KALMAN FILTERING OF LARGE-SCALE GEOPHYSICAL FLOWS BY APPROXIMATIONS BASED ON MARKOV RANDOM FIELD AND WAVELET

TOSHIO M. CHIN AND ARTHUR J. MARIANO

Rosenstiel School of Marine and Atmospheric Science, University of Miami
4600 Rickenbacker Causeway, Miami, Florida, 33149

ABSTRACT

Large-scale extended Kalman filters for atmospheric and oceanic circulation models can readily be approximated using a wavelet transform or a Markov random field model. For a filtering problem where the unknown field of the state variables is highly correlated and the observations are relatively sparse, the wavelet-approximated filter seems more appropriate. For a problem in which the covariance matrix is non-singular and a relatively large quantity of independent observations are processed, the MRF-approximated filter seems more appropriate.

1. INTRODUCTION

Recent technological and economical improvements in computational resources have contributed to growing interests in application of large-scale Kalman filter to atmospheric and oceanic circulation models [1]. By assimilating noisy measurements of various types into these models, meteorologists and oceanographers generally aim to improve the forecasting capability and error estimates. The Kalman filter and associated smoothing algorithms are also expected to provide a sophisticated data analysis tool which can improve initial and boundary conditions as well as insufficiently known model parameters [2].

Estimation of atmospheric and oceanic variables, such as pressure and wind/current velocity fields, with a Kalman filter is computationally demanding, as the number N of state variables tends to be $O(10^5 \sim 10^6)$. The main computational issue is the recursive updating of the covariance matrix, requiring prohibitive $O(N^3)$ flops per time step. Just as important is the issue of storing the N^2 elements of the covariance matrix, as it is more desirable to allocate any gain in computational resources to improve resolution of the variables, i.e., to increase N, rather than to store the entire covariance matrix.

We present two approaches to parameterize the covariance matrix in such large-scale Kalman filters. In one approach the covariance matrix is represented by a truncated set of its wavelet coefficients, while in the other approach the inverse of the covariance matrix (information matrix) is represented only by the elements at selective matrix locations. These parameterization schemes lead to computationally efficient implementations of the extended Kalman filter, in which the storage and computational requirements for the covariances are reduced to $O(N)$.

This work is supported by Office of Naval Research under Random Field in Oceanography ARI grant N00014-91J-1120. T.M. Chin is on leave at the National Center of Atmospheric Research, Boulder, Colorado

2. ALGORITHMS

Let $\mathbf{x}(t)$ denote the N-dimensional vector of state variables at time t of a circulation model. The partial differential equations representing the model can then be expressed generically as

$$\frac{\partial}{\partial t}\mathbf{x} = f(\mathbf{x}) \tag{1}$$

where $f(\cdot)$ is an aggregate of nonlinear (e.g., cross multiplication) yet spatially local (e.g., partial derivative over space) operators. For example, in a layer of constant-density water, the dynamic states of the layer depth and horizontal velocity can be described by a set of nonlinear partial differential equations called the "shallow water equations", which can be stacked to represent a multi-layered (in terms of density) body of water and serve as the basis of an ocean general circulation model [3].

Noisy measurements of the state vector are assumed to be provided at discrete time $t = k\tau$, $k = 1, 2, \ldots,$ as

$$\mathbf{y}_k = \mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k \tag{2}$$

where $\mathbf{x}_k \equiv \mathbf{x}(k\tau)$, τ is the measurement interval, and $\mathbf{v}_k$ is a zero-mean Gaussian noise process with covariance $\mathbf{R}_k$. We require the observation to be spatially local. Namely, the weighted sum operation performed by each row of the observation matrix $\mathbf{C}_k$ must have a local support in the spatial domain. Correspondingly, the observation noise is assumed to be only locally correlated so that $\mathbf{R}_k$ is diagonal or block diagonal with small block sizes. Under this spatial locality requirement, $\mathbf{C}_k$ would be a very sparse matrix and $\mathbf{R}_k$ would be sparse and easily inverted. Most measurements in practice are spatially localized as required (with a notable exception of tomographic measurements).

