A Partitioned Kalman Filter and Smoother

ICHIRO FUKUMORI
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

(Manuscript received 7 June 2001, in final form 16 October 2001)

ABSTRACT

A new approach is advanced for approximating Kalman filtering and smoothing suitable for oceanic and atmospheric data assimilation. The method solves the larger estimation problem by partitioning it into a series of smaller calculations. Errors with small correlation distances are derived by regional approximations, and errors associated with independent processes are evaluated separately from one another. The overall uncertainty of the model state, as well as the Kalman filter and smoother, is approximated by the sum of the corresponding individual components. The resulting smaller dimensionality of each separate element renders application of Kalman filtering and smoothing to the larger problem much more practical than otherwise. In particular, the approximation makes high-resolution global eddy-resolving data assimilation computationally viable. The approach is described and its efficacy demonstrated using a simple one-dimensional shallow water model.

1. Introduction

Data assimilation is a synthesis of measurements into complete descriptions of a dynamic system using numerical models. Data assimilation has long been employed in numerical weather prediction to optimize initial conditions from which model forecasts are initiated. Lately, data assimilation has also become an increasingly important element in oceanography as a tool to describe and understand ocean circulation.

Mathematically, data assimilation can be identified as a problem in estimation and control. Although statistically optimal methods of assimilation are known, their rigorous implementation has been difficult due to prohibitive computational requirements. Kalman filtering and smoothing is one of such approaches (Ghil et al. 1981). The primary computational difficulty of Kalman filtering stems from the enormous dimension of oceanic and atmospheric models. Kalman filtering and smoothing are in effect recursive least squares estimators that perform averages of model and data, weighted according to their respective error covariances. These errors evolve in time according to model dynamics on the one hand and the assimilation process on the other. Typical state-of-the-art general circulation models have millions of prognostic variables. Consequently, the corresponding model’s state error covariance matrix has over a million squared elements. Evaluation of such a matrix and its storage are far in excess of the capabilities of present-day and near-future state-of-the-art supercomputers.

Various approximations of the Kalman filter have been put forth to simplify its computational implementation. One of the earlier attempts was that of Parish and Cohn (1985). Noting the narrow width of stencils used in general circulation models, Parish and Cohn (1985) approximated the model state error covariance matrix as a banded matrix, limiting, a priori, the number of diagonal bands retained as having nonzero elements, thus reducing the computational as well as storage requirements of Kalman filtering. Although computational requirements were much reduced by retaining fewer elements in the covariance matrix, difficulties arose due to the approximated error covariance matrix losing its positive (semi) definiteness and consequently leading to divergent estimates.

More recently, reduced-state (reduced rank) approximations have been explored that approximate uncertainties of the model state with fewer degrees of freedom than those formally in the models (Fukumori and Malanotte-Rizzoli 1995). The smaller dimension of the approximated space allows for the significant computational savings. Combining models with transformations that approximate the model state error covariance matrices to remain positive semidefinite. For instance, Fukumori (1995) approximated the model state error by its large-scale structure defined on a coarse horizontal grid and Fukumori et al. (1999) employed the gravest vertical dynamic modes to reduce the model’s vertical dimensionality. A set of empirical orthogonal functions (EOFs) was employed by Cane et al. (1996) as a basis for reduction. Pham et al.
The method’s fidelity using a simple one-dimensional
computations taking advantage of each estimate’s assumed
errors of the model state into in-
duced-state approximations that physically and statis-
tically approximate errors of the model state into in-
dependent elements. The estimation amounts to dividing
the estimation problem into a collection of smaller com-
putations taking advantage of each estimate’s assumed
near independence from the others.

Approximating model state error covariances with
limited degrees of freedom is sensible because, for in-
stance, gravest errors of the model state occur at scales
much larger than the model grid resolution, often dom-
inated by a few of the model’s most unstable modes. A
number of studies also suggest that models with even
large formal degrees of freedom often evolve around a
limited number of attractors. Moreover, the amount of
extent observations is insufficient for estimating all for-
mal degrees of freedom (the number of independent
variables) present in most model states and thus a strict
application of estimation theory may not be warranted.

Reduced rank approximations have made application
of Kalman filtering and smoothing feasible to many oce-
anic and atmospheric estimation problems. However, the
fidelity of such approximation is limited to the number
of degrees of freedom in the approximation which in
practice is dictated by available computational resourc-
es. Models have other significant errors than what are
retained in typical state reductions and these errors are
not necessarily dominated by the model’s instabilities.
Furthermore, the dimension of computationally tracta-
cible covariance matrix estimates remains much less than
the total number of extant observations and a significant
(or perhaps dominant) fraction of the data’s available
information content often remains unutilized. For in-
cense, meso-scale variabilities measured by satellite alti-
timetry are not resolved by many global ocean data
assimilation experiments (e.g., Fukumori et al. 1999).

The present study extends the reduced-state approx-
imation so as to resolve a larger number of degrees of
freedom commensurate with those of the observations,
yet with minimal increase in computational overhead.
The approach, termed partitioned Kalman filter and par-
titioned smoother, consists of combining a series of re-
duced-state approximations that physically and statis-
tically approximate errors of the model state into in-
dependent elements. The estimation amounts to dividing
the estimation problem into a collection of smaller com-
putations taking advantage of each estimate’s assumed
near independence from the others.

The partitioned Kalman filter and partitioned smooth-
er are described in section 2. Section 3 demonstrates
the method’s fidelity using a simple one-dimensional
shallow water model. We conclude with a summary and
discussion in section 4.

