JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 104, NO. C4, PAGES 7991–8014, APRIL 15, 1999

Spatial regression and multiscale approximations for sequential data assimilation in ocean models

Toshio M. Chin, Arthur J. Mariano, and Eric P. Chassignet
Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida

Abstract. Effects of spatial regularity and locality assumptions in the extended Kalman filter are examined for oceanic data assimilation problems. Biorthogonal wavelet bases are used to implement spatial regularity through multiscale approximations, while a Markov random field (MRF) is used to impose locality through spatial regression. Both methods are shown to approximate the optimal Kalman filter estimates closely, although the stability of the estimates can be dependent on the choice of basis functions in the wavelet case. The observed filter performance is nearly constant over a wide range of values for the scalar weights (uncertainty variances) given to the model and data examined here. The MRF-based method, with its inhomogeneous and anisotropic covariance parameterization, has been shown to be particularly effective and stable in assimilation of simulated TOPEX/POSEIDON altimetry data into a reduced-gravity, shallow-water equation model.

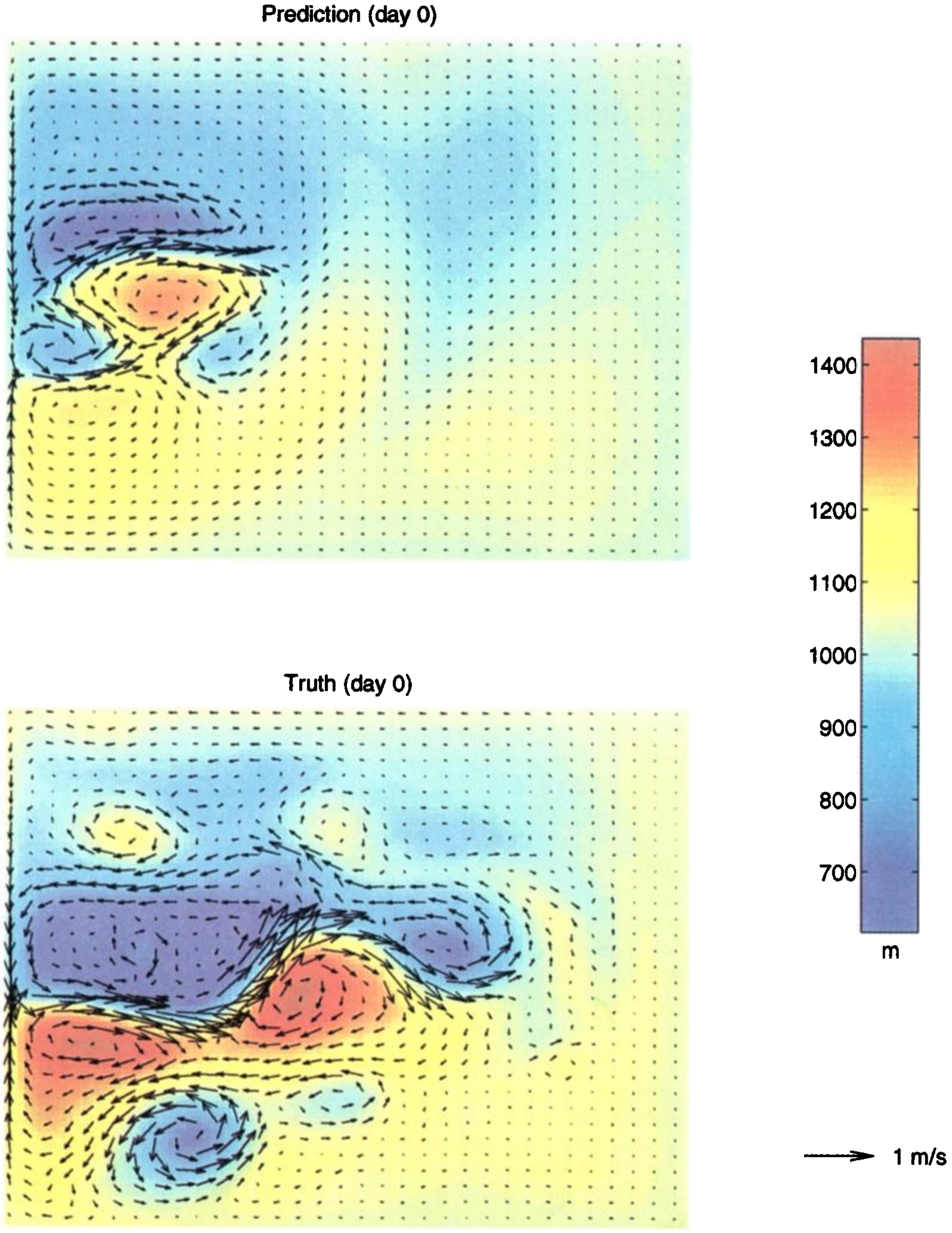
1. Introduction

From a mathematical perspective, data assimilation in physical oceanography implies solution of an overdetermined system of equations for the prognostic variables of a circulation model. In numerical models, data are typically used selectively to provide the necessary values, including the initial and boundary conditions and forcing fields, to establish the forward (time) recursion for the prognostic variables. In data assimilation the time trajectories of the model variables are further constrained by extraneous data, usually inferred from ship, buoy, float, and satellite measurements but also possibly obtained from measured or simulated specifications of open-boundary conditions and air-sea interfaces. A classic approach to evaluate the overconstrained variables objectively is to apply an optimality condition, such as a least squares criterion [*Sasaki*, 1970; *Thacker and Long*, 1988; *Bennett*, 1992; *Wunsch*, 1996], which is employed by most of the sophisticated data assimilation methods under consideration today [*Ghil and Malanotte-Rizzoli*, 1991; *Talagrand*, 1997]. Although solution techniques for a least squares formulation are relatively straightforward in theory, they present serious practical problems with computational speed and storage owing to the large number of variables required to represent a prognostic state in typical ocean circulation models. Approximate solution techniques are thus necessary for objective data assimilation at present. In this paper, such approximation techniques are examined in conjunction with the Kalman filter, a sequential algorithm based on the probability theory, that allows formulation and interpretation of the least squares solutions in terms of the mean and covariance of a normal (Gaussian) distribution [*Anderson and Moore*, 1979]. Assimilation of satellite-based sea-surface height data into primitive equation models is then considered.

When applied to data assimilation, a Kalman filter performs time-sequential statistical interpolation that incrementally corrects for the discrepancy between the observations (data) and model predictions [*Ghil et al.*, 1981]. The statistical interpolation is based on the prediction/analysis error covariances, which are estimated using a forward recursion consistent with the model dynamics. An approximation technique for Kalman filter in ocean data assimilation must address computational efficiency of this forward recursion for the error covariance matrix, whose dimension is the square of the number of the model prognostic variables. Using both physically and algebraically motivated assumptions, various techniques to parameterize the covariance matrix have been explored and practiced to drastically reduce the matrix dimensions. Two qualitative yet prevailing features of a geophysical covariance function are regularity (smoothness) and locality (decay or "*e*-folding") in space. The assumption of regularity concerns maintenance of consistency between the spatial scales of the dominant dynamic modes in the prognostic variables and the correlation scales of the corresponding error process [*Bennett and Budgell*, 1987; *Cane et al.*, 1996], and it tends to allow a coarsely sampled representation of the covariance matrix. The assumption of locality indicates that the correlation between a pair of variables decays steeply (e.g., exponentially) with respect to the distance between them [*Daley*, 1991], and it facilitates parameterization using a small number of regression coefficients.

