers are often compu-

tationally 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). 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 of CST94, at least for the case of linear systems. Correspondence between smoothers obtained by state augmentation and 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.

In the sequel, we first derive the extended FLKS from the EKF in Section 2. Following that, in Section 3 we briefly outline the performance evaluation technique employed to study the behavior of linear suboptimal filter and smoother algorithms. Section 4 gives a summary of the suboptimal filters and smoothers evaluated in Section 5, in the context of a linear shallow–water model. We draw conclusions in Section 6.

## 2. THE EXTENDED FIXED–LAG KALMAN SMOOTHER

### 2.1 The extended Kalman filter

Let us assume that the $n$–dimensional true state $\mathbf{w}_k^t$ of the atmosphere evolves according to the discrete stochastic equation

$$\mathbf{w}_k^t = \mathbf{f}(\mathbf{w}_{k-1}^t) + \mathbf{b}_k^t , \tag{1}$$

where, for the sake of notational simplicity, we omit possible explicit time dependence of the dynamical operator $\mathbf{f}$; also, for simplicity, we consider only an additive state–independent $n$–vector model error $\mathbf{b}_k^t$. We assume $\mathbf{b}_k^t$ to be white in time, with mean zero and $n \times n$ covariance matrix $\mathbf{Q}_k$:

$$\mathcal{E}\{\mathbf{b}_k^t\} = \mathbf{0} , \tag{2a}$$

$$\mathcal{E}\{\mathbf{b}_k^t (\mathbf{b}_{k'}^t)^T\} = \mathbf{Q}_k \delta_{kk'} , \tag{2b}$$

where $\mathcal{E}\{\}$ denotes the expectation operator, superscript $T$ indicates the transpose operation, and $\delta_{kk'}$ is the Kronecker delta.

We also assume the availability of $p$ observations at each time $t_k$, which relate to the true state nonlinearly according to

$$\mathbf{w}_k^o = \mathbf{h}(\mathbf{w}_k^t) + \mathbf{b}_k^o , \tag{3}$$

where again we omit possible explicit time dependence of the observation operator $\mathbf{h}$. We will assume the $p$–vector observational error $\mathbf{b}_k^o$ to be white in time, with mean zero and $p \times p$ covariance matrix $\mathbf{R}_k$:

$$\mathcal{E}\{\mathbf{b}_k^o\} = \mathbf{0} , \tag{4a}$$

$$\mathcal{E}\{\mathbf{b}_k^o (\mathbf{b}_{k'}^o)^T\} = \mathbf{R}_k \delta_{kk'} . \tag{4b}$$

Moreover, we assume the observation and model errors are uncorrelated at all times:

$$\mathcal{E}\{\mathbf{b}_k^t (\mathbf{b}_{k'}^o)^T\} = \mathbf{0} . \tag{5}$$

The extended Kalman filter for the system (1)–(5) is:

$$\mathbf{w}^f_{k|k-1} = \mathbf{f}(\mathbf{w}^a_{k-1|k-1}) \tag{6a}$$

$$\mathbf{P}^f_{k|k-1} = \mathbf{F}_{k-1|k-1}\mathbf{P}^a_{k-1|k-1}\mathbf{F}^T_{k-1|k-1} + \mathbf{Q}_k \tag{6b}$$

$$\mathbf{K}_{k|k} = \mathbf{P}^f_{k|k-1}\mathbf{H}^T_{k|k-1}\left(\mathbf{H}_{k|k-1}\mathbf{P}^f_{k|k-1}\mathbf{H}^T_{k|k-1} + \mathbf{R}_k\right)^{-1} \tag{6c}$$

$$\mathbf{P}^a_{k|k} = \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_{k|k-1}\right)\mathbf{P}^f_{k|k-1} \tag{6d}$$

$$\mathbf{w}^a_{k|k} = \mathbf{w}^f_{k|k-1} + \mathbf{K}_{k|k}\left(\mathbf{w}^o_k - \mathbf{h}(\mathbf{w}^f_{k|k-1})\right), \tag{6e}$$

where the Jacobians, or tangent linear operators $\mathbf{F}_{k-1|k-1}$ and $\mathbf{H}_{k|k-1}$, are defined by

$$\mathbf{F}_{k-1|k-1} \equiv \mathbf{F}(\mathbf{w}^a_{k-1|k-1}) = \left.\frac{\partial \mathbf{f}(\mathbf{w})}{\partial \mathbf{w}^T}\right|_{\mathbf{w}=\mathbf{w}^a_{k-1|k-1}} \tag{7a}$$

$$\mathbf{H}_{k|k-1} \equiv \mathbf{H}(\mathbf{w}^f_{k|k-1}) = \left.\frac{\partial \mathbf{h}(\mathbf{w})}{\partial \mathbf{w}^T}\right|_{\mathbf{w}=\mathbf{w}^f_{k|k-1}}. \tag{7b}$$

In case the operators $\mathbf{f}$ and $\mathbf{h}$ are *linear*, and the noise processes $\mathbf{b}^t_k$ and $\mathbf{b}^o_k$ are *Gaussian-distributed*, the EKF reduces to the standard Kalman filter. Under these circumstances, the forecast and analysis vectors $\mathbf{w}^f_{k|k-1}$ and $\mathbf{w}^a_{k|k}$, respectively, are precisely the *conditional means*

$$\mathbf{w}^f_{k|k-1} \equiv \mathcal{E}\{\mathbf{w}^t_k|\mathbf{W}^o_{k-1}\}, \tag{8a}$$

$$\mathbf{w}^a_{k|k} \equiv \mathcal{E}\{\mathbf{w}^t_k|\mathbf{W}^o_k\}, \tag{8b}$$

where $\mathbf{W}^o_k = \{\mathbf{w}^o_k, \mathbf{w}^o_{k-1}, \ldots, \mathbf{w}^o_1\}$ denotes the set of all observations up to and including time $t_k$. That is, the forecast at time $t_k$ is the mean of $\mathbf{w}^t_k$ conditioned on observations up to time $t_{k-1}$, whereas the analysis is conditioned on observations up to time $t_k$. Similarly, under these circumstances, the forecast and analysis error covariance matrices $\mathbf{P}^f_{k|k-1}$ and $\mathbf{P}^a_{k|k}$, respectively, are the *conditional error covariances*

$$\mathbf{P}^f_{k|k-1} \equiv \mathcal{E}\{(\mathbf{w}^f_{k|k-1} - \mathbf{w}^t_k)(\mathbf{w}^f_{k|k-1} - \mathbf{w}^t_k)^T|\mathbf{W}^o_{k-1}\}, \tag{9a}$$

$$\mathbf{P}^a_{k|k} \equiv \mathcal{E}\{(\mathbf{w}^a_{k|k} - \mathbf{w}^t_k)(\mathbf{w}^a_{k|k} - \mathbf{w}^t_k)^T|\mathbf{W}^o_k\}, \tag{9b}$$

Relations (8) and (9) provide the rationale for the subscript notation adopted here for the forecast and analysis vectors and their corresponding error covariance matrices, employed also in CST94.

Filter schemes which calculate the conditional means (8) are usually called *optimal* because of their numerous special properties (e.g., Jazwinski 1970, Chapter 5; Cohn 1996). For nonlinear operators $\mathbf{f}$ and $\mathbf{h}$, filters that are optimal in this sense generally require an infinite amount of computation: in general, the EKF calculates the conditional means (8) and covariances (9) in only an approximate sense. In fact, the EKF has many well–known drawbacks, and remedies for some of these have been suggested (e.g., Jazwinski 1970, Chapter 9; Cohn 1993; Miller et al. 1994). In this lecture we use the EKF algorithm only as a tool for deriving the extended fixed–lag Kalman smoother. Since the derivation proceeds directly from the EKF algorithm itself, optimality of the smoother algorithm in the linear, Gaussian case follows. Furthermore, since we show that the extended FLKS can be viewed as a disguised EKF, remedies suggested for the various drawbacks of the EKF may apply fairly readily to the extended FLKS.

We remark that suboptimal filter schemes are often obtained by discarding (6b–d) and replacing the Kalman gain matrix $\mathbf{K}_{k|k}$ with some other gain matrix $\tilde{\mathbf{K}}_{k|k}$ in (6e). While (8) cannot hold for such suboptimal filters, in the linear, Gaussian case the conditional error covariances (9) are given by (6b) and

$$\mathbf{P}^a_{k|k} = (\mathbf{I} - \tilde{\mathbf{K}}_{k|k}\mathbf{H}_{k|k-1})\mathbf{P}^f_{k|k-1}(\mathbf{I} - \tilde{\mathbf{K}}_{k|k}\mathbf{H}_{k|k-1})^T + \tilde{\mathbf{K}}_{k|k}\mathbf{R}_k\tilde{\mathbf{K}}^T_{k|k}\,. \tag{10}$$

Equations (6b) and (10) therefore provide a means to evaluate the performance of suboptimal filter schemes in this case, when the size $n$ of the state space is sufficiently small to render these computations feasible.