2.1. A wavelet-based Kalman filter

An extended Kalman filter for the dynamic system (1,2) can be formulated by first discretizing the partial differential equation (1) as

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \tau f(\mathbf{x}_{k-1}). \tag{3}$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

Let the $N \times N$ matrix $\mathbf{F}_k$ be the first-order term in the Taylor expansion of $\tau f(\cdot)$ about the filtered estimate $\widehat{\mathbf{x}}_k$ at time step k, i.e., $\mathbf{F}_k \equiv \tau \left[\frac{\partial f}{\partial \mathbf{x}}(\widehat{\mathbf{x}}_k) \right]$. Note that $\mathbf{F}_k$ is a sparse and spatially local matrix. Given an estimate $\widehat{\mathbf{x}}_{k-1}$ and its associated estimation error covariance $\widehat{\mathbf{P}}_{k-1}$ at time $k-1$, these two quantities can be updated by the measurement at time k as

$$\overline{\mathbf{P}}_k = \rho(\mathbf{I} + \mathbf{F}_{k-1})\widehat{\mathbf{P}}_{k-1}(\mathbf{I} + \mathbf{F}_{k-1})' \quad (4)$$

$$\overline{\mathbf{x}}_k = f(\widehat{\mathbf{x}}_{k-1}) \quad (5)$$

$$\mathbf{K}_k = \overline{\mathbf{P}}_k \mathbf{C}_k' \left(\mathbf{C}_k \overline{\mathbf{P}}_k \mathbf{C}_k' + \mathbf{R}_k \right)^{-1} \quad (6)$$

$$\widehat{\mathbf{P}}_k = \overline{\mathbf{P}}_k - \mathbf{K}_k \mathbf{C}_k \overline{\mathbf{P}}_k \quad (7)$$

$$\widehat{\mathbf{x}}_k = \overline{\mathbf{x}}_k + \mathbf{K}_k (\mathbf{y}_k - \mathbf{C}_k \overline{\mathbf{x}}_k) \quad (8)$$

where the superscript $'$ denotes transpose. The parameter $\rho > 1$ compensates for uncertainty (error) in the dynamic model itself and prevents the filter from becoming insensitive to measurements over time. A more general approach to represent model uncertainty is presented in [4].

As detailed in [4], the covariance matrix $\mathbf{P}$ can be parameterized by orthonormal wavelet bases to approximate the filtering algorithm, based on a physically plausible assumption that correlation between a pair of variables decays exponentially with the distance between the variable locations. The covariance matrix is compressed by truncating the wavelet coefficients of the matrix. Specifically, $\mathbf{P}$ is transformed into the so-called "standard form", and only the elements along certain bands ("fingers") of a given width w are computed and stored [5]. Such a compressed covariance matrix can be represented with only $Nw(3 - 2^{1-\ell})$ coefficients, where ℓ is the resolution level of the transform. In the approximate filtering algorithm, the covariances remain in their compressed form in the transform domain throughout the recursion over time. Availability of compactly supported orthonormal (wavelet) transforms [6] makes this filtering algorithm computationally efficient, compared with a similar algorithm that depends on other covariance-compressing transforms such as the cosine transform. With a version of the shallow water equations implemented over a cyclic domain, we have verified that the wavelet-based approximate filter can produce estimates just as accurate as the optimal Kalman filter [4].

2.2. An MRF-based Kalman filter

In an alternative implementation of the extended Kalman filter, the inverse $\mathbf{L}$ of the covariance matrix $\mathbf{P}$ is recursively updated instead of the covariance matrix itself.

The dynamic model (1) is time-discretized "implicitly" as

$$\mathbf{x}_k + \tau f(\mathbf{x}_k) = \mathbf{x}_{k-1}. \quad (9)$$

The recursion for $\widehat{\mathbf{x}}_k$ and $\widehat{\mathbf{L}}_k$, given $\widehat{\mathbf{x}}_{k-1}$ and $\widehat{\mathbf{L}}_{k-1}$, is

$$\overline{\mathbf{L}}_k = \rho^{-1}(\mathbf{I} - \mathbf{F}_{k-1})'\widehat{\mathbf{L}}_{k-1}(\mathbf{I} - \mathbf{F}_{k-1}) \quad (10)$$

$$\overline{\mathbf{x}}_k = f(\widehat{\mathbf{x}}_{k-1}) \quad (11)$$

$$\widehat{\mathbf{L}}_k = \overline{\mathbf{L}}_k + \mathbf{C}_k' \mathbf{R}_k^{-1} \mathbf{C}_k \quad (12)$$

$$\widehat{\mathbf{z}}_k = \overline{\mathbf{L}}_k \overline{\mathbf{x}}_k + \mathbf{C}_k' \mathbf{R}_k^{-1} \mathbf{y}_k \quad (13)$$

$$\widehat{\mathbf{x}}_k = \overline{\mathbf{L}}_k^{-1} \widehat{\mathbf{z}}_k \quad (14)$$