The structure of the error covariance matrix is partially dependent on the observation network (the sampling patterns associated with the extraneous data) and hence is highly variable from case to case, even with the same dynamic model. In practice, the regularity assumptions are useful in situations where the data density and analysis objective are appropriate for long-distance correlation structures. For data sets sampled sparsely over space (e.g., six tide gauges for the tropical Pacific [*Miller and Cane*, 1989]), only large-scale features can be sampled, for which case a highly "regular" (generally containing strong correlation over long distance) error covariance matrix would be desirable for smooth interpolation uncontaminated with spurious or artificial small-scale features. A climatological analysis also focuses on large-scale events to which smooth and nonlocal covariance structures are relevant. On the other hand, for denser data sets, such as satellite-based measurements whose sampling pattern can usually resolve at much

Paper number 1998JC900075.
0148-0227/99/1998JC900075$09.00


Prediction (day 0)

Truth (day 0)

Plate 1. The two initial states of the identical-twin experiment. Every three velocity vectors, in each of the horizontal and vertical directions, are shown.

finer scales, the locality assumptions would reduce oversmoothing and enhance accuracy in reproduction of features such as mesoscale rings. Some degree of regularity in addition to locality is still important in this case to prevent generation of destabilizing inertia-gravity waves, as in the case with initialization of data for meteorological models [*Daley*, 1991].

Regularity in covariance can be realized numerically by subspace projection, which restricts the dynamic behavior of the prediction/analysis error fields to a small number of dominant modes represented by smooth functions. Several approaches to Kalman filter approximation including those by *Ghil et al.* [1981], *Cohn and Parrish* [1991], *Fukumori and Malanotte-Rizzoli* [1995], and *Cane et al.* [1996] can be examined in this unifying mathematical framework, as detailed in this paper. While locality can be further imposed in this framework by choosing smooth and localized mode functions for the error subspace, the stability of the resulting data assimilation system appears to be dependent on the choice for the mode functions. Alternatively, a numerical realization of locality in covariance can be achieved more directly by spatial regression, a multidimensional generalization of standard autoregression (e.g., for time series analysis), which can be formulated naturally using a Markov random field (MRF). The MRF formalism [*Kindermann and Snell*, 1980] uses a strictly local (in space) prescription of mathematical expressions to characterize statistical dependence over a wide range of spatial scales, finding applications in a variety of fields such as statistical mechanics [*Kindermann and Snell*, 1980], population biology [*Besag*,

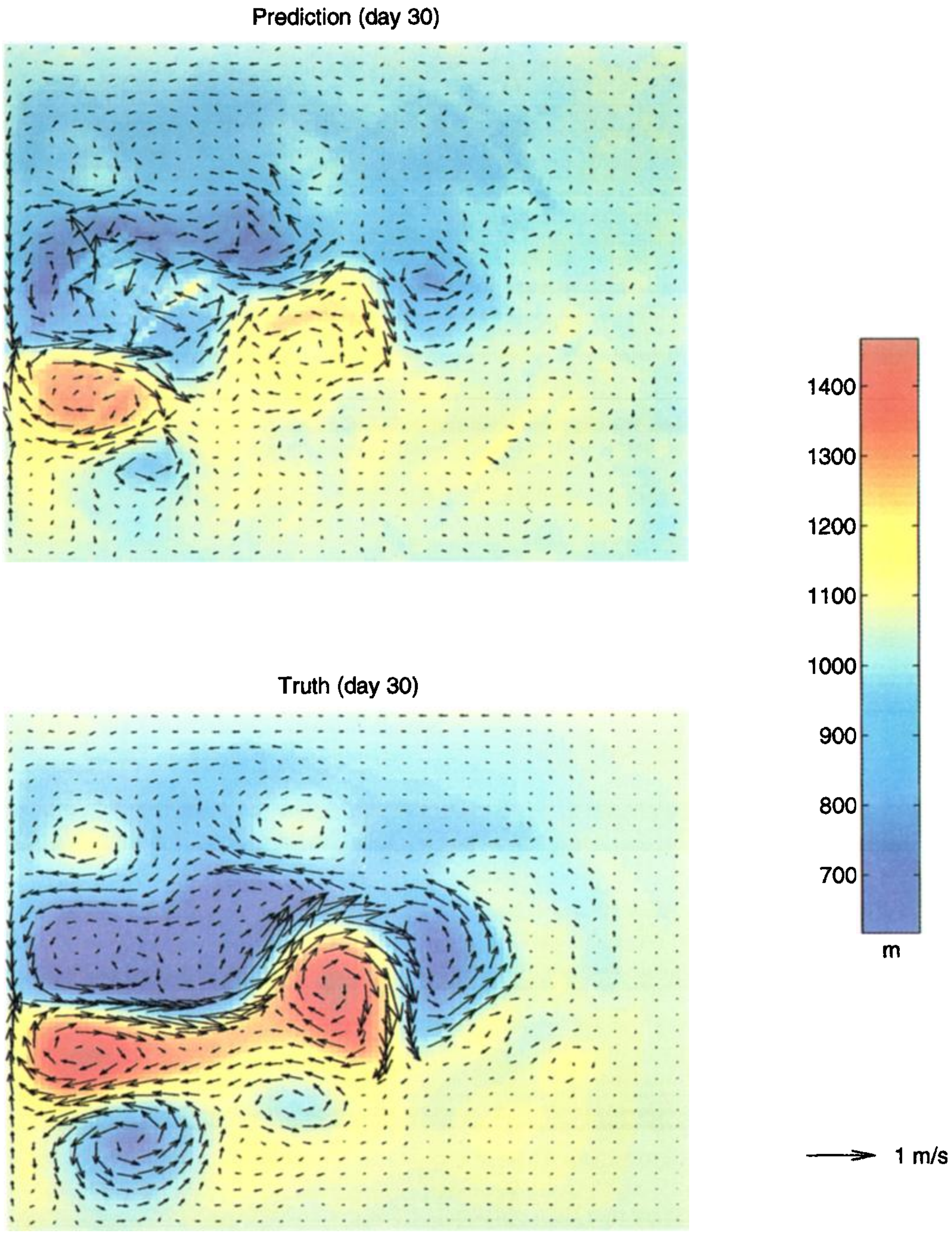


Plate 2. (top) Prediction fields after 30 days (2160 time steps) of assimilating the layer thickness *h* field from the (bottom) truth.

1974], and image processing [*Geman and Geman*, 1984] for statistical description of complex spatial patterns. In our approach to Kalman filter approximation the MRF essentially replaces the covariance matrix with a diagnostic (spatial) operator, or a finite difference operator with a numerical complexity similar to that of a geostrophic balance equation, operating on the prediction/analysis error process. The role of the diagnostic operator then is to encode the correlation structure in the error process, and the order of this finite difference operator determines the correlation scales of the covariance approximation.

The emphasis in this paper is on realization of locality as a strategy for assimilation of relatively dense and widely sampled satellite-based oceanic measurements. A numerical scheme to dynamically update the inhomogeneous MRF description of the covariance matrix in a manner consistent with a given ocean circulation model is presented. Performance of the resulting approximate Kalman filter is then assessed through a variety of simulated data assimilation experiments.