## 2.2 The augmented state system

To derive the extended FLKS algorithm using the augmented state approach we introduce an augmented $n(L+1)$–dimensional state vector $\boldsymbol{w}^t_k$:

$$\boldsymbol{w}^t_k \equiv [\mathbf{w}^{tT}_k\ \mathbf{w}^{tT}_{k-1}\ \cdots\ \mathbf{w}^{tT}_{k-L}]^T\,, \tag{11}$$

where *slant characters* denote augmented vectors and matrices. That is, we append the state vector at time $t_k$ with its values at times $t_{k-1},\ldots,t_{k-L}$ when fixed–lag smoother estimates are sought. Furthermore, we append the equalities $\mathbf{w}^t_{k-\ell} = \mathbf{w}^t_{k-\ell}$, for $\ell = 1,2,\ldots,L$ to the state evolution equation (1), so that we can define the evolution of the augmented state (11) according to

$$\begin{aligned}\boldsymbol{w}^t_k &= \begin{bmatrix}\mathbf{f}(\mathbf{w}^t_{k-1})\\ \mathbf{w}^t_{k-1}\\ \vdots\\ \mathbf{w}^t_{k-L}\end{bmatrix} + \begin{bmatrix}\mathbf{b}^t_k\\ \mathbf{0}\\ \vdots\\ \mathbf{0}\end{bmatrix}\\ &= \boldsymbol{f}(\boldsymbol{w}^t_{k-1}) + \boldsymbol{b}^t_k\,,\end{aligned} \tag{12}$$

where the $n(L+1)$–vector $\boldsymbol{b}^t_k \equiv [\mathbf{b}^{tT}_k\,\mathbf{0}^T\,\ldots\,\mathbf{0}^T]^T$ represents the augmented–state model error, which is white in time with mean zero, since $\mathbf{b}^t_k$ has mean zero, and with covariance matrix

$$\boldsymbol{Q}_k = \mathcal{E}\{\boldsymbol{b}^t_k(\boldsymbol{b}^t_k)^T\} = \begin{bmatrix}\mathbf{Q}_k & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \vdots & \vdots & & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\end{bmatrix}\,. \tag{13}$$

Also, the observation process (3)–(4) can be rewritten as

$$\boldsymbol{w}^o_k \equiv \mathbf{w}^o_k = \boldsymbol{h}(\boldsymbol{w}^t_k) + \mathbf{b}^o_k\,, \tag{14}$$

where $\boldsymbol{h}(\boldsymbol{w}^t_k) = \mathbf{h}(\mathbf{w}^t_k)$ is a $p$–vector function, and consequently this process remains unchanged. It follows from (5) that the augmented model error $\boldsymbol{b}^t_k$ and the observation error $\mathbf{b}^o_k$ are uncorrelated:

$$\mathcal{E}\{\boldsymbol{b}^t_k(\mathbf{b}^o_{k'})^T\} = \mathbf{0}\,. \tag{15}$$

The conditional mean "*forecast*" for the (augmented) state system (12)–(15) at time $t_k$, given observations up to and including time $t_{k-1}$, may be written componentwise as

$$\mathcal{E}\{\boldsymbol{w}^t_k|\mathbf{W}_{k-1}\} = \begin{bmatrix}\mathcal{E}\{\mathbf{w}^t_k|\mathbf{W}_{k-1}\}\\ \mathcal{E}\{\mathbf{w}^t_{k-1}|\mathbf{W}_{k-1}\}\\ \vdots\\ \mathcal{E}\{\mathbf{w}^t_{k-L}|\mathbf{W}_{k-1}\}\end{bmatrix}\,, \tag{16}$$

whereas the conditional mean "*analysis*" at time $t_k$, given observations up to and including time $t_k$, reads

$$\mathcal{E}\{\boldsymbol{w}_k^t|\mathbf{W}_k\} = \begin{bmatrix} \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_k\} \\ \mathcal{E}\{\mathbf{w}_{k-1}^t|\mathbf{W}_k\} \\ \vdots \\ \mathcal{E}\{\mathbf{w}_{k-L}^t|\mathbf{W}_k\} \end{bmatrix} . \tag{17}$$

In the linear, Gaussian case, the standard Kalman filter for the augmented system would calculate these conditional mean forecasts and analyses. Relations (16) and (17) motivate us, in the generally *nonlinear* case, to introduce the following notation for the components of the *a priori* augmented state estimate $\boldsymbol{w}_{k|k-1}^f$ at time $t_k$:

$$\boldsymbol{w}_{k|k-1}^f \equiv \begin{bmatrix} \mathbf{w}_{k|k-1}^f \\ \mathbf{w}_{k-1|k-1}^a \\ \vdots \\ \mathbf{w}_{k-L|k-1}^a \end{bmatrix} , \tag{18}$$

as well as the following notation for the components of the *a posteriori* augmented state estimate $\boldsymbol{w}_{k|k}^a$ at time $t_k$:

$$\boldsymbol{w}_{k|k}^a \equiv \begin{bmatrix} \mathbf{w}_{k|k}^a \\ \mathbf{w}_{k-1|k}^a \\ \vdots \\ \mathbf{w}_{k-L|k}^a \end{bmatrix} . \tag{19}$$

We refer to the analyses $\mathbf{w}_{k-\ell|k}^a$ for $\ell = 1, 2, \ldots, L$, as *retrospective analyses.*

By analogy to (6), the EKF algorithm for the augmented system (12)–(15) can be written as

$$\boldsymbol{w}_{k|k-1}^f = \boldsymbol{f}(\boldsymbol{w}_{k-1|k-1}^a) \tag{20a}$$

$$\boldsymbol{P}_{k|k-1}^f = \boldsymbol{F}_{k-1|k-1}\boldsymbol{P}_{k-1|k-1}^a\boldsymbol{F}_{k-1|k-1}^T + \boldsymbol{Q}_k \tag{20b}$$

$$\boldsymbol{K}_{k|k} = \boldsymbol{P}_{k|k-1}^f\boldsymbol{H}_{k|k-1}^T\left(\boldsymbol{H}_{k|k-1}\boldsymbol{P}_{k|k-1}^f\boldsymbol{H}_{k|k-1}^T + \mathbf{R}_k\right)^{-1} \tag{20c}$$

$$\boldsymbol{P}_{k|k}^a = \left(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_{k|k-1}\right)\boldsymbol{P}_{k|k-1}^f \tag{20d}$$

$$\boldsymbol{w}_{k|k}^a = \boldsymbol{w}_{k|k-1}^f + \boldsymbol{K}_{k|k}\left(\boldsymbol{w}_k^o - \boldsymbol{h}(\boldsymbol{w}_{k|k-1}^f)\right) , \tag{20e}$$

in which all the vectors and matrices are merely changed to slant characters, save the observation error covariance matrix $\mathbf{R}_k$ in (20c). The $n(L+1) \times n(L+1)$ Jacobian matrix $\boldsymbol{F}_{k-1|k-1}$ of the augmented dynamics operator is now defined as

$$\boldsymbol{F}_{k-1|k-1} \equiv \left.\frac{\partial \boldsymbol{f}(\boldsymbol{w})}{\partial \boldsymbol{w}^T}\right|_{\boldsymbol{w}=\boldsymbol{w}_{k-1|k-1}^a} , \tag{21}$$

and similarly, the $p \times n(L+1)$ Jacobian matrix $\boldsymbol{H}_{k|k-1}$ of the augmented observation operator is defined as

$$\boldsymbol{H}_{k|k-1} \equiv \left.\frac{\partial \boldsymbol{h}(\boldsymbol{w})}{\partial \boldsymbol{w}^T}\right|_{\boldsymbol{w}=\boldsymbol{w}_{k|k-1}^f} . \tag{22}$$

### 2.3 Matrix–partition unfolding

We now unfold the augmented vector–matrix formulation given above. From (20e) it follows that

$$\begin{bmatrix} \mathbf{w}^a_{k|k} \\ \mathbf{w}^a_{k-1|k} \\ \vdots \\ \mathbf{w}^a_{k-L|k} \end{bmatrix} = \begin{bmatrix} \mathbf{w}^f_{k|k-1} \\ \mathbf{w}^a_{k-1|k-1} \\ \vdots \\ \mathbf{w}^a_{k-L|k-1} \end{bmatrix} + \begin{bmatrix} \mathbf{K}_{k|k} \\ \mathbf{K}_{k-1|k} \\ \vdots \\ \mathbf{K}_{k-L|k} \end{bmatrix} \left( \mathbf{w}^o_k - \mathbf{h}(\mathbf{w}^f_{k|k-1}) \right) , \tag{23}$$

where we noticed that $\boldsymbol{w}^o_k = \mathbf{w}^o_k$ and that $\boldsymbol{h}(\boldsymbol{w}^f_{k|k-1}) = \mathbf{h}(\mathbf{w}^f_{k|k-1})$, and we have also partitioned the $n(L+1) \times p$ gain matrix $\boldsymbol{K}_{k|k}$ of the augmented filter as

$$\boldsymbol{K}_{k|k} = [\mathbf{K}^T_{k|k} \; \mathbf{K}^T_{k-1|k} \; \cdots \; \mathbf{K}^T_{k-L|k}]^T . \tag{24}$$