2. The partitioned Kalman filter and smoother

The state error covariance matrix defines the Kalman
filter and smoother and is the fundamental element of
the estimation procedure. The gains of the Kalman filter
(K) and the fixed-interval Rauch–Tung–Striebel (RTS)
smoother (S) can be written as (Gelb 1974),

\[ K = PH^T R^{-1} \]  

\[ S = PA^T P(-)^{-1} \]  

where \( P(-) \) and \( P \) denote the model state error co-
variance matrices immediately before and after the se-
quential Kalman filter assimilation. Here, \( P(-) \) and \( P \)
are also often referred to as the forecast and analysis
error covariance matrices, respectively. The \( H \) is the
“observation” matrix that provides the expected value
of the observations \( Hx \) given the model state vector \( x \).
Here, \( R \) is the observation plus representation error co-
variance matrix and \( A \) is the model state transition ma-
trix. We assume a linear model for simplicity. The filter
and smoother and the approximations described below
can be extended to nonlinear models using suitable ap-
proximations.

Error evaluation in Kalman filtering (Riccati equa-
tion) consists of matrix multiplications. Computational
requirements associated with such calculation are pro-
portional to the cube of the state dimension. The re-
duced-state Kalman filter and smoother approximate the
state error covariances with limited degrees of freedom
that capture the errors’ gravest structures such as un-
certainties associated with large basin-scale processes.
Such approximation greatly reduces the dimensionality
of the problem and consequently makes application
of Kalman filtering more practical. However, the approx-
imation may contain elements that may not be signifi-
cant while failing to resolve other components of model
uncertainties.

For instance, Fig. 1 shows an example of some ele-
ments of the correlation matrix associated with the esti-
mation error of a reduced-state Kalman filter used by
Fukumori et al. (1999). The example approximates the
wind-driven errors of a global ocean general circulation
model in terms of amplitudes of barotropic and first
baroclinic modes on a coarse horizontal grid (10° by 5°
zonai and meridional grid spacing). The structure of
correlation is generally confined locally and there is only
weak correlation between uncertainties in barotropic and
baroclinic modes. On the one hand the approximation
resolves the large-scale errors, but on the other hand it
also retains correlations that are not significant. More-
over, the approximation does not capture errors with
scales smaller than the coarse horizontal grid (null space
of the state reduction), which may be more significant
than the weak correlations that are retained.
A significant computational savings can also be achieved if independent error components are evaluated separately from one another without limiting the overall range space (i.e., the dimension or the number of degrees of freedom) of the estimation as in the reduced-state approximation described above. For instance, if the total state dimension (n) were partitioned into l equal elements and if errors of each element were solved independently of others, each element will require \((nl)^3\) operations. Then when combined with the other \(l - 1\) elements in the partition, the total operation count amounts to \((nl)^3 \times l = n^4l^2\), resulting in savings by a factor of \(1/l^2\). Likewise, storage requirements that are proportional to the square of the state dimension would be reduced by a factor of \(1/l\) by such approximation. For instance, if the global ocean or atmosphere is divided into 10 equal segments in each of the zonal, meridional, and vertical dimensions, and if errors of each segment are assumed to be independent of other segments and are solved separately from one another, such approximation will lead to a computational savings by a factor of 1 000 000 and storage by 1000. Additional state reductions could be applied to separate segments, further reducing the computational requirements of individual error evaluations.

For the example in Fig. 1, separation could be achieved by evaluating the barotropic component separately from the baroclinic mode, and/or evaluating each error in a suitable local area instead of over the entire model domain. Furthermore, an additional regional approximation with scales smaller than those retained in the large-scale approximation would allow estimation at those smaller scales without incurring excessive computational overhead. Errors associated with uncertainties in diabatic forcing might also be estimated separately using a different approximation than those due to errors in wind forcing.

The Kalman filter and smoother based on such partitioning can be described mathematically in terms of a series of model state transformations, and will be termed the partitioned Kalman filter (PKF) and partitioned smoother (PS), respectively. The model state, \(x\), is approximated by a sum of \(l\) independent elements (or processes), \(x_1', x_2', \ldots, x_{l}'\), each with much fewer formal degrees of freedom than the model state \(x\) itself; namely, \[x = B_1x_1' + B_2x_2' + \cdots + B_{l}x_{l}',\] (3)
where $\mathbf{B}$ denotes a transformation approximating particular elements of the model state $\mathbf{x}$ by $\mathbf{x}'$. Primes, henceforth, denote variables in the approximation, while subscripts identify particular elements of the partition. Each partition element can be regarded as a separate reduced-state approximation of the model state. For instance, in the example of Fukumori et al. (1999), $\mathbf{x}_1$ and $\mathbf{x}_2$ may consist of amplitudes of the barotropic and baroclinic modes on the coarse grid, respectively, with $\mathbf{B}_1$ and $\mathbf{B}_2$ being corresponding transformations relating these reduced-state vectors ($\mathbf{x}_1'$, $\mathbf{x}_2'$) to the model variables ($\mathbf{x}$). Alternatively, $\mathbf{x}_i'$ could represent other modal amplitudes or different approximations on a regional basis.

Given the linear relationship in (3), the uncertainty of the model state $\Delta \mathbf{x}$ may be approximated by the sum of the errors of each element,

$$\delta \mathbf{x} = \mathbf{B}_1 \delta \mathbf{x}_1' + \mathbf{B}_2 \delta \mathbf{x}_2' + \cdots + \mathbf{B}_i \delta \mathbf{x}_i'$$  \hspace{1cm} (4)

where $\Delta \mathbf{x}_i'$ denotes uncertainty of the $i$th partition element. One can recognize Eq. (4) as a single-stage reduced-state approximation,

$$\delta \mathbf{x} = \begin{pmatrix} \delta \mathbf{x}_1' \\ \delta \mathbf{x}_2' \\ \vdots \\ \delta \mathbf{x}_i' \end{pmatrix} = \mathbf{B} \delta \mathbf{x}'$$  \hspace{1cm} (5)

concatenating different elements $\mathbf{B}_i$ and $\mathbf{x}_i'$ as $\mathbf{B}$ and $\mathbf{x}'$, respectively. Uncertainties of the model state $\mathbf{P}$ and its inverse $\mathbf{P}^{-1}$ can be approximated as (Fukumori 1995),