2. Data Assimilation With Kalman Filter

We first formulate the data assimilation task as a weighted least squares problem and designate the Kalman filter as a means to solve this minimization problem. The readers familiar with the Kalman filter may skip to section 3, after glancing at the mathematical notations in this section.

Let the equations for a numerical ocean circulation model be represented by the algebraic recursion

$$\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}) + \varepsilon_k^{\text{model}} \tag{1}$$

where the state vector $\mathbf{x}_k$, containing all the dynamically independent prognostic variables, represents the synoptic ocean in the model at time $k\Delta t$, $k = 0, 1, 2, \cdots$, for a given model time step Δt. The function $\mathbf{f}_k$ is a generic representation of computation of a single time step in a primitive equation model, while $\varepsilon_k^{\text{model}}$ is the aggregate of physical uncertainty and numeric errors in the model. Typically, $\varepsilon_k^{\text{model}}$ can only be characterized statistically. Let the vector $\mathbf{y}_k$ be a representation of the observable variables at time $k\Delta t$ such that

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \varepsilon_k^{\text{data}} \tag{2}$$

with an additive noise/error term $\varepsilon_k^{\text{data}}$. The operator $\mathbf{H}_k$ samples the synoptic map $\mathbf{x}_k$ by linear combinations and thus tends to be a diagonally structured and sparsely banded matrix. The observed values ("data") of $\mathbf{y}_k$ are used to constrain the prognostic variables in $\mathbf{x}_k$ to adjust the dynamic trajectory toward the observation. The general case where $\mathbf{y}_k$ is a nonlinear function of $\mathbf{x}_k$ can also be formulated in this framework through local linearization [*Cohn and Parrish*, 1991] but is not addressed in this paper, as practical effectiveness of such linearization is yet to be documented.

The Kalman filter is intimately associated with the time-dependent least squares problem (for a given time index k)

$$\min_{(\mathbf{x}_0,\mathbf{x}_1,\cdots,\mathbf{x}_k)} \|\mathbf{x}_0 - \mathbf{x}_0^d\|_{\mathbf{L}_0}^2 + \sum_{j=1}^{k} [\|\mathbf{x}_j - \mathbf{f}_j(\mathbf{x}_{j-1})\|_{\mathbf{M}_j}^2 + \|\mathbf{y}_j^d - \mathbf{H}_j\mathbf{x}_j\|_{\mathbf{N}_j}^2] \tag{3}$$

where $\|\mathbf{v}\|_{\mathbf{W}}^2 \equiv \mathbf{v}^T\mathbf{W}\mathbf{v}$ denotes the square norm of a column vector $\mathbf{v}$ weighted by a square matrix $\mathbf{W}$. The initial condition $\mathbf{x}_0^d$ for the model (1) and the full set of observations $\mathbf{y}_j^d$, $j = 1, 2, \cdots, k$, are provided externally and denoted with the superscript d, indicating "data." The weighting matrices $\mathbf{L}_0$, $\mathbf{M}_j$, and $\mathbf{N}_j$ (for $j = 1, 2, \cdots, k$) are the parameters of this particular least squares formulation. The two terms weighted by $\mathbf{M}_j$ and $\mathbf{N}_j$, defining the relative confidence in the model (1) and data (2), tend to dominate the initial condition term weighted by $\mathbf{L}_0$ as more observations become available with increasing k. We denote the minimizing solution for (3) as $\mathbf{x}_0^a, \mathbf{x}_1^a, \cdots, \mathbf{x}_k^a$ (superscript a for "analysis") and remind readers that the optimality condition (3) is dependent on the time index k. The Kalman filter then is a time-recursive algorithm to compute only the latest member $\mathbf{x}_k^a$ of the optimal solution for (3) for each $k = 1, 2, \cdots$.

The time recursion for the Kalman filter (initialized as $\mathbf{x}_0^a = \mathbf{x}_0^d$) is

$$\mathbf{x}_k^f = \mathbf{f}_k(\mathbf{x}_{k-1}^a) \tag{4}$$

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^d - \mathbf{H}_k\mathbf{x}_k^f) \tag{5}$$

where the superscript f denotes "forecast." The last additive term in (5) computes the correction necessary to reduce the difference between the forecast and measured values of $\mathbf{y}_k$. The matrix $\mathbf{K}_k$, known as the "filter gain," controls how this correction is incorporated into the prognostic variables. Note that if $\mathbf{K}_k$ is null (zero), the recursion (4)–(5) is equivalent to the circulation model in (1).

The essence of the Kalman filter is in statistically optimal computation of the filter gain $\mathbf{K}_k$, which is obtained through a separate recursion of the error covariance matrix, defined as $\mathbf{P}_k \equiv \langle \mathbf{e}_k\mathbf{e}_k^T\rangle$, where $\mathbf{e}_k$ is the estimation error, i.e., $\mathbf{e}_k^f \equiv \mathbf{x}_k^f - \mathbf{x}_k$ and $\mathbf{e}_k^a \equiv \mathbf{x}_k^a - \mathbf{x}_k$, and angle brackets denote ensemble average. To obtain an optimal recursion for $\mathbf{P}_k$ consistent with the least squares (3), each inverse of the weight matrices, $\mathbf{P}_0 \equiv \mathbf{L}_0^{-1}$, $\mathbf{Q}_j \equiv \mathbf{M}_j^{-1}$, and $\mathbf{R}_j \equiv \mathbf{N}_j^{-1}$, is treated as the covariance matrix of an independent random vector representing the respective errors in initialization $(\mathbf{x}_0 - \mathbf{x}_0^d)$, modeling $(\varepsilon_j^{\text{model}})$, and observation $(\varepsilon_j^{\text{data}})$. Such a statistical framework facilitates determination of the weighting parameters and interpretation of the filter prediction/analysis.

The optimal time recursion for the error covariance $\mathbf{P}_k$ (initialized as $\mathbf{P}_0^a = \mathbf{P}_0 \equiv \mathbf{L}_0^{-1}$) can be given as

$$\mathbf{P}_k^f = \mathbf{F}_k\mathbf{P}_{k-1}^a\mathbf{F}_k^T + \mathbf{Q}_k \tag{6}$$

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{7}$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \tag{8}$$

where the matrix $\mathbf{F}_k \equiv \partial\mathbf{f}_k/\partial\mathbf{x}(\mathbf{x}_{k-1}^a)$ is a tangent linear approximation of the model dynamics about the most recent estimate $\mathbf{x}_{k-1}^a$. The filter analysis $\mathbf{x}_k^a$ thus corresponds exactly to the minimizing solution to the least squares (3) only when the model dynamics $\mathbf{f}_k(\mathbf{x}_k)$ is linear in $\mathbf{x}_k$ for all k. The tangent linear approximation of model dynamics (known as the extended Kalman filter) invokes theoretical issues involving "closure" of statistical moments [*Cohn*, 1993; *Miller et al.*, 1994; *Evensen*, 1994]; however, in practice, the approximation appears to be suitable for a typical ocean circulation model that tends to have a small time step Δt relative to evolution timescales of circulation features [*Menemenlis and Wunsch*, 1997].