Since $\mathbf{F}$, $\mathbf{C}$, and $\mathbf{R}$ are sparse and spatially local, all matrix multiplications in this filtering algorithm can be implemented efficiently (and in parallel), as long as the information matrix $\mathbf{L}$ is sparse and spatially local. (The matrix inversion in (14) can be performed efficiently by an iterative inversion scheme, because a favorable initial condition $\overline{\mathbf{x}}_k$ is available.) Fortunately, the information matrix of a space-time process governed by local constraints tends to be sparse and local, and the filtering algorithm can be approximated efficiently by simply constraining the structure of the matrix $\mathbf{L}$ [7]. (Without such a structure constraint, $\mathbf{L}$ loses its sparseness over time due to the steps (12) and, particularly, (10).) Namely, a sparse and spatially local structure is pre-determined for the matrix $\mathbf{L}$, and only the matrix elements in this structure are actually computed and stored in the approximate filter. The approximate filtering algorithm is shown to produce near-optimal estimates in [8] (which also presents a general method to deal with model uncertainty).

As detailed in [7], the information matrix $\mathbf{L}$ can be considered as a spatial model for the corresponding estimation error process $\widehat{\mathbf{x}} - \mathbf{x}$ or $\overline{\mathbf{x}} - \mathbf{x}$ at a given time instant. Conceptually, the approximate filtering algorithm described above is obtained by treating each of the steps (10) and (12) as a model realization process for the error field and then approximating the resulting optimal field models with appropriate ones of reduced order. Such reduced order field models may be viewed as arising from the imposition of a Markov random field (MRF) structure on the estimation error processes.

2.3. Numerical tradeoffs

In the first filtering algorithm (4–8), the covariance matrix $\mathbf{P}$ is *explicitly* updated at each time step. In the second algorithm (10–14), on the other hand, covariances are *implicitly* updated as the filter performs recursion on the inverse $\mathbf{L}$ of the covariance matrix. Success of the approximate filtering strategies depends heavily on how well the non-approximated forms of extended Kalman filter matches with the particular filtering problem at hand. For example, the initial condition $\widehat{\mathbf{x}}_0$ is often assumed to be a highly correlated, or even deterministic, field in practice. Only the explicit algorithm (4–8) would be suitable for such a filtering problem, as the nearly-singular $\mathbf{P}$ would make the implicit algorithm (10–14) unstable.

Computational efficiency of the explicit algorithm depends partly on the density and volume of the observation. Specifically, as detailed in [4], the matrix inversion in step (6) is made efficient by exploiting spatially local and independent properties of the measurements at a given time step k (as reflected by the diagonal structure of $\mathbf{R}_k$). For each k, the steps (6–8) are repeated as many times as the number of the statistically independent units of measurements. The wavelet parameterized version of the explicit algorithm inherits this performance dependency on measurement volume.

The implicit algorithm (10–14) seems quite suitable for over-constrained filtering problems, in which large quantities of statistically independent measurements are processed. Namely, such measurements can prevent the field of state variables from acquiring deterministic (singular) characteristic and improve the condition number of the matrix **L** and, hence, convergence of an iterative solution to the inversion step (14).

3. APPLICATIONS

Kalman filters are expected to prevent circulation models from diverging. To present an extreme example, a straightforward discretization of the shallow water equations tends to be highly unstable as shown in Fig. 1. We have simulated observation of the "true ocean" by using a stable numerical implementation [3] of the equations, which produces the depth and velocity fields at a given time interval. Using the MRF-approximated filtering algorithm, a part of the simulated observations is assimilated into the *unstably discretized model.* Assimilation of only a quarter of the depth field at each time step leads to dramatic improvements in *both* the depth *and* velocity fields, as shown in Fig. 2.

Also, to demonstrate a flexible use of the filtering algorithms, a boundary condition is incorporated into the model as an "observation" of the filter. Namely, a coastline is introduced into an ocean basin by making observation of the depth h and velocity $\mathbf{u}$ as $h =$ constant and $\mathbf{n} \cdot \mathbf{u} = 0$, where $\mathbf{n}$ is the unit vectors normal to the given coastline, along the coastline and over the assumed "land" area. Using the MRF-approximated filter, the observations are gradually incorporated into a land-free circulation pattern, by controlling the observation noise variances at each time step. As shown in Fig. 3, the resulting circulation pattern appears to have reorganized itself as though the coastline had always been there.