$$\mathbf{P} = \langle \delta \mathbf{x} \delta \mathbf{x}^T \rangle = \mathbf{B} (\delta \mathbf{x}' \delta \mathbf{x}'^T) \mathbf{B}^T = \mathbf{B} \mathbf{P}' \mathbf{B}^T$$  \hspace{1cm} (6)

$$\mathbf{P}^{-1} = \mathbf{B}^{\dagger T} \mathbf{P}^{\dagger} \mathbf{B}^\ast$$  \hspace{1cm} (7)

where $\mathbf{P}' = \langle \delta \mathbf{x}' \delta \mathbf{x}'^T \rangle$ is the state error covariance matrix of the reduced-state $\mathbf{x}'$ and $\mathbf{B}^\ast$ is the pseudoinverse of transformation $\mathbf{B}$. The $\langle \cdot \rangle$ denotes statistical expectation. In terms of approximation (3), Eqs. (7) and (8) may be written as,

$$\mathbf{P} = (\mathbf{B}_1 \mathbf{B}_2 \ldots \mathbf{B}_i) \begin{pmatrix} \delta \mathbf{x}_1' \\ \delta \mathbf{x}_2' \\ \vdots \\ \delta \mathbf{x}_i' \end{pmatrix} \begin{pmatrix} \mathbf{B}_1^T \\ \mathbf{B}_2^T \\ \vdots \\ \mathbf{B}_i^T \end{pmatrix}$$  \hspace{1cm} (9)

$$\mathbf{P}^{-1} \approx (\mathbf{B}_1^{\dagger T} \mathbf{B}_2^{\dagger T} \ldots \mathbf{B}_i^{\dagger T}) \begin{pmatrix} \delta \mathbf{x}_1' \\ \delta \mathbf{x}_2' \\ \vdots \\ \delta \mathbf{x}_i' \end{pmatrix} (\mathbf{B}_1^{\dagger T} \mathbf{B}_2^{\dagger T} \ldots \mathbf{B}_i^{\dagger T})^{-1} \begin{pmatrix} \mathbf{B}_1^T \\ \mathbf{B}_2^T \\ \vdots \\ \mathbf{B}_i^T \end{pmatrix}$$  \hspace{1cm} (10)

where $\mathbf{B}_i^\ast$ is the block of $\mathbf{B}^\ast$ corresponding to the $i$th element of approximation (3). Assuming statistical independence among different components in (4), namely,

$$\mathbf{P} = \mathbf{B}_i \delta \mathbf{x}_i' \delta \mathbf{x}_i'^T + \mathbf{B}_j \delta \mathbf{x}_j' \delta \mathbf{x}_j'^T \cdots + \mathbf{B}_l \delta \mathbf{x}_l' \delta \mathbf{x}_l'^T$$

$$\beta = \mathbf{B}_i \delta \mathbf{x}_i' \delta \mathbf{x}_i'^T$$

$$\mathbf{P} \approx \mathbf{B}_i \mathbf{P}^{\dagger} \mathbf{B}_i^T + \mathbf{B}_j \mathbf{P}^{\dagger} \mathbf{B}_j^T \cdots + \mathbf{B}_l \mathbf{P}^{\dagger} \mathbf{B}_l^T$$  \hspace{1cm} (11)

Eq. (9) may be approximated as

$$\mathbf{P} = \mathbf{B}_i \delta \mathbf{x}_i' \delta \mathbf{x}_i'^T + \mathbf{B}_j \delta \mathbf{x}_j' \delta \mathbf{x}_j'^T \cdots + \mathbf{B}_l \delta \mathbf{x}_l' \delta \mathbf{x}_l'^T$$

$$\mathbf{P} = \mathbf{B}_i \mathbf{P}^{\dagger} \mathbf{B}_i^T + \mathbf{B}_j \mathbf{P}^{\dagger} \mathbf{B}_j^T \cdots + \mathbf{B}_l \mathbf{P}^{\dagger} \mathbf{B}_l^T$$  \hspace{1cm} (12)

$$\approx \mathbf{B}_i \mathbf{P}^{\dagger} \mathbf{B}_i^T + \mathbf{B}_j \mathbf{P}^{\dagger} \mathbf{B}_j^T \cdots + \mathbf{B}_l \mathbf{P}^{\dagger} \mathbf{B}_l^T$$  \hspace{1cm} (13)

where

$$\mathbf{P}^{\dagger} = \langle \delta \mathbf{x}_i' \delta \mathbf{x}_i'^T \rangle$$  \hspace{1cm} (14)

is the state error covariance matrix of the $i$th partition element in (3). Likewise, Eq. (10) may be expanded as,

$$\mathbf{P}^{-1} = \mathbf{B}_i^{\dagger T} \mathbf{P}_i^{\dagger} \mathbf{B}_i^\ast + \mathbf{B}_j^{\dagger T} \mathbf{P}_j^{\dagger} \mathbf{B}_j^\ast \cdots + \mathbf{B}_l^{\dagger T} \mathbf{P}_l^{\dagger} \mathbf{B}_l^\ast$$

$$\approx \sum_i \mathbf{B}_i \mathbf{P}_i^{\dagger} \mathbf{B}_i^\ast$$  \hspace{1cm} (15)