The optimal filter gain $\mathbf{K}_k$ given in (7) would make the data-update equation (5) for a given k equivalent to the optimal interpolation [*Daley*, 1991] and Gauss-Markov estimator [*Wunsch*, 1996] applied to a variety of operational geophysical data analysis. The Kalman filter can thus be naturally interpreted as a time sequence of spatial interpolators, in which the optimality criteria evolve in a manner consistent with the model dynamics and data distribution (observation network).

Practicality of the extended Kalman filter in oceanic data assimilation depends on two procedures: choice of the filter parameters and reduction of the variable dimensions. The filter parameters to be specified are the matrices $\mathbf{P}_0$, $\mathbf{Q}_k$, and $\mathbf{R}_k$ (equivalently, $\mathbf{L}_0$, $\mathbf{M}_k$, and $\mathbf{N}_k$), $k = 1, 2, \cdots$. Determination of realistic spatial correlation structures in $\mathbf{Q}_k$ for the model dynamics error is still an open issue, although a high degree of regularity in $\mathbf{Q}_k$ has been determined crucial [*Bennett and Budgell*, 1987; *Dee*, 1991; *Cane et al.*, 1996], especially for assimilation problems with sparsely sampled data. The need for a reduction in the dimensions of the filter variables is due primarily to demands for computational resources to store and time update the large prediction and analysis error covariance matrices, $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$. If N denotes the dimension of the prognostic variable set $\mathbf{x}_k$, the corresponding error covariance matrix $\mathbf{P}_k$ then contains N^2 variables (elements), a tremendous storage requirement for a typical value of $N = 10^4$–10^8. As the computational demands are already at a premium for the N-dimensional recursion of the model dynamics (4) [*Bleck et al.*, 1995], the N^2-dimensional covariance recursion (6)–(8) must be greatly simplified numerically, in practice. A dimensionally reduced representation for the covariance matrix is intended to serve such a computational purpose, but it can also influence the matrix structures (and hence correlation structures) of the filter parameters to be determined. The remainder of this paper thus focuses on approaches to reduce the covariance matrix $\mathbf{P}_k$ in dimension by parameterization.

3. Reduction of Covariance Dimension

Approximation of a covariance matrix through parameterization of its correlation structure is a common practice in geophysical data interpolation [e.g., *Mariano and Brown*, 1992]. For the inhomogeneous and anisotropic covariance matrix $\mathbf{P}_k$ in a Kalman filter, the regularity (smoothness) and locality (decay) assumptions are realized mathematically using "subspace projection" and "spatial regression" schemes, respectively. The mathematical issue here is to develop a parameterized arithmetic recursion consistent with the optimal covariance dynamic equations (6)–(8).

3.1. Subspace Projection for Spatial Regularity

A systematic approach to impose regularity in the error covariance, essentially taken by *Fukumori and Malanotte-Rizzoli* [1995] and *Cane et al.* [1996] among others, is to constrain evolution of the error fields $\mathbf{e}_k$ within a subspace defined by a set of smooth, linearly independent, normalized basis ("mode") vectors $\mathbf{b}_i$, as

$$\mathbf{e}_k \approx \sum_{i=1}^{n} \tilde{e}_{ki}\mathbf{b}_i = \mathbf{B}\tilde{\mathbf{e}}_\mathbf{k} \tag{9}$$

where $\tilde{\mathbf{e}}_k$ is the column vector of the modal coefficients $\tilde{e}_{ki}$ and $\mathbf{B}$ is the $N \times n$ matrix whose columns are the basis vectors $\mathbf{b}_i$, $i = 1, 2, \cdots, n$. The idea is to constrain evolution of $\mathbf{e}_k$ using a small number $n < N$ (where N is the dimension of the state $\mathbf{x}_k$) of carefully selected modes $\mathbf{b}_i$, so that the resulting reduction in degrees of dynamic freedom maintains spatial regularity in $\mathbf{P}_k$ and accuracy in the estimate $\mathbf{x}_k^a$ while decreasing computational costs.

The error vector $\mathbf{e}_k$ is transformed into the modal subspace as $\tilde{\mathbf{e}}_k = \mathbf{T}\mathbf{e}_k$, by the pseudoinverse $\mathbf{T}$ of the mode matrix $\mathbf{B}$, given as $\mathbf{T} \equiv (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$. Substitution into (9) then reveals the essence of the approximation as $\mathbf{e}_k \approx \mathbf{B}\mathbf{T}\mathbf{e}_k$, in which the matrix operator $\mathbf{BT}$ performs a spatial low-pass filtering for a low-resolution approximation of the error fields [*Fukumori and Malanotte-Rizzoli*, 1995]. The time recursion of the covariance matrix $\tilde{\mathbf{P}}_k$ in the modal subspace can be derived from (6)–(8) by straightforward linear transformations (note that $\mathbf{TB} = \mathbf{I}$, the identity matrix) as

$$\tilde{\mathbf{P}}_k^f = \tilde{\mathbf{F}}_k\tilde{\mathbf{P}}_{k-1}^a\tilde{\mathbf{F}}_k^T + \tilde{\mathbf{Q}}_k \tag{10}$$

$$\tilde{\mathbf{K}}_k = \tilde{\mathbf{P}}_k^f\tilde{\mathbf{H}}_k^T(\tilde{\mathbf{H}}_k\tilde{\mathbf{P}}_k^f\tilde{\mathbf{H}}_k^T + \mathbf{R}_k)^{-1} \tag{11}$$

$$\tilde{\mathbf{P}}_k^a = (\mathbf{I} - \tilde{\mathbf{K}}_k\tilde{\mathbf{H}}_k)\tilde{\mathbf{P}}_k^f(\mathbf{I} - \tilde{\mathbf{K}}_k\tilde{\mathbf{H}}_k)^T + \tilde{\mathbf{K}}_k\mathbf{R}_k\tilde{\mathbf{K}}_k^T \tag{12}$$

where the transformed matrices are given as $\tilde{\mathbf{P}}_k = \mathbf{T}\mathbf{P}_k\mathbf{T}^T$, $\tilde{\mathbf{Q}}_k = \mathbf{T}\mathbf{Q}_k\mathbf{T}^T$, $\tilde{\mathbf{F}}_k = \mathbf{T}\mathbf{F}_k\mathbf{B}$, and $\tilde{\mathbf{H}}_k = \mathbf{H}_k\mathbf{B}$. In principle, the gain matrix can be computed as $\mathbf{K}_k = \mathbf{B}\tilde{\mathbf{K}}_k$, with which the data update (5) can then be performed. In practice, the data update can be accomplished more efficiently without such an explicit evaluation of the gain matrix $\mathbf{K}_k$, i.e., by first computing the n-dimensional vector $\tilde{\mathbf{K}}_k(\mathbf{y}_k^d - \mathbf{H}_k\mathbf{x}_k^f)$ and then multiplying the resulting vector by $\mathbf{B}$ (inverse transform). For the typical case of $n \ll N$, reduction in computational costs is immediately apparent as the transformed covariance matrix $\tilde{\mathbf{P}}_k$ is dimensionally reduced to $n \times n$.

3.1.1. Subspaces of dynamic modes. The key issue in the subspace projection approach is determination of the approximation subspace $\mathbf{B}$. Choosing a small subspace (small n) could make the approximated covariance recursion (10)–(12) highly efficient; however, the danger of introducing spurious spatial structures (estimation biases) is higher with a smaller number of basis functions.