The state dimension N in these examples was $O(10^4)$. Each recursion of (10–14) took 1 $\sim$ 3 minutes on a DEC Alpha workstation.

4. CONCLUSION

Two generic forms of extended Kalman filter for atmospheric and oceanic circulation models have been presented. Each form can be approximated for a computationally efficient implementation. There are numerical tradeoffs between the two approximate filtering algorithms. At present, we are evaluating these tradeoffs in order to develop a practical filter implementation for a specific ocean general circulation model (Miami Isopycnal Coordinate Ocean Model [3]).

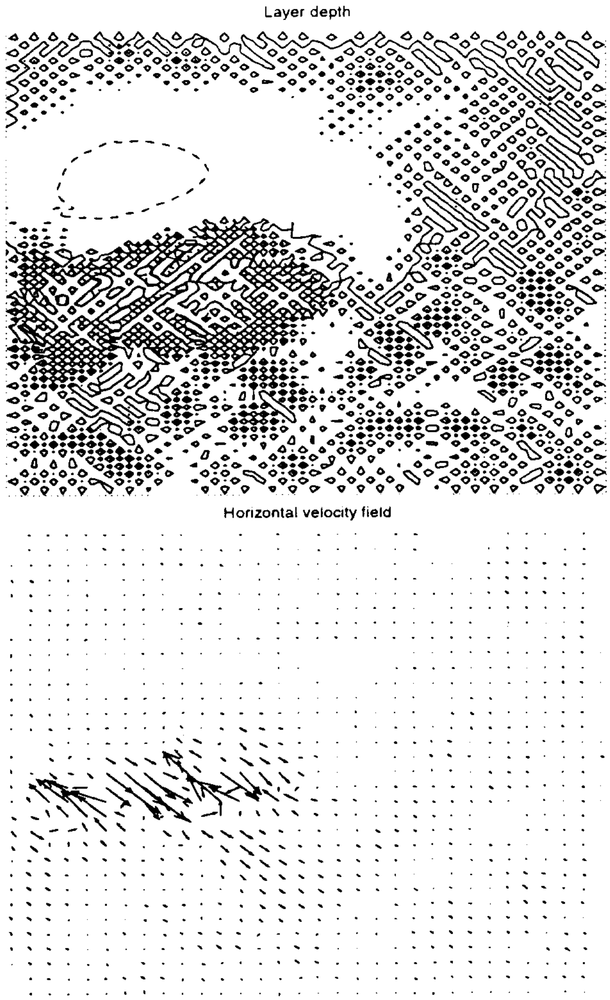

Figure 1: Depth (top) and velocity (bottom) fields produced by an unstable implementation of the shallow water equations, after 72 time steps (corresponding to a single simulation day).

5. REFERENCES

[1] M. Ghil and P. Malanotte-Rizzoli. Data assimilation in meteorology and oceanography. *Adv. Geophys.*, 33:141–266, 1991.

[2] G. Evensen. Using the extended Kalman filter with multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research*, 97:17905–17924, 1992.

[3] R. Bleck and D. Boudra. Wind-driven spin-up in eddy-resolving ocean models formulated in isopycnic and isobaric coordinates. *Journal of Geophysical Research*, 91:7611–7621, 1986.

[4] T. M. Chin and A. J. Mariano. Wavelet-based compression of covariance in Kalman filtering of geophysical flows. *Proceedings of SPIE*, 2242:842–850, 1994.

[5] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Comm. on Pure and Applied Math.*, 44:141–183, 1991.

[6] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. on Pure and Applied Math.*, 41:909–996, 1988.

[7] T. M. Chin, W. C. Karl, and A. S. Willsky. Sequential filtering for multi-frame visual reconstruction. *Signal Processing*, 28:311–333, 1992.

[8] T. M. Chin, W. C. Karl, and A. S. Willsky. Sequential optical flow estimation using temporal coherence. *IEEE Transactions on Image Processing*, 3(6), 1994.