Given the approximation of the state error covariance matrix [Eq. (13)], the Kalman filter gain [Eq. (1)] can be approximated as,

$$\mathbf{K} = \mathbf{P}^{\dagger T} \mathbf{R}^{-1} \approx \left( \sum_i \mathbf{B}_i \mathbf{P}_i^{\dagger} \mathbf{B}_i^\ast \right) \mathbf{H} \mathbf{R}^{-1}$$

$$= \sum_i \mathbf{B}_i \mathbf{K}_i$$  \hspace{1cm} (16)

where

$$\mathbf{H}_i = \mathbf{H} \mathbf{B}_i$$

$$\mathbf{K}_i = \mathbf{P}_i^{\dagger} \mathbf{R}^{-1}$$  \hspace{1cm} (17)

are the observation matrix and the Kalman gain for the approximation’s $i$th element [Eq. (3)], respectively. Similarly, the smoother gain ($\mathbf{S}$) can be approximated using (13) and (15) as,

$$\mathbf{S} = \mathbf{P}^{\dagger T} \mathbf{P}^{\dagger} (\mathbf{P}^{\dagger T} \mathbf{P})^{-1}$$

$$= \left( \sum_i \mathbf{B}_i \mathbf{P}_i^{\dagger} \mathbf{B}_i^\ast \right) \mathbf{A} \left[ \sum_i \mathbf{B}_i^{\dagger T} \mathbf{P}_i^{\dagger} \mathbf{P}^{\dagger} (\mathbf{P}^{\dagger T} \mathbf{P})^{-1} \mathbf{B}_i^\ast \right]$$

$$\approx \sum_i \mathbf{B}_i \mathbf{S}_i^\ast \mathbf{B}_i^\ast$$  \hspace{1cm} (19)

Equation (19) can be further approximated by assuming that variations within a given partition remain independent from the space spanned by other elements in (3), namely,

$$\mathbf{B}_i^\ast \mathbf{A} \mathbf{B}_j^\ast = 0 \quad (\text{when } i \neq j)$$  \hspace{1cm} (20)

Then Eq. (19) can be further expanded as,

$$\mathbf{S} \approx \sum_i \mathbf{B}_i \mathbf{P}_i^{\dagger} \mathbf{A} \mathbf{P}_i^{\dagger} \mathbf{P}^{\dagger} (\mathbf{P}^{\dagger T} \mathbf{P})^{-1} \mathbf{B}_i^\ast \approx \sum_i \mathbf{B}_i \mathbf{S}_i^\ast \mathbf{B}_i^\ast$$  \hspace{1cm} (21)

Here,

$$\mathbf{A}_i^\ast = \mathbf{B}_i^\ast \mathbf{A} \mathbf{B}_i^\ast$$  \hspace{1cm} (22)
where

\[ \mathbf{S}_i' = \mathbf{P}_i'\mathbf{A}_i'\mathbf{P}_i'(-1)^{-1} \]  \hspace{1cm} (23)

is the filter increment for the \( i \)th element of (3). Equations (27) and (28) show that the PKF approximates the Kalman filter increments by the sum of the equivalent increments of individual elements in (3).

Likewise, the smoother increment can also be evaluated element by element in the smaller partitioned space. The RTS smoothed state at time \( t \), \( x(t, +) \), where the argument “+” denotes a smoothed estimate, is derived backwards in time, recursively, by

\[ x(t, +) = x(t) + S[x(t + 1, +) - x(t + 1, -)] \]  \hspace{1cm} (29)

(Gelb 1974). The second term in (29) provides the smoother increment,

\[ \delta_s(t) = x(t, +) - x(t). \]  \hspace{1cm} (30)

Equation (29) can be rewritten, using (25) and (30), as

\[ \delta_s(t) = S[x(t + 1, +) - x(t + 1, -)] \]

\[ = S[x(t + 1) - x(t + 1, -)] \]

\[ + [x(t + 1) - x(t + 1, -)] \]

\[ = S[\delta_s(t + 1) + \delta_s(t + 1)], \]  \hspace{1cm} (31)

which reveals that the smoother increment is obtained by an integration of the model \([\mathbf{S}, \mathbf{E}, \mathbf{Q}]\) backwards in time forced by the Kalman filter increments, \( \delta_k \) (26). Substituting (21) and (27), (31) can be expanded as

\[ \delta_s(t) = \sum \mathbf{B} \mathbf{S}_i' \mathbf{B}_i^* [\delta_s(t + 1) + \sum \mathbf{B}_i \delta_k(t + 1)] \]

\[ = \sum \mathbf{B}_i \mathbf{S}_i' [\mathbf{B}_i^* \delta_s(t + 1) + \delta_k(t + 1)]. \]  \hspace{1cm} (32)

Equation (32) can be written as

\[ \delta_s(t) = \sum \mathbf{B}_i \delta_s(t). \]  \hspace{1cm} (33)

where

\[ \delta_s(t) = \mathbf{S}_i' [\mathbf{B}_i^* \delta_s(t + 1) + \delta_k(t + 1)] \]  \hspace{1cm} (34)

\[ \delta_k(t) = \mathbf{S}_i' [\delta_k(t + 1) + \delta_k(t + 1)] \]  \hspace{1cm} (35)

using (33) at time \( t + 1 \). As with the Kalman filter, Eqs. (33) and (35) show that the PS increments can be evaluated separately in their respective partitioned space and combined linearly together to form the total smoother increment. Note that smoothing given by Eqs. (23) and (35) is defined entirely within the reduced space, which allows for a considerable reduction in computational requirements compared to implementing (19).