When the dynamic model is linear and time invariant; that is, $\mathbf{f}_k(\mathbf{x}_k) = \mathbf{F}\mathbf{x}_k$ for all k, the range space of the dynamics is spanned by the column vectors of the matrix $\mathbf{U}$ given by the singular value decomposition (SVD) $\mathbf{F} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices (matrices whose columns are normalized and mutually orthonormal) and $\mathbf{S}$ is a diagonal matrix of singular values. Using linearized shallow-water equation models, *Ghil et al.* [1981] and *Cohn and Parrish* [1991] have partitioned the dynamic modes as $\mathbf{U} = [\mathbf{U}_{\text{slow}}, \mathbf{U}_{\text{fast}}]$ into the numerically desirable "slow modes" $\mathbf{U}_{\text{slow}}$ and nondesirable "fast modes" $\mathbf{U}_{\text{fast}}$, respectively, representing the Rossby waves and inertial gravity waves, leading to a Kalman filter approximation with the subspace $\mathbf{B} = \mathbf{U}_{\text{slow}}$.

For nonlinear dynamics, direct modal decomposition is not as straightforward. Alternatively, *Cane et al.* [1996] and *Pham et al.* [1998] have used the empirical orthogonal functions (EOFs) for the variability in the state $\mathbf{x}_k$ as the mode vectors for the error process $\mathbf{e}_k$ (which is difficult to measure or simulate directly). Typically, the total variance of the model variability is contributed by a small number n of the EOFs, and using such EOFs as the error modes $\mathbf{b}_i$ often results in an efficient covariance recursion due to this small n. These statistically dominant dynamic modes of the model variability tend to have inherent large-scale regularity. Constraining the error dynamics using these smooth mode functions has an advantage of greatly reducing the chance of exciting numerically undesirable small-scale modes such as inertia-gravity waves. An outstanding issue in the EOF subspace approach is computation and selection of the appropriate EOF modes. Both the choice of the empirical statistics (e.g., design of simulation experiments) and partitioning of the corresponding EOFs into "desirable" and "undesirable" modes can affect the filter analysis to a large extent. A lack of sensitivity to smaller-scale (e.g., mesoscale) features in data can also be expected for typical EOF modes owing to their large-scale and ensemble (statistical) natures.

3.1.2. Subspaces based on wavenumber and scale. To prevent oversmoothing or large-scale biases that might be introduced by statistical modes like EOFs, nonempirical mode functions have also been considered. For example, the dynamic modes of linearized shallow-water equations over a periodic domain are the Fourier basis functions with which an approximation subspace $\mathbf{B}$ can be defined in terms of wavenumber spectra [*Ghil et al.*, 1981]. Also, *Fukumori and Malanotte-Rizzoli* [1995] have used a cubic spline function (a locally supported cubic polynomial) for spatial smoothing and coarse-scale approximation of Kalman filter. In these approaches, rudimentary geometric notions such as wavenumber and scale, respectively, are used to characterize the approximation subspaces.

Specifying the error mode functions in terms of a characteristic scale is especially appealing, as some notion of "locality" may be imposed using the scale parameter (in addition to regularity already assumed in the mode functions). Smooth and locally supported functions like spline bases seem to possess qualities that satisfy both the regularity and locality assumptions. In particular, the spline functions associated with the wavelet-based "multiresolution analysis" [*Daubechies*, 1992] offer a systematic approach to construct an approximation subspace **B** defined by a characteristic scale length. These basis functions are self-similar functions that change their scales diadically (by a factor of 2) between two adjacent levels in scale space (Figure 1). Each level of coarse-scale approximation thus reduces the dimension of the original subspace by half, so that using ℓ levels of approximation would result in $n = N2^{-\ell}$. A numerically convenient set of multiresolution basis functions is the compactly supported, biorthogonal wavelet ("scaling") functions [*Daubechies*, 1992]. With these functions, both **B** and **T** become sparse and banded matrices, offering computational advantages (Appendix A). We use such a subspace **B** in the numerical experiments (section 4).

3.2. Spatial Regressions for Spatial Locality

Alternatively, compact representations of a large covariance matrix can also be obtained using regression formulas. For example, an autoregression of order n (AR-n) applied to the elements e_j ($j = 1, 2, \cdots, N$) of the error vector $\mathbf{e}_k$ can be written as

$$e_j = \sum_{i=1}^{n} \alpha_i e_{j-i} + \nu_j \tag{13}$$

where ν_j, $j = 1, \cdots, N$, are mutually uncorrelated noise (zero-mean stochastic) processes with variance β_j^2. As ν_j are uncorrelated, the statistical dependence among the variables e_j are encoded solely by the regression coefficients α_i. The N^2 covariances in $\mathbf{P}_k$ can then be represented by only the $N + n$ parameters, α_i and β_j. The standard autoregression formulas are applicable only to stochastic processes over a one-dimensional (1-D) domain such as a time series. For multidimensionality, noncausality, and inhomogeneity expected in the spatial process $\mathbf{e}_k$, we consider the generalized regression formula

$$e_j = \sum_{i \in Z} \alpha_{ij} e_{j-1} + \nu_j \tag{14}$$

where α_{ij} is now inhomogeneous and hence dependent on j, while a multidimensional and noncausal (i.e., no directional preference) index set Z is used for i to define the extent of statistical (conditional) dependence. Examples of Z in two dimensions are the stencils for the numerical Laplacian and biharmonic operators (crosses in Figures 2a and 2b). The number of parameters necessary to represent the N^2 covariances is approximately nN, where n is the size of the index set Z.

Both the AR-n model (13) and its spatial generalization (14) use a local (the site j and its neighbors on the respective grid) parameterization to represent the covariance structure for the entire domain. This notion of being "local" is consistent with the probabilistically more rigorous definition of locality in term of the (nth order) Markov condition [e.g., *Wilks*, 1995] and Markov random field [*Kindermann and Snell*, 1980; *Chin et al.*, 1995] for the AR-n model (13) and its spatial generalization (14), respectively. We mention that the MRF paradigm is a multidimensional generalization of the better known 1-D Markov condition and refer the interested readers to cited references for more details. The particular MRF model of interest to us and implied by (14) is a Gaussian-distributed and spatially heterogeneous version. The size and topology of the index set Z (Figures 2a and 2b) determine the complexity or order of the MRF model. The accuracy of MRF parameterization of the covariance $\mathbf{P}_k$ increases with the order of model, while computational efficiency of the corresponding Kalman filter decreases with the order. Our numerical experiments (section 4) show that a drastically low order MRF is sufficient for sequential assimilation of satellite altimetry data.