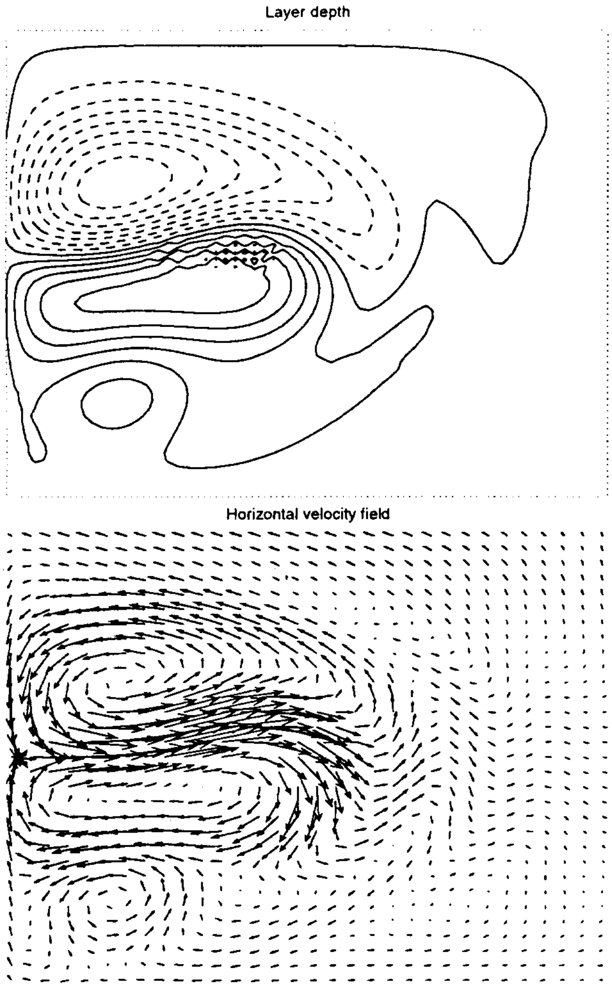

Figure 2: Depth and velocity fields produced by the same unstable model as Fig. 1, supplemented by partial observations of the correct depth field (after 72 steps). The solid contours represent peaks in the depth field, while the dashed contours represent depressions. Except for the small area of spurious oscillations in the center of the depth field, the depth field observation has overcome numerical unstability in the model and has led to accurate estimate of both the depth and velocity fields.

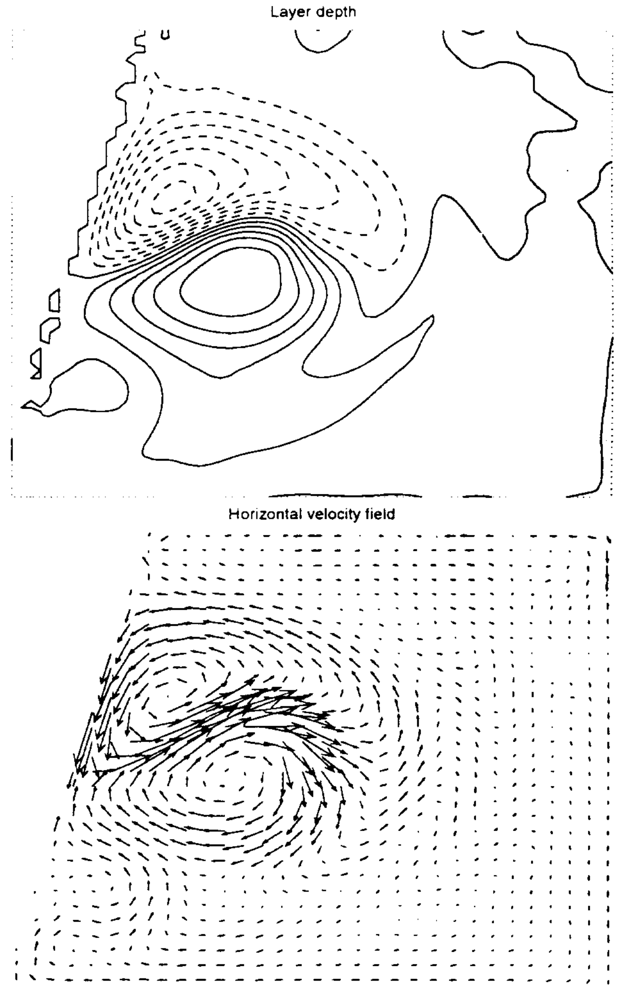

Figure 3: Output fields of the dynamic model supplemented by the "coastline model" after 800 time steps. A triangular area extending from the upper left corner and along the left edge of the field is considered to be "land". As the land has been gradually, instead of abruptly, incorporated into a flow field similar to one displayed in Fig. 2, the gyres in the depth and velocity fields have reorganized to retain their characteristic circulatory patterns, instead of truncating the patterns.