The inverse transformation \( \mathbf{B}_i^* \) is employed in the implementation \([\text{Eqs. (18), (27), and (28) for PKF and Eqs. (23), (33), and (35) for PS}]\) only in deriving the reduced-state transition matrix \( \mathbf{A}_i' \), which in turn is used to derive \( \mathbf{P}_i' \) and \( \mathbf{S}_i' \). Given the assumed independence among the approximation’s different elements in (3), \( \mathbf{B}_i^* \) can be approximated individually by the pseudoinverse of the corresponding transformation \( \mathbf{B}_i^* \).
3. Example: One-dimensional shallow water model

In this section, we examine the fidelity of the partitioned Kalman filter and partitioned smoother in a so-called twin experiment; namely, pseudo-observations are sampled from a model simulation ("true ocean") and assimilated into a different simulation ("false ocean") to examine how well the former's states can be estimated. The model employed is a nonrotating, one-dimensional 1\frac{1}{2}-layer nonlinear shallow water model driven by random winds. The momentum and continuity equations are given by, using standard notation,

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = -g' \frac{\partial \eta}{\partial x} + \nu \frac{\partial^2 u}{\partial x^2} + \tau - \kappa u \eta, \quad (36)$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x} [u(H + \eta)] = -\kappa \eta \eta, \quad (37)$$

where $u$, $\eta$, and $H$ are velocity, layer thickness perturbation, and model depth at rest, respectively. The model is configured with parameters roughly corresponding to simulating the first baroclinic mode of the equatorial Pacific Ocean (e.g., Cane 1984). Layer thickness $H$ is 281 m and the reduced gravity $g'$ is 0.0302 m s$^{-2}$. The model domain is 38 400 km long (close to the circumference of the earth) with cyclical boundary conditions, and is discretized in a staggered grid of 256 equally spaced grid points. The model has a spatial grid spacing of 150 km and uses a 1-h leapfrog time stepping scheme. The model employs Laplacian friction ($\nu = 2 \times 10^4$ m$^2$ s$^{-1}$). A weak Newtonian damping is added to both equations for computational stability ($\kappa = 30$ days and $\kappa = 60$ days, respectively). Wind stress, $\tau$, consists of two independent autocorrelated Gaussian random processes (see appendix) with covariance

$$\langle \tau(x_1) \tau(x_2) \rangle = a_1 e^{-(x_1-x_2)^2/2w_1^2} + a_2 e^{-(x_1-x_2)^2/2w_2^2}. \quad (38)$$

One has a large spatial correlation distance ($w_1 = 12000$ km) while the other has a smaller correlation distance ($w_2 = 1200$ km), with variances $a_1 = 3 \times 10^{-3}$ and $a_2 = 7 \times 10^{-3}$ (N$^2$ m$^{-4}$), respectively. These winds are varied randomly every 10 h. The random winds are otherwise time invariant during the 10-h intervals.

Figure 2 describes the space–time evolution of the model sea level and velocity for a 250-day period starting from rest (true ocean). Waves are generated randomly and propagate over a significant distance.
throughout the model domain before being dissipated. The black curves in Fig. 3 show the model sea level and velocity time series at \( x = 31\,950 \) km and their respective spatial structures at day 165. Note the presence of large- and small-scale variabilities in the spatial structures (lower panels of Fig. 3) corresponding to the two separate forcings [Eq. (38)].

This simulation constitutes the “truth” from which pseudo-observations are obtained. Specifically, sea level is sampled every 10 h on a nonuniform grid denoted by lines shown in Fig. 2, simulating time series measurements in a region well sampled and another that is not. The samples are made at 4200-km intervals across the entire domain except between \( x = 6900 \) km and \( x = 12\,900 \) km where the samples are made at 600-km resolution. The coarser pseudo-observation grid allows adequate sampling of the large-scale signal. However, the smaller-scale variations are measured but not resolved by the pseudo-observations (are aliased) except in the higher-resolution sampling region (Fig. 3). Random errors are added to the sampled sea level to simulate data constraint errors. The added noise variance is 2.8 cm\(^2\) and is equal to the variance of the “true” sea level itself (Fig. 2).

The sampled sea levels are assimilated into the same model but one that is forced with additional winds given by another independent set of random noise of the form Eq. (38). The control run with different forcing from the truth, and with no assimilation, constitutes the false ocean. The additional wind has the same covariance as the true forcing and simulates sources of errors for the model state (process noise). Examples of this false ocean are shown as gray curves in Fig. 3.

The partitioning of the Kalman filter and smoother is achieved in this example by a set of “cells” described in Fig. 4. A global reduced-state filter and smoother is defined to resolve model state errors due to the large-scale component of Eq. (38). This global filter/smooother is based on a global cell defined by a set of coarse grid points spanning the entire model domain (Fig. 4a). The coarse global grid consists of eight points with a 4800-km grid resolution (32 model-grid-points wide). Transformation \( B \) in Eq. (3) is defined by interpolation of the model state represented on this coarse grid to those on the model grid. Such coarse grids-based global reduced-state filters have been described and used in the past (e.g., Fukumori 1995). However, such coarse grids do not permit resolving scales smaller than the coarse grid resolution, such as variabilities due to the more energetic smaller scales of Eq. (38).

A series of local cells is additionally defined (Fig. 4b) with finer resolution than the global cell so as to
resolve errors due to the smaller-scale component of Eq. (38) and to take advantage of the information content of the high-resolution observations (Fig. 2). The local cells consist of 14 fine-resolution grid points spaced at 600-km intervals. Each local fine cell extends 8400-km wide and is centered around each separate coarse grid point defining the global cell. The local cells total eight (Fig. 4a) and are defined so as to best resolve the small-scale errors across the entire model domain. In particular, the local cells are chosen to overlap neighboring cells as shown in Fig. 4b to minimize possible inaccuracies in the assimilation arising from cell boundaries as described later.