It follows immediately from (23) that the analysis at times $t_k$ is

$$\boxed{\mathbf{w}^a_{k|k} = \mathbf{w}^f_{k|k-1} + \mathbf{K}_{k|k} \left( \mathbf{w}^o_k - \mathbf{h}(\mathbf{w}^f_{k|k-1}) \right) ,} \tag{25}$$

and the retrospective analyses at times $t_{k-\ell}$ are

$$\boxed{\mathbf{w}^a_{k-\ell|k} = \mathbf{w}^a_{k-\ell|k-1} + \mathbf{K}_{k-\ell|k} \left( \mathbf{w}^o_k - \mathbf{h}(\mathbf{w}^f_{k|k-1}) \right) ,} \tag{26}$$

for $\ell = 1, 2, \ldots, L$, where the equation for the forecast $\mathbf{w}^f_{k|k-1}$ follows from (20a) and the definition of $\boldsymbol{f}$:

$$\boxed{\mathbf{w}^f_{k|k-1} = \mathbf{f}(\mathbf{w}^a_{k-1|k-1}) .} \tag{27}$$

Equations (25) and (27) are identical to equations (6e) and (6a), respectively, as expected.

To unfold the remaining equations of the augmented filter we look next at the expression for the augmented gain (20c). It is appropriate to introduce the following partitioning of the forecast error covariance matrix for the augmented system:

$$\boldsymbol{P}^f_{k|k-1} = \begin{bmatrix} \mathbf{P}^f_{k|k-1} & \mathbf{P}^{fa}_{k,k-1|k-1} & \mathbf{P}^{fa}_{k,k-2|k-1} & \cdots & \mathbf{P}^{fa}_{k,k-L|k-1} \\ \mathbf{P}^{af}_{k-1,k|k-1} & \mathbf{P}^a_{k-1|k-1} & \mathbf{P}^{aa}_{k-1,k-2|k-1} & \cdots & \mathbf{P}^{aa}_{k-1,k-L|k-1} \\ \mathbf{P}^{af}_{k-2,k|k-1} & \mathbf{P}^{aa}_{k-2,k-1|k-1} & \mathbf{P}^a_{k-2|k-1} & \cdots & \mathbf{P}^{aa}_{k-2,k-L|k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{P}^{af}_{k-L,k|k-1} & \mathbf{P}^{aa}_{k-L,k-1|k-1} & \mathbf{P}^{aa}_{k-L,k-2|k-1} & \cdots & \mathbf{P}^a_{k-L|k-1} \end{bmatrix} , \tag{28}$$

where $\mathbf{P}^{af}_{i,j|k} = \mathbf{P}^{faT}_{j,i|k}$.

To unfold (20c) following the partitioning of the augmented gain matrix $\boldsymbol{K}_{k|k}$ in (24), we need to calculate a more explicit expression for the Jacobian (22) of the observation operator $\boldsymbol{h}$. Hence, we write

$$\begin{aligned} \boldsymbol{H}_{k|k-1} &= \left[ \frac{\partial \boldsymbol{h}(\boldsymbol{w})}{\partial \mathbf{w}^T_1} \; \frac{\partial \boldsymbol{h}(\boldsymbol{w})}{\partial \mathbf{w}^T_2} \; \cdots \; \frac{\partial \boldsymbol{h}(\boldsymbol{w})}{\partial \mathbf{w}^T_{L+1}} \right] \Big|_{\boldsymbol{w}=\boldsymbol{w}^f_{k|k-1}} \\ &= \left[ \frac{\partial \mathbf{h}(\mathbf{w}_1)}{\partial \mathbf{w}^T_1} \; \frac{\partial \mathbf{h}(\mathbf{w}_1)}{\partial \mathbf{w}^T_2} \; \cdots \; \frac{\partial \mathbf{h}(\mathbf{w}_1)}{\partial \mathbf{w}^T_{L+1}} \right] \Big|_{\boldsymbol{w}=\boldsymbol{w}^f_{k|k-1}} \\ &= \left[ \mathbf{H}_{k|k-1} \; \mathbf{0} \; \mathbf{0} \; \ldots \; \mathbf{0} \right] , \end{aligned} \tag{29}$$

where we partitioned an arbitrary augmented vector $\boldsymbol{w} = \left[\mathbf{w}_1^T\, \mathbf{w}_2^T \ldots \mathbf{w}_{L+1}^T\right]^T$ into $(L+1)$ component $n$–vectors.

An explicit expression for the gains can be obtained by noticing first that

$$\boldsymbol{P}^f_{k|k-1}\boldsymbol{H}^T_{k|k-1} = \begin{bmatrix} \mathbf{P}^f_{k|k-1}\mathbf{H}^T_{k|k-1} \\ \mathbf{P}^{af}_{k-1,k|k-1}\mathbf{H}^T_{k|k-1} \\ \mathbf{P}^{af}_{k-2,k|k-1}\mathbf{H}^T_{k|k-1} \\ \vdots \\ \mathbf{P}^{af}_{k-L,k|k-1}\mathbf{H}^T_{k|k-1} \end{bmatrix}, \tag{30}$$

and second that

$$\boldsymbol{H}_{k|k-1}\boldsymbol{P}^f_{k|k-1}\boldsymbol{H}^T_{k|k-1} = \mathbf{H}_{k|k-1}\mathbf{P}^f_{k|k-1}\mathbf{H}^T_{k|k-1}\,. \tag{31}$$

Consequently, from (20c) we have

$$\begin{bmatrix} \mathbf{K}_{k|k} \\ \mathbf{K}_{k-1|k} \\ \vdots \\ \mathbf{K}_{k-L|k} \end{bmatrix} = \begin{bmatrix} \mathbf{P}^f_{k|k-1} \\ \mathbf{P}^{af}_{k-1,k|k-1} \\ \vdots \\ \mathbf{P}^{af}_{k-L,k|k-1} \end{bmatrix} \mathbf{H}^T_{k|k-1}\boldsymbol{\Gamma}^{-1}_{k|k-1}\,, \tag{32}$$

where we define the $p \times p$ matrix $\boldsymbol{\Gamma}_{k|k-1}$ as

$$\boldsymbol{\Gamma}_{k|k-1} \equiv \mathbf{H}_{k|k-1}\mathbf{P}^f_{k|k-1}\mathbf{H}^T_{k|k-1} + \mathbf{R}_k\,, \tag{33}$$

for convenience.

Thus, we see that the filter gain is given by

$$\boxed{\mathbf{K}_{k|k} = \mathbf{P}^f_{k|k-1}\mathbf{H}^T_{k|k-1}\boldsymbol{\Gamma}^{-1}_{k|k-1}\,,} \tag{34}$$

which is identical to (6c), and the expressions for the smoother gains can be written compactly as

$$\boxed{\mathbf{K}_{k-\ell|k} = \mathbf{P}^{faT}_{k,k-\ell|k-1}\mathbf{H}^T_{k|k-1}\boldsymbol{\Gamma}^{-1}_{k|k-1}\,,} \tag{35}$$

for $\ell = 1, 2, \cdots, L$, which are identical to those for the linear case (cf. CST94), except that the observation matrix is now replaced by the Jacobian of the observation operator $\mathbf{h}$ evaluated at the current forecast.