To relate the spatial regression coefficients with the Kalman filter covariance matrix, (14) is rewritten in term of a matrix operator $\boldsymbol{\Gamma}_k$ as

$$\boldsymbol{\Gamma}_k \mathbf{e}_k = \boldsymbol{\delta} \tag{15}$$

such that $\boldsymbol{\delta}$ denotes a zero-mean process whose covariance is the identity matrix **I** (a spatially uncorrelated covariance matrix with unit variance). Only the elements (denoted as γ) along and near the main diagonal of the matrix $\boldsymbol{\Gamma}_k$ are nonzero and given as $\gamma_{jj} = 1/\beta_j$ and $\gamma_{j,j-i} = -\alpha_{ij}/\beta_j$, $i \in Z$ for each row j. Sparseness of the matrix $\boldsymbol{\Gamma}_k$ depends on the size of the index set Z (Figures 2c and 2d). It is insightful to interpret $\boldsymbol{\Gamma}_k$ as a numerical differential operator in space. A deterministic version of (15) would be $\boldsymbol{\Gamma}_k \mathbf{e}_k = 0$, a balancing equation such as the geostrophic balance (see (31)) that conforms the residual $\mathbf{e}_k$ for optimal interpolation [*Daley*, 1991]. The stochastic expression (15) can be considered as a weakly imposed version of such a balancing constraint. Equating the covariance matrices for the vectors on both sides of (15) would yield $\boldsymbol{\Gamma}_k \mathbf{P}_k \boldsymbol{\Gamma}_k^T = \mathbf{I}$. The square of the regression operator $\mathbf{L}_k \equiv \boldsymbol{\Gamma}_k^T \boldsymbol{\Gamma}_k$, often referred to as the information matrix, is then just the inverse of the covariance matrix, $\mathbf{L}_k = \mathbf{P}_k^{-1}$ (when $\mathbf{P}_k$ is nonsingular). Without the requirement to be sparse (i.e., using the highest possible order for MRF), the regression operator $\boldsymbol{\Gamma}_k$ or the information matrix $\mathbf{L}_k$ can thus encode the error covariance $\mathbf{P}_k$ exactly. Indeed, in several algorithmic variants of the Kalman filter, the time recursion (6)–(8) for $\mathbf{P}_k$ are replaced by a recursion of $\boldsymbol{\Gamma}_k$ [*Chin et al.*, 1995] or of $\mathbf{L}_k$ (Appendix B), known as the square root information filter and information filter, respectively.

In conjunction with the MRF-based approximation, the information filter is found to execute faster than its square root counterpart on serial computers in practice. We thus focus on time recursion of $\mathbf{L}_k$ instead of $\boldsymbol{\Gamma}_k$ to implement the spatial regression approximation. A spatially local matrix $\boldsymbol{\Gamma}_k$ indicative of a low-order MRF parameterization imposes a sparse, block-banded, and symmetric structure on the corresponding information matrix $\mathbf{L}_k$. In our implementation of Kalman (information) filter, $\mathbf{L}_k$ is thus truncated accordingly at each time step. Figure 2 exemplifies the symmetric and "banded" truncation patterns used. Note that direct truncation of $\mathbf{L}_k$ results in different MRF from one obtained by truncation of $\boldsymbol{\Gamma}_k$ using the same pattern. This is not a problem because a sparse $\boldsymbol{\Gamma}_k$ is always associated with a sparse $\mathbf{L}_k$. ($\mathbf{L}_k$ is also preferred over $\boldsymbol{\Gamma}_k$ in standard MRF terminology, e.g., "neighborhood," which indicates the locations of the nonzero bands in $\mathbf{L}_k$.) We emphasize that, unlike direct truncation of long-distance covariances by a banded approximation of $\mathbf{P}_k$ [*Parrish and Cohn*, 1985], the sparsely banded approximation of $\mathbf{L}_k$ would preserve and smoothly taper the long-distance covariances [*Habibi*, 1972], as evident in Figure 3.

The time recursion for the truncated $\mathbf{L}_k$, initialized as $\mathbf{L}_0^a = \mathbf{L}_0$, is

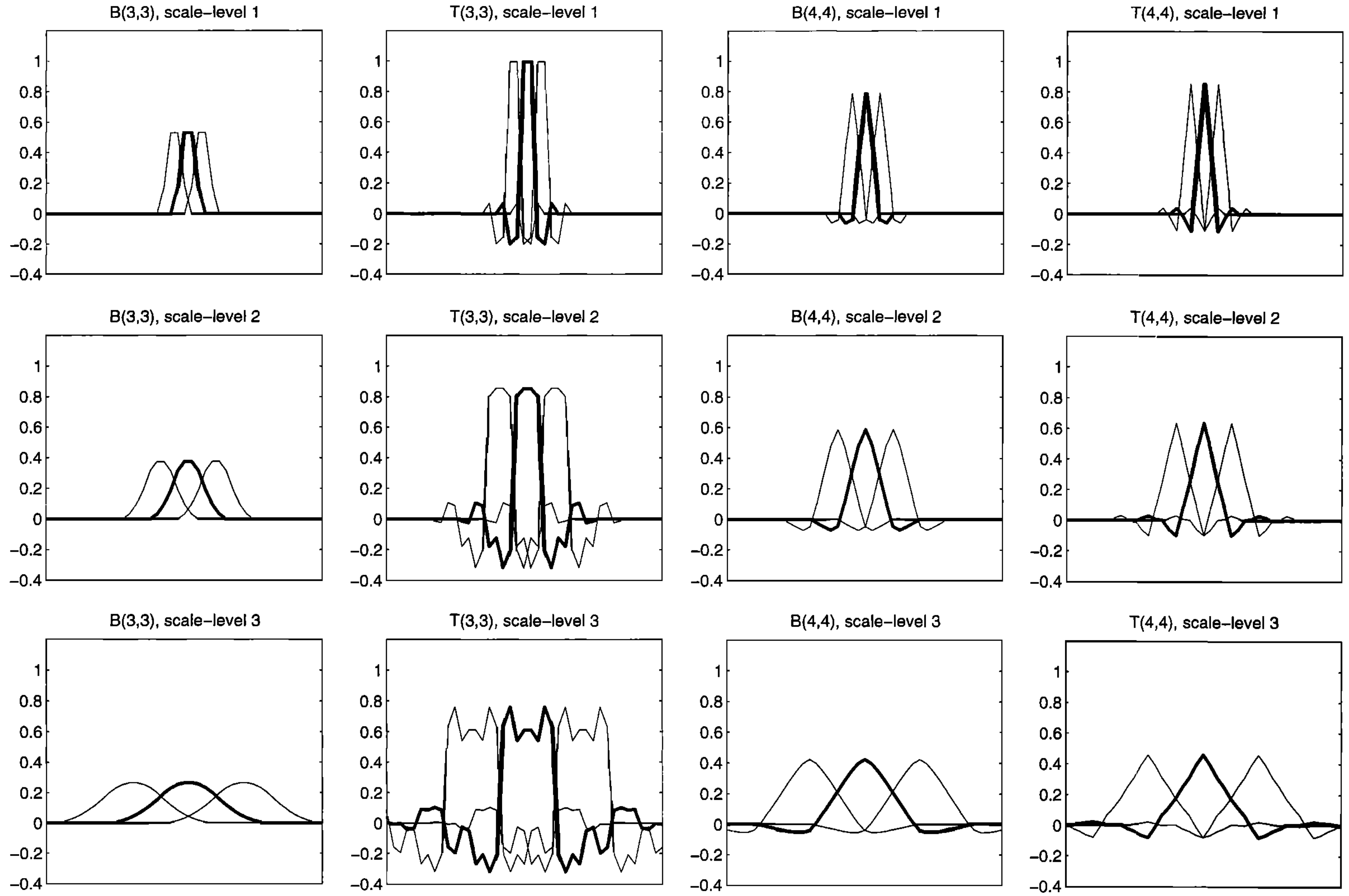


Figure 1. The two biorthogonal wavelet basis functions described in Appendix A and used in section 4. B(3,3) indicates the function $\tilde{\phi}^{3,3}$, while T(3,3), B(4,4), and T(4,4) denote functions $\phi^{3,3}$, $\tilde{\phi}^{4,4}$, and $\phi^{4,4}$, respectively. Three spatially adjacent basis functions are shown, with the middle function highlighted with thick curves. The functions with the length scales of $\ell = 1, 2$, and 3 are shown from top to bottom, respectively.