Process noise of the individual cells is defined by approximating the overall model process noise into separate partitions. In terms of wind error (source of process noise), the overall wind error covariance matrix, $W$, is partitioned into those of separate cells, such that,

$$ W = W_s + W_1 + W_2 + \cdots + W_n, $$

where $W_s$ and $W_i$ ($i = 1, 8$) are wind error covariance matrices for the global coarse and local fine cells, respectively. Here, $W_s$ and the sum of $W_i$ over all local cells are given, respectively, by the larger- and smaller-scale components of (38). The smaller-scale component of (38) is further split among different local cells using a trapezoidal tapering function, $d_i(x)$ (Fig. 4). Namely, the wind error covariance matrix of local cell $i$ is defined by,

$$ W_i = DW_{\text{small}}D_i^T, $$

where $D_i$ is a diagonal matrix with its diagonal elements given by $d_i$ and $W_{\text{small}}$ is the covariance matrix of the smaller-scale component of (38). Examples of the tapering function $d_i$ (its square) are shown in Fig. 4. In particular, $d_i$ for the local fine cell $i$ centered at $x = 23400$ km in Fig. 4b, is defined by,

$$ d_i(x) = \begin{cases} 
0 & \text{for } x < 20400 \text{ km} \\
\sqrt{(x - 20400)/1200} & \text{for } 20400 \leq x < 21600 \text{ km} \\
1 & \text{for } 21600 \leq x < 25200 \text{ km} \\
\sqrt{(26400 - x)/1200} & \text{for } 25200 \leq x < 26400 \text{ km} \\
0 & \text{for } 26400 \leq x.
\end{cases} $$

Tapering the assumed wind error to zero near the edge of the local partitions [Eq. (41)] minimizes inaccuracies in evaluating the local model state error covariances due to the open cell boundaries. To satisfy Eq. (39) with such tapering, local cells are defined to overlap neighboring cells. Namely, the tapering function for neighboring cells are identical except shifted in $x$ by the distance between centers of the local cells as shown by the gray trapezoidal curves in Fig. 4b. Within the overlapping regions of the local fine cells, the small-scale wind error is split between neighboring cells. The tapering functions thus defined satisfy

$$ \sum_i d_i^2(x) = 1, $$

where the sum is taken over all local cell partitions, assuring that the sum of small-scale wind error variances projected to each cell by (40) equals that of $W_{\text{small}}$. The process noise covariance matrix of each partition is obtained by projecting the individual wind error covariance matrices onto the corresponding partitioned state according to the model dynamics and the mapping operators $B_i^T$.

Objective mapping (Bretherton et al. 1976) is employed as the transformation operator $B_i$ [Eq. (3)] for both local and global cells. The mapping assumes no sampling error and employs a Gaussian covariance function with correlation distance equaling the grid resolution of each cell (partition). Figure 5 shows an example expanding and reconstructing the “false” model state shown in Fig. 3 using mapping operators of the cellular reduced state. Note the global cell resolving the large-scale variability while the additional series of local cells are needed to capture the smaller scales.

The assimilation is performed after spinup from day 83 (2000th time step from rest), and employs the time-asymptotic approximation (Fukumori et al. 1993) in addition to the state reduction so as to further reduce the computational complexity of Kalman filtering and smoothing. The asymptotic Kalman filter/smoothers approximates the state error covariance matrix with its time-invariant (or limit-cycle) asymptotic limit (i.e., solution to the Riccati matrix equation), thus eliminating the individual temporal integration of the error covariance matrix. The approximation’s reduced-state transition matrix and other matrices of the system used in deriving the filters and smoothers are computed numerically by a Green’s function method (Fukumori and Malanotte-Rizzoli 1995).

To assess the fidelity of the PKF and PS, results are compared with the skill of estimates made by other filters and smoothers. Comparisons are made with a single-stage reduced-state filter and smoother consisting of only the global coarse cell (Fig. 4a) and another filter and smoother pair that consists of a fine grid with a
The resolution of the local cell (Fig. 4b) extending over the entire model domain. These two additional filters and smoothers will be referred to as global coarse and global fine, respectively. The two represent the lower and upper bounds of the possible skill of the PKF and PS.

The skills of the various assimilations are compared in Table 1 and in Figs. 6 and 7. The skills are measured in terms of the variance of their errors, that is, differences between the true solution and the various estimates:

$$\sum_{i=1}^{N} (v_{\text{true}} - v_{\text{estimate}})^2 / N.$$  (43)

Here, $v$ is either sea level or velocity, and $N$ is the total number of elements in the summation which is computed either in space and/or time. In particular, Figs. 6 and 7 compare such skill as functions of time and space, respectively, whereas Table 1 summarizes the skill averaged over both space and time.

The solutions obtained by the smoothers generally have smaller errors than those of the corresponding filtered estimates, because of the former utilizing more observations than the latter, consistent with theoretical expectations. In comparison, the smoother estimates often have larger model-data residuals (last row of Table 1) than filtered estimates owing to data errors and the latter overfitting the (pseudo) observations as it is not constrained by formally future measurements.

Table 1 also provides a measure of the assimilations’ skills as a function of length scales. A simple boxcar running average with a width equaling the global coarse grid (4800 km) is employed as a simple means to evaluate the states’ large-scale errors. The residuals of the boxcar average constitute estimates of the smaller scale.

The filter and smoother of the global coarse approximation successfully resolve the large-scale errors. However, as expected, errors of the smaller scale are comparable to those of the false model, indicating a failure of the global coarse filter/smoother in correcting the model at scales smaller than the approximation’s spatial resolution (Table 1). Consequently the total skill of the global coarse approximation is limited when compared with the less approximated global fine filter and smoother (Figs. 6 and 7).

In comparison, skills of the PKF and PS are nearly indistinguishable, respectively, from the filter and
Table 1. Error variance of sea level and velocity averaged over space and time of different model estimates. Variance of the model data sea level differences is also given for reference (last row). Large-scale errors are based on a boxcar spatial average of the solution with a width of 4800 km. The small-scale errors are defined as the difference between the total error and this large-scale component. Values for the assimilation denote errors of the filtered estimates while those of the smoothed estimates are given in parentheses. Units are in cm$^2$ and cm$^2$ s$^{-2}$ for sea level and velocity, respectively.