By appropriately partitioning the augmented analysis error covariance matrix (20d), it follows that

$$\begin{aligned} \boldsymbol{P}^a_{k|k} &= \begin{bmatrix} \mathbf{P}^a_{k|k} & \mathbf{P}^{aa}_{k,k-1|k} & \mathbf{P}^{aa}_{k,k-2|k} & \cdots & \mathbf{P}^{aa}_{k,k-L|k} \\ \mathbf{P}^{aa}_{k-1,k|k} & \mathbf{P}^a_{k-1|k} & \mathbf{P}^{aa}_{k-1,k-2|k} & \cdots & \mathbf{P}^{aa}_{k-1,k-L|k} \\ \mathbf{P}^{aa}_{k-2,k|k} & \mathbf{P}^{aa}_{k-2,k-1|k} & \mathbf{P}^a_{k-2|k} & \cdots & \mathbf{P}^{aa}_{k-2,k-L|k} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{P}^{aa}_{k-L,k|k} & \mathbf{P}^{aa}_{k-L,k-1|k} & \mathbf{P}^{aa}_{k-L,k-2|k} & \cdots & \mathbf{P}^a_{k-L|k} \end{bmatrix} \\ &= \begin{bmatrix} \left[\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_{k|k-1}\right] & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{K}_{k-1|k}\mathbf{H}_{k|k-1} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{K}_{k-2|k}\mathbf{H}_{k|k-1} & \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ -\mathbf{K}_{k-L|k}\mathbf{H}_{k|k-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \end{bmatrix} \end{aligned}$$

$$\times \begin{bmatrix} \mathbf{P}^{f}_{k|k-1} & \mathbf{P}^{fa}_{k,k-1|k-1} & \mathbf{P}^{fa}_{k,k-2|k-1} & \cdots & \mathbf{P}^{fa}_{k,k-L|k-1} \\ \mathbf{P}^{af}_{k-1,k|k-1} & \mathbf{P}^{a}_{k-1|k-1} & \mathbf{P}^{aa}_{k-1,k-2|k-1} & \cdots & \mathbf{P}^{aa}_{k-1,k-L|k-1} \\ \mathbf{P}^{af}_{k-2,k|k-1} & \mathbf{P}^{aa}_{k-2,k-1|k-1} & \mathbf{P}^{a}_{k-2|k-1} & \cdots & \mathbf{P}^{aa}_{k-2,k-L|k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{P}^{af}_{k-L,k|k-1} & \mathbf{P}^{aa}_{k-L,k-1|k-1} & \mathbf{P}^{aa}_{k-L,k-2|k-1} & \cdots & \mathbf{P}^{a}_{k-L|k-1} \end{bmatrix} . \tag{36}$$

From the main diagonal of the expression above we get

$$\boxed{\mathbf{P}^{a}_{k|k} = \left(\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_{k|k-1}\right)\mathbf{P}^{f}_{k|k-1}} , \tag{37}$$

which is the analysis error covariance expression (6d) of the usual extended Kalman filter, and

$$\begin{aligned} \mathbf{P}^{a}_{k-1|k} &= \mathbf{P}^{a}_{k-1|k-1} - \mathbf{K}_{k-1|k}\mathbf{H}_{k|k-1}\mathbf{P}^{fa}_{k,k-1|k-1} && \text{(38a)} \\ \mathbf{P}^{a}_{k-2|k} &= \mathbf{P}^{a}_{k-2|k-1} - \mathbf{K}_{k-2|k}\mathbf{H}_{k|k-1}\mathbf{P}^{fa}_{k,k-2|k-1} && \text{(38b)} \\ &\vdots \quad \vdots \\ \mathbf{P}^{a}_{k-L|k} &= \mathbf{P}^{a}_{k-L|k-1} - \mathbf{K}_{k-L|k}\mathbf{H}_{k|k-1}\mathbf{P}^{fa}_{k,k-L|k-1} , && \text{(38c)} \end{aligned}$$

which are the retrospective analysis error covariances of the *extended* fixed–lag Kalman smoother. Expressions (38a)–(38c) can be written more compactly as

$$\boxed{\mathbf{P}^{a}_{k-\ell|k} = \mathbf{P}^{a}_{k-\ell|k-1} - \mathbf{K}_{k-\ell|k}\mathbf{H}_{k|k-1}\mathbf{P}^{fa}_{k,k-\ell|k-1}} , \tag{39}$$

for $\ell = 1, 2, \ldots, L$.

From the first row of the upper triangle of (36) we get

$$\begin{aligned} \mathbf{P}^{aa}_{k,k-1|k} &= \left(\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_{k|k-1}\right)\mathbf{P}^{fa}_{k,k-1|k-1} && \text{(40a)} \\ \mathbf{P}^{aa}_{k,k-2|k} &= \left(\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_{k|k-1}\right)\mathbf{P}^{fa}_{k,k-2|k-1} && \text{(40b)} \\ &\vdots \quad \vdots \\ \mathbf{P}^{aa}_{k,k-L|k} &= \left(\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_{k|k-1}\right)\mathbf{P}^{fa}_{k,k-L|k-1} , && \text{(40c)} \end{aligned}$$

which can be written compactly as

$$\boxed{\mathbf{P}^{aa}_{k,k-\ell|k} = \left(\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_{k|k-1}\right)\mathbf{P}^{fa}_{k,k-\ell|k-1}} , \tag{41}$$

for $\ell = 1, 2, \ldots, L$.

From the first column of the lower triangle of (36) we get

$$\begin{aligned} \mathbf{P}^{aa}_{k-1,k|k} &= \mathbf{P}^{af}_{k-1,k|k-1} - \mathbf{K}_{k-1|k}\mathbf{H}_{k|k-1}\mathbf{P}^{f}_{k|k-1} && \text{(42a)} \\ \mathbf{P}^{aa}_{k-2,k|k} &= \mathbf{P}^{af}_{k-2,k|k-1} - \mathbf{K}_{k-2|k}\mathbf{H}_{k|k-1}\mathbf{P}^{f}_{k|k-1} && \text{(42b)} \\ &\vdots \quad \vdots \\ \mathbf{P}^{aa}_{k-L,k|k} &= \mathbf{P}^{af}_{k-L,k|k-1} - \mathbf{K}_{k-L|k}\mathbf{H}_{k|k-1}\mathbf{P}^{f}_{k|k-1} , && \text{(42c)} \end{aligned}$$

which can be written compactly as

$$\mathbf{P}^{aa}_{k-\ell,k|k} = \mathbf{P}^{af}_{k-\ell,k|k-1} - \mathbf{K}_{k-\ell|k}\mathbf{H}_{k|k-1}\mathbf{P}^{f}_{k|k-1} , \tag{43}$$

for $\ell = 1, 2, \ldots, L$. At first glance this expression does not seem to be the transpose of (41) as it should be since the augmented matrix $\boldsymbol{P}^a_{k|k}$ is symmetric. To show that (43) is indeed the transpose of (41) let us replace the smoother gain in (43) by its explicit expression (35), that is,

$$\begin{aligned}
\mathbf{P}^{aa}_{k-\ell,k|k} &= \mathbf{P}^{af}_{k-\ell,k|k-1} - \mathbf{P}^{af}_{k-\ell,k|k-1}\mathbf{H}^T_{k|k-1}\boldsymbol{\Gamma}^{-1}_{k|k-1}\mathbf{H}_{k|k-1}\mathbf{P}^f_{k|k-1} \\
&= \mathbf{P}^{af}_{k-\ell,k|k-1} - \mathbf{P}^{af}_{k-\ell,k|k-1}\mathbf{H}^T_{k|k-1}\mathbf{K}^T_{k|k} \\
&= \left\{\left(\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_{k|k-1}\right)\mathbf{P}^{fa}_{k,k-\ell|k-1}\right\}^T ,
\end{aligned} \tag{44}$$

which is just the transpose of (41). All other off–diagonal elements of (36) reduce to a time–shift of expression (41) for $\mathbf{P}^{aa}_{k,k-\ell|k}$.

All we have left to unfold now is (20b), the equation for the error covariance propagation of the augmented system. To do so, we first calculate a more explicit expression for the Jacobian (21) of $\boldsymbol{f}$:

$$\begin{aligned}
\boldsymbol{F}_{k-1|k-1} &= \left[\begin{array}{c} \frac{\partial \mathbf{f}(\mathbf{w}_1)}{\partial \boldsymbol{w}^T} \\ \frac{\partial \mathbf{w}_1}{\partial \boldsymbol{w}^T} \\ \frac{\partial \mathbf{w}_2}{\partial \boldsymbol{w}^T} \\ \vdots \\ \frac{\partial \mathbf{w}_L}{\partial \boldsymbol{w}^T} \end{array}\right]\Bigg|_{\boldsymbol{w}=\boldsymbol{w}^a_{k-1|k-1}} \\
&= \left[\begin{array}{cccc} \frac{\partial \mathbf{f}(\mathbf{w}_1)}{\partial \mathbf{w}_1^T} & \frac{\partial \mathbf{f}(\mathbf{w}_1)}{\partial \mathbf{w}_2^T} & \cdots & \frac{\partial \mathbf{f}(\mathbf{w}_1)}{\partial \mathbf{w}_L^T} \quad \frac{\partial \mathbf{f}(\mathbf{w}_1)}{\partial \mathbf{w}_{L+1}^T} \\ \frac{\partial \mathbf{w}_1}{\partial \mathbf{w}_1^T} & \frac{\partial \mathbf{w}_1}{\partial \mathbf{w}_2^T} & \cdots & \frac{\partial \mathbf{w}_1}{\partial \mathbf{w}_L^T} \quad \frac{\partial \mathbf{w}_1}{\partial \mathbf{w}_{L+1}^T} \\ \frac{\partial \mathbf{w}_2}{\partial \mathbf{w}_1^T} & \frac{\partial \mathbf{w}_2}{\partial \mathbf{w}_2^T} & \cdots & \frac{\partial \mathbf{w}_2}{\partial \mathbf{w}_L^T} \quad \frac{\partial \mathbf{w}_2}{\partial \mathbf{w}_{L+1}^T} \\ \vdots & \vdots & & \vdots \quad \vdots \\ \frac{\partial \mathbf{w}_L}{\partial \mathbf{w}_1^T} & \frac{\partial \mathbf{w}_L}{\partial \mathbf{w}_2^T} & \cdots & \frac{\partial \mathbf{w}_L}{\partial \mathbf{w}_L^T} \quad \frac{\partial \mathbf{w}_L}{\partial \mathbf{w}_{L+1}^T} \end{array}\right]\Bigg|_{\boldsymbol{w}=\boldsymbol{w}^a_{k-1|k-1}} \\
&= \left[\begin{array}{ccccc} \mathbf{F}_{k-1|k-1} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I} & \mathbf{0} \end{array}\right] .
\end{aligned} \tag{45}$$