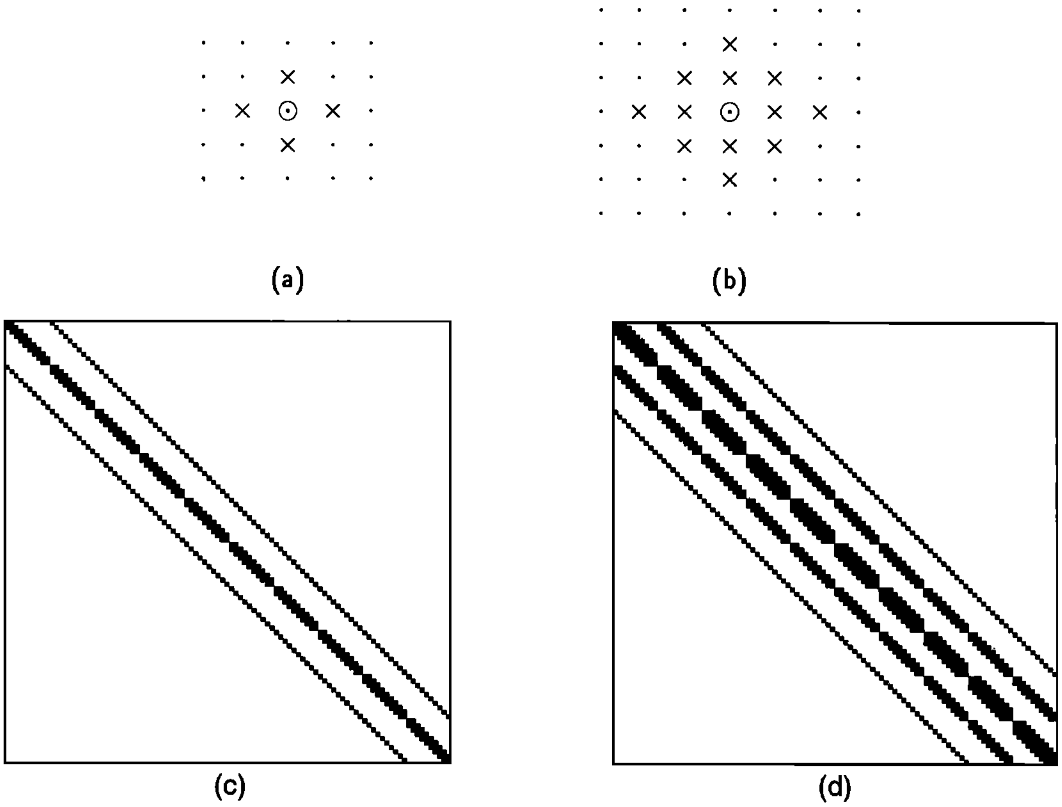

(a) (b) (c) (d)

Figure 2. The Markov random field stencils ("neighborhoods"), marked with crosses, for the (a) first- and (b) second-order configurations about the center dots (circled). The corresponding locations of nonzero elements in the information matrix, assuming a 10 × 10 two-dimensional domain leading to a 100 × 100 matrix dimension, for (c) the first-order case resulting in a block tridiagonal structure and (d) the second-order case resulting in a block pentadiagonal structure.

$$\mathbf{\Theta}_k = \mathbf{\Lambda}_k^{-1}\mathbf{F}_k^T\mathbf{M}_k \tag{16}$$

$$\mathbf{L}_k^f = (\mathbf{I} - \mathbf{F}_k\mathbf{\Theta}_k)^T\mathbf{M}_k(\mathbf{I} - \mathbf{F}_k\mathbf{\Theta}_k) + \mathbf{\Theta}_k^T\mathbf{L}_{k-1}^a\mathbf{\Theta}_k \tag{17}$$

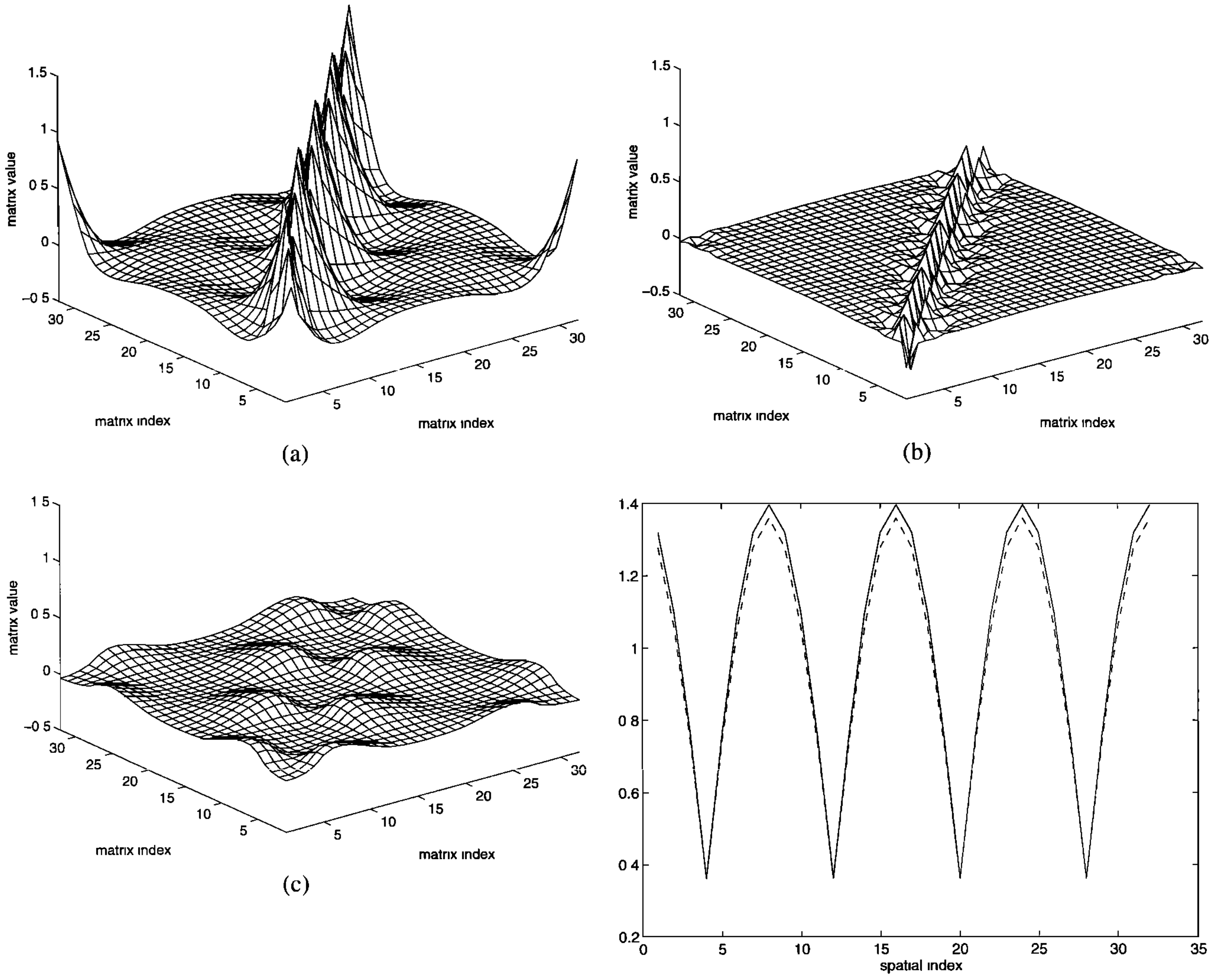
$$\mathbf{L}_k^a = \mathbf{L}_k^f + \mathbf{H}_k^T\mathbf{N}_k\mathbf{H}_k \tag{18}$$