<table>
<thead>
<tr>
<th>Scale</th>
<th>&quot;False&quot; ocean</th>
<th>Global fine</th>
<th>Partitioned</th>
<th>Global coarse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sea level</td>
<td>3.0</td>
<td>1.0 (0.7)</td>
<td>1.2 (1.0)</td>
<td>2.4 (2.3)</td>
</tr>
<tr>
<td>Large</td>
<td>0.9</td>
<td>0.2 (0.1)</td>
<td>0.3 (0.2)</td>
<td>0.5 (0.4)</td>
</tr>
<tr>
<td>Small</td>
<td>1.8</td>
<td>0.7 (0.5)</td>
<td>0.8 (0.7)</td>
<td>1.8 (1.8)</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>26 (16)</td>
<td>28 (20)</td>
<td>40 (36)</td>
</tr>
<tr>
<td>Velocity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>18</td>
<td>15 (9)</td>
<td>15 (11)</td>
<td>17 (14)</td>
</tr>
<tr>
<td>Small</td>
<td>20</td>
<td>9 (6)</td>
<td>10 (8)</td>
<td>20 (20)</td>
</tr>
<tr>
<td>Model data difference</td>
<td>6.0</td>
<td>2.7 (2.8)</td>
<td>2.3 (2.7)</td>
<td>4.8 (4.8)</td>
</tr>
</tbody>
</table>

smoother of the global fine approximation. The error variances are similar at different scales between the two approximations (Table 1), and their dependence on time and space are nearly the same (Figs. 6 and 7). For both the partitioned and global fine approximations, errors are generally much smaller near observation points, particularly in the high-resolution observation region between $x = 6900$ km to $x = 12900$ km (Fig. 7).

In spite of the assimilated pseudo-observations being limited to sea level, velocity errors are similarly reduced as those of sea level. In the absence of geostrophy (the model is nonrotating), velocity estimates from sea level are nontrivial. In particular, velocities cannot be uniquely determined from instantaneous sea level measurements. The velocity estimates here owe their improvements to the model’s time dependency and its successful inversion.

The error variances of the PKF and PS are comparable.
to their respective formal uncertainty estimates, demonstrating the consistency of the assimilations (Figs. 6 and 7). Moreover, the formal error estimates of the partitioned approximation are practically the same as those of the global fine approximation. Small-scale differences in the formal error estimates, such as those of the partitioned estimates’ additional small-scale local error minima and maxima in the sparse observation region (Fig. 7b), are due to inaccuracies associated with the cells’ open boundaries.

Finally, Fig. 8 shows examples of the PS estimate corresponding to those of Fig. 3. The PS estimates are visibly closer to the true solutions than the false solutions are (Fig. 3) and do not show any evidence of cell boundaries. The true solutions generally lie within the uncertainty limits of the PS estimates, illustrating the estimates’ first-order consistency and validity.

4. Discussion

The main difficulty in implementing oceanographic and atmospheric data assimilation lies in the enormous computational requirements that stem from the large number of formal degrees of freedom in the respective dynamic systems. Reduced-state Kalman filters and smoothers have made feasible application of sequential estimation theory to such problems by limiting the number of estimated degrees of freedom within a given model (Fukumori et al. 1999; Gourdeau et al. 2000; Balabera-Poy et al. 2001). However, the fidelity of such an approach has been restricted by the very nature of the approximation. Namely, the limitation in the estimated degrees of freedom leaves unutilized a significant fraction of the information content available in observations. For instance, advances in modeling and computational capabilities have enabled global eddy-permitting simulations (Maltrud et al. 1998), and sea level observations from multiple altimetric satellites now allow mapping eddy variabilities over the global ocean (Ducet et al. 2000). However, global eddy-resolving data assimilation has not yet been achieved.

The partitioned Kalman filter and partitioned smoother introduced in this study significantly increase the possible number of degrees of freedom that can be estimated and thereby enable utilization of more of the observations’ information content. The PKF and PS partition data assimilation into a series of smaller estimation problems by approximating the model errors into independent elements and by evaluating the individual errors separately. The partitioning is in essence a divide and conquer approach that solves the larger assimilation problem by dividing it into more manageable smaller calculations. Because computational requirements are proportional to the cube of the problem size, the indi-
Fig. 8. Examples of the PS estimates. Variables correspond to those in Fig. 3. The shaded area denotes the standard errors of the PS estimates (dashed curve). The solid curve denotes the “true” solution.

Individual components’ smaller dimension enables estimation of the much larger overall state with dramatically less computational overhead than otherwise.

Many processes in the ocean and atmosphere can reasonably be approximated as being independent of others. For instance, much of the smaller-scale errors are likely uncorrelated over a large distance, and thus can be estimated by regional approximations. It is unlikely that errors associated with mesoscale eddies in the Kuroshio region are correlated with those in the Gulf Stream area, and vice versa. Likewise, to first approximation, errors in placing the location of thunderstorms in Kansas may have little to do with inaccuracies in simulating squall lines across Africa. Because of differences in ocean physics, errors due to inaccuracies in wind and thermodynamic forcing of the ocean may be approximated as being independent of each other and thus could be estimated separately. Baroclinic and barotropic modes have large differences in timescales and therefore could be treated independently as suggested in Fig. 1. Moreover, linear theory in uniform media predicts independence among separate vertical dynamic modes, which would allow estimating separate modes independently of others. Separation could also be achieved across other timescales. For instance, to first approximation, errors associated with seasonal-to-interannual processes may be regarded as being independent of errors on daily timescales, and thus could be approximated separately from each other.

That such separation is sensible is evidenced by the success of many simplified models in simulating various aspects of the ocean and the atmosphere. Coarse resolution global ocean models have been used to study large-scale ocean circulation as well as to simulate climate change without resolving mesoscale eddies (Manabe et al. 2001). Basin-scale ocean models have been employed to study ocean circulation of particular basins in finer detail than otherwise possible (Smith et al. 2000). Smaller regional models are employed to analyze circulation in bays and coastal regions (Haidvogel et al. 1991). Reduced-gravity shallow water models of the tropical Pacific Ocean have been successful in simulating some of the salient features of El Niño and the Southern Oscillation (Zebiak and Cane 1987).