Using this Jacobian matrix in expression (20b), it follows that

$$\boxed{\mathbf{P}^f_{k|k-1} = \mathbf{F}_{k-1|k-1}\mathbf{P}^a_{k-1|k-1}\mathbf{F}^T_{k-1|k-1} + \mathbf{Q}_k ,} \tag{46}$$

and

$$\boxed{\mathbf{P}^{fa}_{k,k-\ell|k-1} = \mathbf{F}_{k-1|k-1}\mathbf{P}^{aa}_{k-1,k-\ell|k-1} ,} \tag{47}$$

for $\ell = 1, 2, \ldots, L$, which give the forecast error covariance propagation for the extended Kalman filter, and the propagation of the fixed–lag smoother forecast–analysis error cross–covariance, under the tangent linear approximation, respectively. The expressions in framed boxes above are the formulas for the extended FLKS.

## 3. PERFORMANCE ANALYSIS FOR LINEAR SYSTEMS

In this section we summarize the suboptimal schemes to be evaluated in the context of the *linear* shallow–water model of the following section. These schemes approximate only the filter and smoother gains (34) and (35), along with the innovation covariance (33) on which they depend, by replacing them with gains $\tilde{\mathbf{K}}_{k|k}$ and $\tilde{\mathbf{K}}_{k-\ell|k}$ identical in form to (34) and (35) but involving approximate expressions for $\mathbf{P}^f_{k|k-1}$ and $\mathbf{P}^{fa}_{k,k-\ell|k-1}$. Thus we will be concerned with approximate expressions for the propagated (predictability) error covariance matrix

$$\mathbf{P}^p_{k|k-1} \equiv \mathbf{A}_{k,k-1}\mathbf{P}^a_{k-1|k-1}\mathbf{A}^T_{k,k-1}\,, \tag{48}$$

and approximate expressions for the forecast–analysis error cross–covariance matrix

$$\mathbf{P}^{fa}_{k,k-\ell|k-1} = \mathbf{A}_{k,k-1}\mathbf{P}^{aa}_{k-1,k-\ell|k-1}\,, \tag{49}$$

where for the linear case the tangent linear propagator is a state–independent operator, indicated by $\mathbf{F}_{k-1|k-1} = \mathbf{A}_{k,k-1}$. These expressions correspond to the most computationally demanding part of the filter and smoother algorithms (cf. Todling 1995). To focus on the issue of performance due to approximating the filter and smoother, 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.

For linear systems, performance evaluation can be accomplished following the procedure of Todling and Cohn (1994), but now extended to incorporate the smoother performance analysis equations as well. These equations can be obtained from an augmented–state version of the performance analysis expression (10), valid for general (filter and smoother) gain matrices:

$$\mathbf{P}^a_{k|k} = (\mathbf{I} - \tilde{\mathbf{K}}_{k|k}\mathbf{H}_k)\mathbf{P}^f_{k|k-1}(\mathbf{I} - \tilde{\mathbf{K}}_{k|k}\mathbf{H}_k)^T + \tilde{\mathbf{K}}_{k|k}\mathbf{R}_k\tilde{\mathbf{K}}^T_{k|k} \tag{50a}$$

$$\begin{aligned}\mathbf{P}^a_{k-\ell|k} = {} & \mathbf{P}^a_{k-\ell|k-1} + \tilde{\mathbf{K}}_{k-\ell|k}\boldsymbol{\Gamma}_{k|k-1}\tilde{\mathbf{K}}^T_{k-\ell|k} \\ & - \tilde{\mathbf{K}}_{k-\ell|k}\mathbf{H}_k\mathbf{P}^{fa}_{k,k-\ell|k-1} - (\mathbf{H}_k\mathbf{P}^{fa}_{k,k-\ell|k-1})^T\tilde{\mathbf{K}}^T_{k-\ell|k}\end{aligned} \tag{50b}$$

$$\begin{aligned}\mathbf{P}^{aa}_{k,k-\ell|k} = {} & \mathbf{P}^{fa}_{k,k-\ell|k-1} + \tilde{\mathbf{K}}_{k|k}\boldsymbol{\Gamma}_{k|k-1}\tilde{\mathbf{K}}^T_{k-\ell|k} \\ & - \tilde{\mathbf{K}}_{k|k}\mathbf{H}_k\mathbf{P}^{fa}_{k,k-\ell|k-1} - \mathbf{P}^f_{k|k-1}\mathbf{H}^T_k\tilde{\mathbf{K}}^T_{k-\ell|k}\,,\end{aligned} \tag{50c}$$

where for the linear case we write $\mathbf{H}_{k|k-1} = \mathbf{H}_k$, since the Jacobian of the observation operator is then independent of the state estimate. Together with (48) and (49), these equations give the update and evolution of the *actual* filter and smoother error covariances. Expression (50a) [same as (10)] is the well–known Joseph equation, and gives the performance of the filter analysis for a general gain matrix $\tilde{\mathbf{K}}_{k|k}$, while (50b) and (50c) give the performance of the smoother analyses for general gains $\tilde{\mathbf{K}}_{k|k}$ and $\tilde{\mathbf{K}}_{k-\ell|k}$. These performance analysis equations also appear in the derivation of the FLKS of CST94 [cf. eqs. (2.33), (2.39) and (2.45) in CST94], in a slightly different form.

## 4. SUMMARY OF SUBOPTIMAL FILTERS AND SMOOTHERS

The following are the suboptimal *filter* schemes considered in this lecture (see Cohn and Todling 1996, CT96 hereafter, and also Todling et al. 1996):

*(1.a) Constant Error Covariance (CEC):*

Here the predictability error covariance $\mathbf{P}^p_{k|k-1}$ is replaced in the gain expression (34) by

$$\mathbf{S}^p_{k|k-1} = \alpha_k \bar{\mathbf{S}}\,, \tag{51}$$

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

*(1.b) Partial Singular–Value Decomposition Filter (PSF):*

In the PSF, the dynamical operator $\mathbf{A}_{k,k-1}$ is replaced by the leading part of its singular value decomposition, here abbreviated by $\tilde{\mathbf{A}}_{k,k-1}$, and the predictability error covariance is simplified to:

$$\mathbf{S}^p_{k|k-1} = \tilde{\mathbf{A}}_{k,k-1}\mathbf{S}^a_{k-1|k-1}\tilde{\mathbf{A}}^T_{k,k-1} + \mathbf{T}_{k|k-1}\,, \tag{52}$$

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.

*(1.c) Partial Eigendecomposition Filter (PEF):*

In the PEF, the entire predictability error covariance is replaced by the leading part of its eigendecomposition, which ideally explains most of the variance, that is,

$$\mathbf{S}^p_{k|k-1} = (\mathbf{W}_N\hat{\mathbf{S}}_N\mathbf{W}^T_N)_{k,k-1} + \mathbf{T}'_{k|k-1}\,, \tag{53}$$

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 $\hat{\mathbf{S}}_N$, and $\mathbf{T}'_{k|k-1}$ is an estimate representing the trailing error covariance matrix of this approximation, not necessarily equal to $\mathbf{T}_{k|k-1}$. This approach resembles the reduced–rank square–root filter of Verlaan and Heemink (1995).

*(1.d) Reduced Resolution Filter (RRF):*

This approximation follows the approach of Fukumori and Malanotte–Rizzoli (1995; see also Todling and Cohn 1996) and involves carrying the error covariances at lower resolution than that of the state estimates. In this case, the predictability error covariance is written as

$$\tilde{\mathbf{S}}^p_{k|k-1} = (\mathbf{B}^+\mathbf{A}_{k,k-1}\mathbf{B})\tilde{\mathbf{S}}^a_{k-1|k-1}(\mathbf{B}^+\mathbf{A}_{k,k-1}\mathbf{B})^T + \mathbf{T}''_{k|k-1}\,, \tag{54}$$

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 $\tilde{\mathbf{S}}^a_{k-1|k-1}$ and $\tilde{\mathbf{S}}^p_{k|k-1}$ are represented to the $n$–dimensional space of the state estimates; the matrix $\mathbf{B}^+$ represents an inverse of the interpolation operator $\mathbf{B}$, which in our experiments is taken to be the Moore–Penrose pseudo–inverse (e.g., Campbell and Meyer 1991).