The success of the approaches used in traditional data assimilation, such as with 3D-variations and optimal interpolation methods (e.g., Ghil and Malanotte-Rizzoli 1991), lends support to the effectiveness of local partitioning of the estimation. In realistic applications in the atmosphere and ocean, these traditional methods are often simplified for computational efficiency by implementing them locally either by prescribing background state error statistics to be of small scales and/or utilizing data only within some fixed local radii of the analysis.
points. Local truncations have also been employed in ensemble Kalman filtering (Houtekamer and Mitchell 1998).

The nature of the partitioning approach introduced in this study allows flexible modeling of errors of different scales and diverse processes simultaneously in an objective hierarchical framework. The method provides a means to perform data assimilation at resolutions and scope commensurate with the information content of available observations and with known or assumed error statistics without incurring the debilitating computational requirements of direct applications of estimation theory. In effect, the PKF and PS provides a solution to the fundamental computational problem associated with the enormous dimension of oceanographic and atmospheric data assimilation. In particular, the approach makes state estimation computationally viable with global eddy-resolving models of the ocean and state-of-the-art high-resolution models of the atmosphere from their smallest grid resolutions to the size of their entire model domain (planetary scales).

The introduction of partitioning shifts the principal issue in data assimilation from one of computational approximation to that of physical approximation. The fidelity of the PKF and PS hinges on identifying an effective partitioning of error processes, in particular the individual transformation operators $\mathbf{B}$, that define the approximation [Eq. (3)]. The statistical independence among the partitioned elements is a particular approximation rather than an assertion, and the effectiveness of the partitioning depends on the goodness of this assumption. The hierarchical cellular approximation utilized in section 3 provides an example of dividing model uncertainties into separate scales (large scale vs small scale) and different regions (overlapping local cells). Issues concerning how best to partition a given model, such as how to define and combine local approximations, require careful physical as well as numerical and statistical consideration. Local cells could be strategically defined near observation points and not in data-void regions. Partitioning can also be achieved in terms of different physics and dynamics in addition to geographic locations.

Identifying efficient and effective representations of the various physical processes underlying uncertainties of the model state is a central outstanding problem, and should be pursued vigorously in future studies. The PKF and PS extend and build upon the reduced-state approximation. In particular, each of the individual components in the partitioning [Eq. (3)] constitutes a separate reduced-state approximation of the larger estimation problem. Consequently, each element in the partitioned state, as in reduced-state approximations, must define dynamically closed systems such that any perturbation in the approximated state remains approximately within the range space of the given transformation (Fukumori and Malanotte-Rizzoli 1995). Application of the PKF and PS to oceanographic state estimation is presently under way and will be reported elsewhere.

Acknowledgments. Thoughtful comments by Chris Snyder and another anonymous reviewer are gratefully acknowledged. This study is a contribution of the Consortium for Estimating the Circulation and Climate of the Ocean (ECCO) funded by the National Oceanographic Partnership Program. Support was also provided by the TOPEX/Poseidon and the Jason-1 Projects. This research was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

APPENDIX

Generating correlated random numbers

Correlated normal random numbers with a specified mean and covariance can be generated by a transformation of uncorrelated random numbers. For instance, let $\mathbf{a}$ be a column vector of uncorrelated normal random numbers with zero mean and unit variance; namely,

$$\langle \mathbf{a} \rangle = 0 \quad (A1)$$
$$\langle \mathbf{aa}^T \rangle = \mathbf{I}, \quad (A2)$$

where $\mathbf{I}$ is an identity matrix. Let vector $\mathbf{b}$ be a linear transformation of $\mathbf{a}$ defined by a matrix $\mathbf{S}$ and a scalar constant $m$;

$$\mathbf{b} = \mathbf{Sa} + m \mathbf{o} \quad (A3)$$

where $\mathbf{o}$ is a vector with all elements being one. Then the mean and covariance matrix of $\mathbf{b}$ are given by

$$\langle \mathbf{b} \rangle = \mathbf{S} \langle \mathbf{a} \rangle + m \mathbf{o} = m \mathbf{o} \quad (A4)$$
$$\langle \mathbf{b} - m \mathbf{o} \rangle (\mathbf{b} - m \mathbf{o})^T = \mathbf{S} \langle \mathbf{aa}^T \rangle \mathbf{S}^T = \mathbf{SS}^T \quad (A5)$$

using Eqs. (A1) and (A2), respectively. Equations (A4) and (A5) show that $\mathbf{b}$ is a set of random numbers with mean $m$ and covariance matrix $\mathbf{C}$ if $\mathbf{S}$ in (A3) satisfies,

$$\mathbf{C} = \mathbf{SS}^T \quad (A6)$$

For positive definite matrices $\mathbf{C}$, the Cholesky decomposition provides a particular decomposition of the form (A6), often referred to as the square root of matrix $\mathbf{C}$. The singular value decomposition provides an alternative decomposition of form (A6) even when $\mathbf{C}$ is rank deficient. The singular value decomposition of the symmetric covariance matrix $\mathbf{C}$ can be written as

$$\mathbf{C} = \mathbf{UU}^T, \quad (A7)$$

where the columns of $\mathbf{U}$ are the column (row) singular vectors of $\mathbf{C}$ with nonzero singular values. Here, $\mathbf{U}$ is a diagonal matrix of corresponding nonzero singular values. Defining $\mathbf{U}^{1/2}$ as a diagonal matrix with square root of the nonzero singular values, (A7) can be written as
\[ C = \mathbf{U} \Lambda^{1/2} \Lambda^{1/2} \mathbf{U}^T = (\mathbf{U} \Lambda^{1/2}) (\mathbf{U} \Lambda^{1/2})^T \quad (A8) \]

from which we can identify the matrix product

\[ \mathbf{U} \Lambda^{1/2} \quad (A9) \]

providing an alternate definition for \( S \) satisfying Eq. (A6).

REFERENCES