The following are the suboptimal *smoothers* considered here:

*(2.a) Partial Singular–Value Decomposition Smoother (PSS):*

The PSS approximation simplifies the evolving of the $n$ columns of $\mathbf{P}^{aa}_{k-1,k-\ell|k-1}$ by replacing the propagator by a reduced–rank propagator similar to that used in the PSF:

$$\mathbf{S}^{fa}_{k,k-\ell|k-1} = \tilde{\mathbf{A}}_{k,k-1}\mathbf{S}^{aa}_{k-1,k-\ell|k-1} + \mathbf{X}_{k,k-\ell|k-1}\,, \tag{55}$$

where $\mathbf{X}_{k,k-\ell|k-1}$ is a trailing error cross–covariance matrix. Notice that, in principle, the reduced–rank dynamics here does not have to be the same as that in the PSF. However, in the experiments discussed below they are chosen to be equal, whenever the PSF and PSS are employed simultaneously. Also, in the experiments reported here we take $\mathbf{X}_{k,k-\ell|k-1} = \mathbf{0}$, at all times $t_k$.

*(2.b) Reduced Resolution Smoother (RRS):*

In the RRS, by analogy with the RRF approximation, we compute the smoother forecast–analysis error cross–covariance at reduced resolution:

$$\tilde{\mathbf{S}}^{fa}_{k,k-\ell|k-1} = (\mathbf{B}^{+}\mathbf{A}_{k,k-1}\mathbf{B})\tilde{\mathbf{S}}^{aa}_{k-1,k-\ell|k-1} + \mathbf{X}'_{k,k-\ell|k-1}\,, \tag{56}$$

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

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

We can construct approximate RDASs by combining different strategies for approximating the filter and the smoother. For instance, one could choose to approximate the filter and the smoother equally, i.e., with two similar schemes like the RRF and RRS at the same resolution; or one could choose to approximate the filter and calculate the smoother equations exactly, that is, to approximate (48) and use (49); one could also build hybrid approximations in which the filter and the smoother employ different strategies. In any case, since our formulation of the smoother is based on the filter, whenever the filter is approximated the smoother becomes suboptimal. Notice that the converse is not true, that is, if the filter is kept exact and the smoother is approximated — if we use (48) and approximate (49) — only the smoother becomes suboptimal, but not the filter. This may not be a very useful approach, since major computational requirements are associated with the filter equation (48). Moreover, as discussed before, in the general nonlinear case the EKF does not always represent a good solution to the filtering problem, thus the smoother inherits the EKF's weaknesses.

## 5. 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, p. 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 hours and distributed outside the strongest part of the jet. The tick marks in the figure indicate the $25 \times 16$ model grid. In the experiments referring to a trailing error covariance matrix we construct it, as in CT96, using the slow eigenmodes of the autonomous unstable dynamics of our shallow–water model.

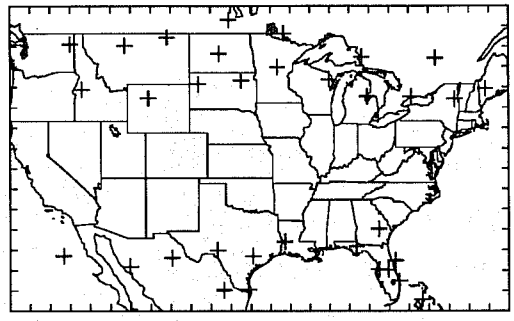

Figure 1: Radiosonde network composed of 33 stations observing winds and heights every 12 hours (same as Fig. 2 of CT96).

Before evaluating the performance of a few suboptimal RDASs, we discuss the results obtained by 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) error in the total energy as a function of time. The top curve corresponds to the filter result while successive retrospective analysis results are given by successively lower curves, which refer to analyses including data 12, 24, 36 and 48 hours ahead in time, that is, lags $\ell = 1, 2, 3$, and 4. The filter curve is the same as that seen 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.

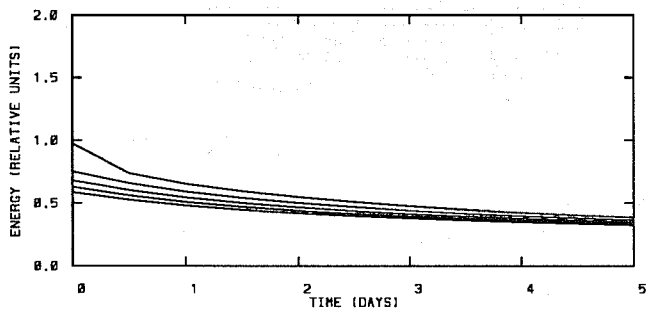


Figure 2: ERMS analysis error in total energy for the Kalman filter (upper curve) and fixed-lag Kalman smoother (lower curves).

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 [panel (a)], and for the smoother analysis errors for lags $\ell = 1$ [panel (b)] and $\ell = 4$ [panel (c)]. Thus, in panels (b) and (c) the analysis errors are reduced by incorporating data 12 and 48 hours ahead of the current analysis time, respectively. We see not only the overall decrease in error levels from panel (a) to panel (c), as expected from Fig. 2, but also that within each panel errors are highest in the central band of the domain, where there are no observations and where the jet is strongest. Furthermore, we notice that 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 panels (a) to (c)]. This property of the

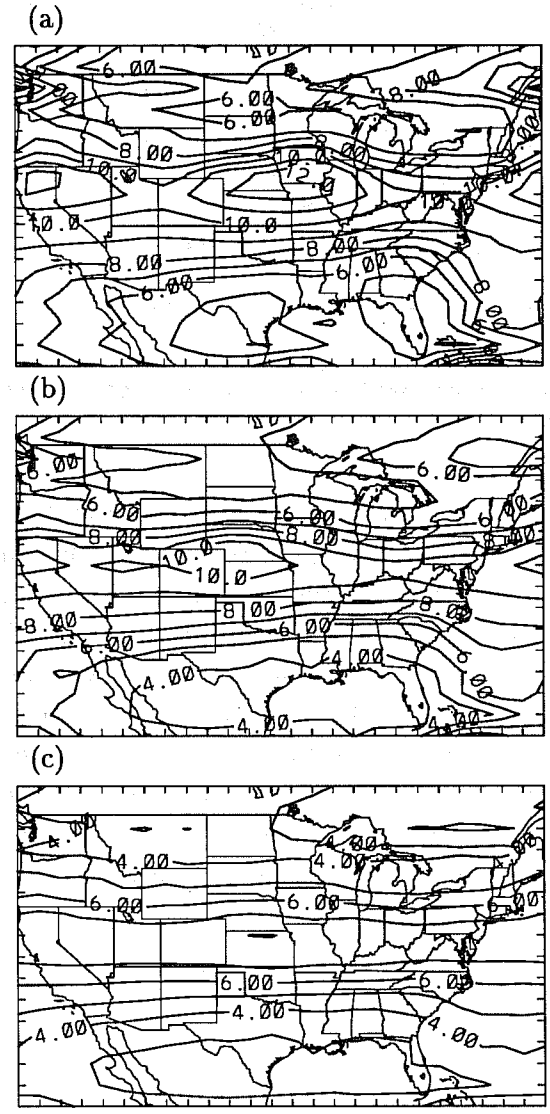


Figure 3: Analysis error standard deviation in the height field at time $t = 0.5$ days. Panel (a) is for the filter analysis; panels (b) and (c) are for the smoother analyses with lags $\ell = 1$ and 4, respectively.

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 both the filter and the smoother similarly. In this category, we investigate the behavior of the RRF–and–RRS corresponding to expressions (54)–and–(56), respectively, as well as the behavior of the PSF–and–PSS corresponding to expressions (52)–and–(55), respectively.

The results of Todling and Cohn (1996) show that the RRF described above, with resolutions $13 \times 16$ and $13 \times 12$, provides good filter performance in our shallow–water model context. This is mainly attributed 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 $13 \times 16$. In these experiments, each column of the interpolation matrix $\mathbf{B}$ in (54) and (56) corresponds to a spline interpolant with period boundary conditions in the east–west direction and an Akima spline in the north–south direction. Hence, in Fig. 4 we show results of the performance analysis

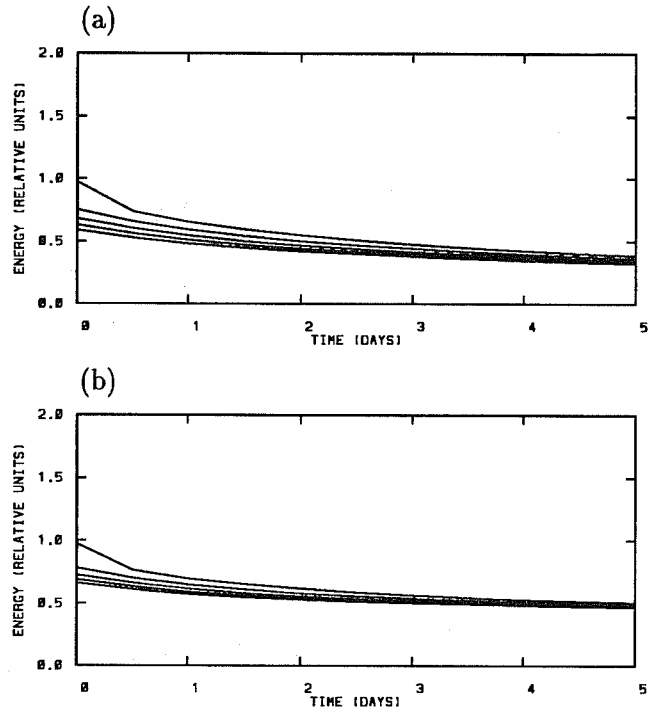


Figure 4: RRF/RRS for resolutions: (a) $13 \times 16$, (b) $13 \times 12$.

for the RRF and RRS algorithms at these resolutions [panel (a) for $13 \times 16$; panel (b) for $13 \times 12$]. As in Fig. 2, the upper curve in each panel is for the performance of the filter while the lower curves in each panel are for the performance of the corresponding RDAS. A comparison of panel (a) with the optimal FLKS results of Fig. 2 shows remarkable agreement when the jet is fully resolved. The agreement for the coarse resolution result in panel (b) is still quite good, especially during the transient part of the assimilation. Asymptotically, the error levels for the case with $13 \times 12$ resolution are somewhat high, with the smoother having less of an impact than at $13 \times 16$ resolution.

Along similar lines, we investigate the performance of an RDAS using the PSF as the algorithm to compute the predictability error covariance for the filter part, and the PSS algorithm as the procedure to compute the forecast–analysis error cross–covariance for the smoother part. From the experiments of CT96, we know that using the first 54 singular modes of the 12–hour propagator of our linear shallow–water model is enough to produce a stable suboptimal filter in this context. 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, that is, we take $\mathbf{X}_{k,k-\ell|k-1} = \mathbf{0}$ at all times.

Fig. 5 shows performance results for the PSF–PSS suboptimal RDAS when 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. The PSS smoother, on the other hand, does not perform nearly as well as the optimal smoother (Fig. 2), with little difference among results for lag $\ell = 1$ and those for higher lags $\ell = 2, 3$, and 4. This poor smoother performance may be attributed in part to the neglected trailing forecast–analysis error cross–covariance $\mathbf{X}_{k,k-\ell|k-1}$.

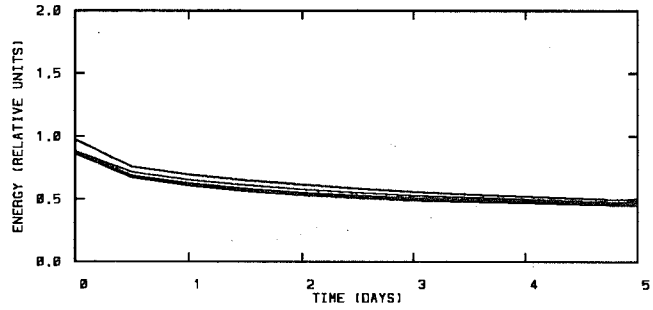


Figure 5: As in Fig. 2, but for an approximate RDAS using the PSF and PSS simultaneously, both with 54 modes.

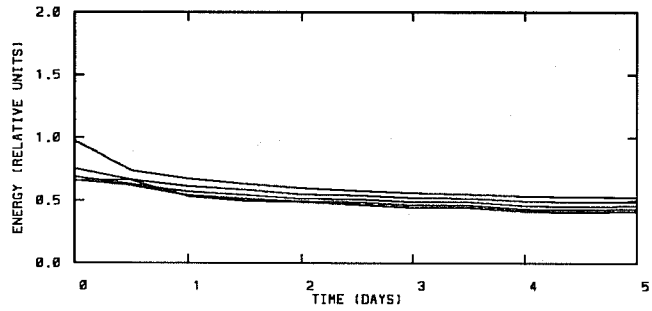


Figure 6: As in Fig. 2, but for the adaptive CEC filter and exact smoother.

We evaluate next the performance of schemes that approximate only the filter but leave the smoother calculations unchanged. That is, the approximations now take one of the suboptimal filters (51)–(54), and the exact smoother expression (49). We start with an approximate RDAS in which the CEC scheme is utilized for the filter part. Fig. 6 shows the evolution of the actual ERMS errors up to day 5 (same as Fig. 3 of Todling et al. 1996). While the performance of the CEC filter (top curve) is worse than that seen in Fig. 2 for the optimal KF, it is not dramatically worse. As a consequence of suboptimality of the CEC filter, the performance of the CEC–based retrospective analyses shown in Fig. 6 is also suboptimal. However, a comparison between Figs. 2 and 6 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–1 suboptimal retrospective analysis of Fig. 6 has about the same error level as that of the optimal filter analysis of Fig. 2.

When comparing the RDAS using the CEC filter (Fig. 6) with the suboptimal RDASs using the RRF–RRS of Fig. 4 and the PSF–PSS of Fig. 5, we see that the performance of the CEC filter itself is not much different than that of the RRF with $13 \times 12$ resolution and that of the PSF with 54 modes. The performance of the CEC–based smoother, however, exceeds that of the $13 \times 12$ RRF–RRS and the 54–mode PSF–PSS, for every lag, beyond the initial transient assimilation period. During the transient assimilation period, the RRS shows better performance, for high lags, than either the CEC–based RDAS or the PSS.

Analogously to Fig. 3, we show in Fig. 7 maps of the actual analysis error standard deviation for the height field at day 0.5, for the RDAS experiments of Fig. 6. The panels are arranged as before: (a) filter analysis; (b) lag $\ell = 1$ smoother analysis; and (c) lag $\ell = 4$ smoother analysis. Comparing panels (a) and (b) with the corresponding panels in Fig. 3, we see that the CEC

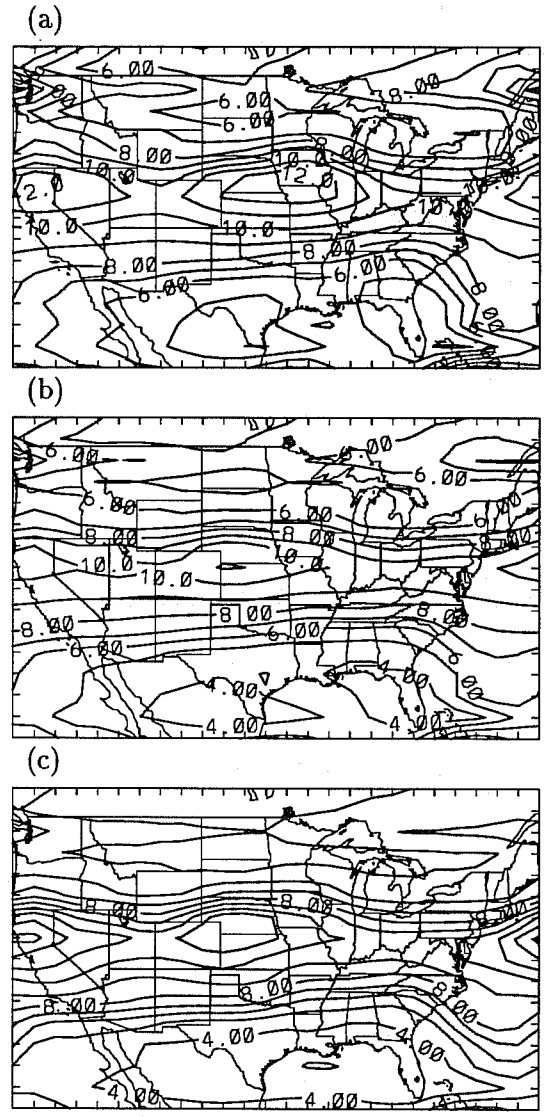


Figure 7: As in Fig. 3, but using the CEC–based RDAS of Fig. 6.

filter and the lag $\ell = 1$ smoother perform relatively well. However, the retrospective analysis for lag $\ell = 4$ (Fig. 7c) is not significantly better than for lag $\ell = 1$ (Fig. 7b), as one might expect from Fig. 6 at day 0.5, and in fact compares poorly with the optimal case (Fig. 3c), particularly over the data–void central band.

Next we examine the performance of the more sophisticated PSF and PEF suboptimal filters and the corresponding suboptimal RDASs, using (49) for the smoother portion. In both cases we retain only 54 leading modes and we adaptively tune a trailing error covariance matrix as in CT96. In Fig. 8, the top curve in panel (a) refers to the performance of the PSF, while that in panel (b) 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. 8a, which use the exact smoother formulation (49), and Fig. 5, where the smoother was approximated by the PSS, shows the superior performance of the exact smoother formulation. This result comes as no surprise, and the PSF–based RDAS incurs higher computational cost. The PSF–based RDAS performance (Fig. 8a) is similar to, and the PEF–based RDAS performance (Fig. 8b) is superior to, that of the CEC–based RDAS

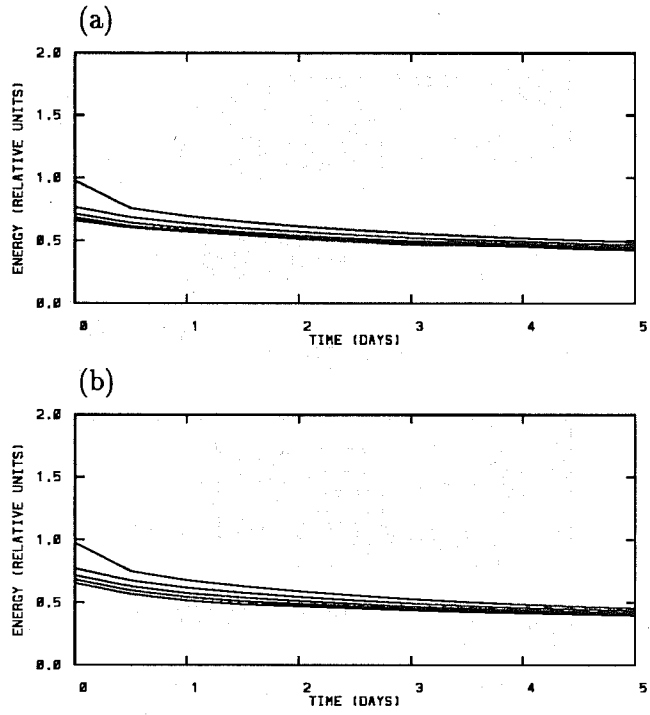


Figure 8: As in Fig. 2, but using the PSF [panel (a)] and the PEF [panel (b)], both with 54 modes.

of Fig. 6. The RDAS using the PEF (Fig. 8b) presents very good long–term performance, with its results being very close to those of the optimal FLKS in Fig. 2, and only slightly inferior to those of the $13 \times 16$ RRF–RRS of Fig. 4a.

Finally, in Fig. 9 we show the maps of actual height analysis error standard deviations for the experiment using the PSF of Fig. 8a. Compared to the maps of Fig. 7, there is improvement in the analyses over specific regions of the domain. In particular, the lag $\ell = 4$ analysis in panel (c) shows a considerable error reduction over the central part of the domain and the Atlantic Ocean.

## 6. CONCLUSIONS

In this lecture we derived an algorithm for performing retrospective analysis for nonlinear systems. The extended fixed–lag Kalman smoother (FLKS) was obtained from the extended Kalman filter (EKF) via the approach of state augmentation. The properties of the extended FLKS follow directly from those of the EKF. In particular, remedies proposed elsewhere to cure EKF weaknesses should apply to the extended FLKS. Feasibility of this algorithm to construct an operational retrospective data assimilation system (RDAS) is constrained by its large computational requirements and presumed knowledge of the required statistics, in much the same way as the EKF is constrained for practical implementation. This motivates the search for feasible and reliable approximate RDAS schemes based on the extended FLKS.

For linear dynamics and observing systems, performance analysis equations for approximate RDASs based on the FLKS follow directly from the approach of state augmentation and the usual performance analysis equation for linear filters utilizing general gain matrices. In this context, we examined the performance of a variety of possible suboptimal RDASs for a barotropically

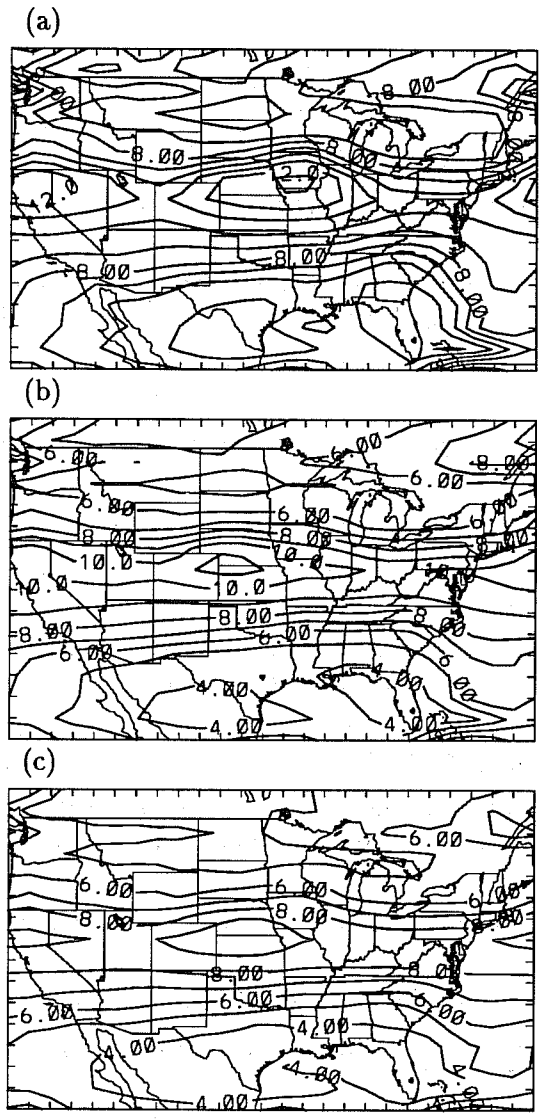


Figure 9: As in Fig. 3, but using the PSF–based RDAS of Fig. 8a.

unstable shallow–water model. We concentrated on evaluating the performance of alternative expressions for the error covariance propagation in the filtering part of the RDAS, as well as for the error cross–covariance propagation in the smoothing part of the RDAS. Our experiments indicate that successful retrospective analysis schemes can be designed by approximating either the filter alone or by approximating both the filter and smoother simultaneously. An important conclusion from these experiments is that a few lags of suboptimal retrospective analysis may accomplish 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.

*Acknowledgments.* It is a pleasure to thank N.S. Sivakumaran and D. 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).

## References


Anderson, B.D.O., and J.B. Moore, 1979: *Optimal Filtering.* Prentice–Hall, 357 pp.

Biswas, K.K., and A.K. Mahalanabis, 1972: An approach to fixed–point smoothing problems. *IEEE Trans. Aerospace and Electronic Systems,* **8**, 676–682.

———, and ———, 1973: Suboptimal algorithms for nonlinear smoothing. *IEEE Trans. Aerospace and Electronic Systems,* **9**, 529–534.

Bryson, A.E., and M. Frazier, 1963: Smoothing for linear and nonlinear dynamic systems. TDR 63–119, *Aero. Sys. Div.*, 353–364, Wrigth–Patterson Air Force Base, Ohio.

Campbell, S.L., and C.D. Meyer, Jr., 1991: *Generalized inverses of linear transformations.* Dover Publications, 272 pp.

Cohn, S.E., 1993: Dynamics of short–term univariate forecast error covariances. *Mon. Wea. Rev.,* **121**, 3123–3149.

———, 1996: An introduction to estimation theory. *J. Meteor. Soc. Japan, submitted.*

———, S.E., 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.

Haltiner, G.J., and R.T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology*, John Wiley & 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.

Leondes, C.T., J.B. Peller, and E.B. Stear, 1970: Nonlinear smoothing theory. *IEEE Trans. Sys. Sci. Cybernetics,* **6**, 63–71.

Meditch, J.S., 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.

Miller, R.N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51**, 1037–1056.

Moore, J.B., 1973: Discrete–time fixed–lag smoothing algorithms. *Automatica,* **9**, 163–173.

Sage, A.P., 1970: Maximum a *posteriori* filtering and smoothing algorithms. *Intl. J. Control,* **11**, 641–658.

———, and W.S. Ewing, 1970: On filtering and smoothing for non–linear state estimation. *Intl. J. Control,* **11**, 1–18.

———, and J.L. Melsa, 1970: *Estimation Theory with Applications to Communications and Control.* McGraw–Hill Book Co., 529 pp.

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, in press.

_____, and S.E. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.,* **122**, 2530–2557.

_____, and S.E. Cohn, 1996: Two reduced resolution filter approaches to data assimilation. *Proc. 9th Brazilian Meteorological Congress*, São José dos Campos, Brazil, Brazil. Meteor. Soc., in press.

_____, N.S. Sivakumaran, and S.E. Cohn, 1996: Some strategies for retrospective data assimilation: Approximate fixed–lag Kalman smoothers. *Proc. 11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meterol. Soc., 238–240.

Verlaan, M, and A. W. Heemink, 1995: Reduced rank square root filters for large scale data assimilation problems. *Proc. Intl. Symp. on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan, WMO, Vol. 1, 247–252.

Willman, W.W., 1969: On the linear smoothing problem. *IEEE Trans. Auto. Control,* **14**, 116–117.