where $\mathbf{\Lambda}_k$ is the diagonally truncated version of the matrix $\mathbf{F}_k^T\mathbf{M}_k\mathbf{F}_k + \mathbf{L}_{k-1}^a$. This recursion preserves the sparsely banded structure and positive-definite property in $\mathbf{L}_k$ (Appendix B). In particular, by using a diagonally truncated version of $\mathbf{\Theta}_k$ in the last additive term $\mathbf{\Theta}_k^T\mathbf{L}_{k-1}^a\mathbf{\Theta}_k$ of (17), the structure of $\mathbf{L}_k$ is constrained to be just as sparsely banded as the maximum structural extent of the symmetric matrix $\mathbf{F}_k\mathbf{F}_k^T\mathbf{M}_k^2\mathbf{F}_k\mathbf{F}_k^T + \mathbf{H}_k^T\mathbf{N}_k\mathbf{H}_k$ for all k. To ensure such sparseness, we assume that the system and parameter matrices $\mathbf{F}_k$, $\mathbf{H}_k$, $\mathbf{L}_0$, $\mathbf{M}_k$, and $\mathbf{N}_k$ are sparsely banded, an assumption usually satisfied in practice. For example, $\mathbf{F}_k$ is always local as it is a numerical version of a differential operator. Also, a spatially regular covariance matrix $\mathbf{Q}_k(\equiv \mathbf{M}_k^{-1})$ of modeling error, satisfying physical constraints such as geostrophy [*Jiang and Ghil*, 1993], can be specified using a locally structured weighting matrix $\mathbf{M}_k$ (see section B1). The optimal filter gain can be derived from this recursion as $\mathbf{K}_k = (\mathbf{L}_k^a)^{-1}\mathbf{H}_k^T\mathbf{N}_k$, in principle. As in the case with the subspace filter (10)–(12), however, an explicit computation of the gain matrix is not computationally efficient. Instead, by exploiting the imposed sparseness of the information matrix, the data-update step (5) is accomplished by computing $\mathbf{H}_k^T\mathbf{N}_k(\mathbf{y}_k^d - \mathbf{H}_k\mathbf{x}_k^f)$ before inverting it with the sparse matrix operator $\mathbf{L}_k^a$. In our numerical experiments (section 4), only about 20 of the standard Gauss-Seidel iterations are required for an effective inversion.

Note that the filter parameters in the recursion (16)–(18) are the weighting matrices $\mathbf{L}_0$, $\mathbf{M}_k$, and $\mathbf{N}_k$ of the least squares formulation (3), instead of the standard covariance parameters ($\mathbf{P}_0$, $\mathbf{Q}_k$, and $\mathbf{R}_k$, which would have to be inverted). In the MRF framework the filter parameters are thus more naturally interpreted as the balancing constraints (regression operators similar to one in (15)) applied to the model and data residual fields. More technical details of the MRF-based Kalman filter approximation can be found in Appendix B and references therein.

4. Numerical Experiments

Synthetic data assimilation problems have been formulated to observe the behavior of the approximate Kalman filter schemes in a controlled setting. Our approach is to first examine the filter performances in simple problems in which various controls are possible and then gradually increase the complexity for more relevance to realistic ocean data assimilation problems, such as incorporation of satellite altimetry data. The prognostic equations (4) in these problems are given by dynamic components of the Navier-Stokes equations with progressively less simplifications. The first set of experiments consists of univariate diffusion and advection dynamics in which the exact evolution of the prognostic variable is known analytically. In the second set of experiments a linearized shallow-



(a) (b) (c) (d)

Figure 3. Covariance approximation using the wavelet and Markov random field (MRF) approaches: (a) nonapproximated covariance matrix (over a one-dimensional periodic domain); (b) error resulting from approximation using the wavelet approach ($\tilde{\phi}^{3,3}$, $\ell = 1$); (c) error resulting from approximation using the MRF approach (second-order structure); and (d) diagonal elements (variances) of the original matrix (solid line), wavelet approximation (dotted line), and MRF approximation (dashed line). The periodic advection dynamics of equation (20) for $N = 32$ was used with four regularly spaced observation points and integrated over $k = 128$ time steps. The noise parameter was $q/r = 0.3$.

water equation over a periodic domain, as studied by *Ghil et al.* [1981], with a known analytic solution is used. In the final set, identical twin experiments with a reduced-gravity, shallow-water dynamics configured for an idealized midlatitude basin are conducted using the sampling pattern of the TOPEX/POSEIDON satellite. A multilayer version of this dynamic configuration has also been examined but will be discussed in a separate paper (in preparation) to elaborate on the vertical correlation structures due to a relatively implicit nature of cross-layer interactions in shallow-water systems.

Realization of the locality assumption is of particular interest here. The approximation schemes thus examined are spatial regression realized by the MRF-based information filter (section 3.2) and subspace projection using biorthogonal wavelets as the mode functions (section 3.1.2), hereafter referred to as the "MRF approximation" and "wavelet approximation," respectively. The biorthogonal wavelet functions used here are depicted in Figure 1, middle ($\ell = 2$, $n = N/4$) and described in Appendix A. The second-order MRF structure (a block pentadiagonal matrix, e.g., Figure 2d) is used in the MRF approximation. All computations in the experiments presented here have been performed on desktop workstations (300 MHz DEC-Alpha).

The filter parameters are given by the homogeneously diagonal matrices $\mathbf{M}_k = q^{-1}\mathbf{I}$ and $\mathbf{N}_k = r^{-1}\mathbf{I}$ or, equivalently, $\mathbf{Q}_k = q\mathbf{I}$ and $\mathbf{R}_k = r\mathbf{I}$. The correlation structures in $\mathbf{Q}_k$ and $\mathbf{R}_k$ are inherently dependent on the application (particularities of the model, data, and observation network). While using the uncorrelated parameter matrices may lead to slight computational advantages for their sparsity, it can also affect the regularity of the computed covariance $\mathbf{P}_k$ and hence present a stringent test for stability of the filter. Our study focuses on asymptotic filter performance, to which the effects of the initialization parameters $\mathbf{L}_0$ or $\mathbf{P}_0$ would become negligible with

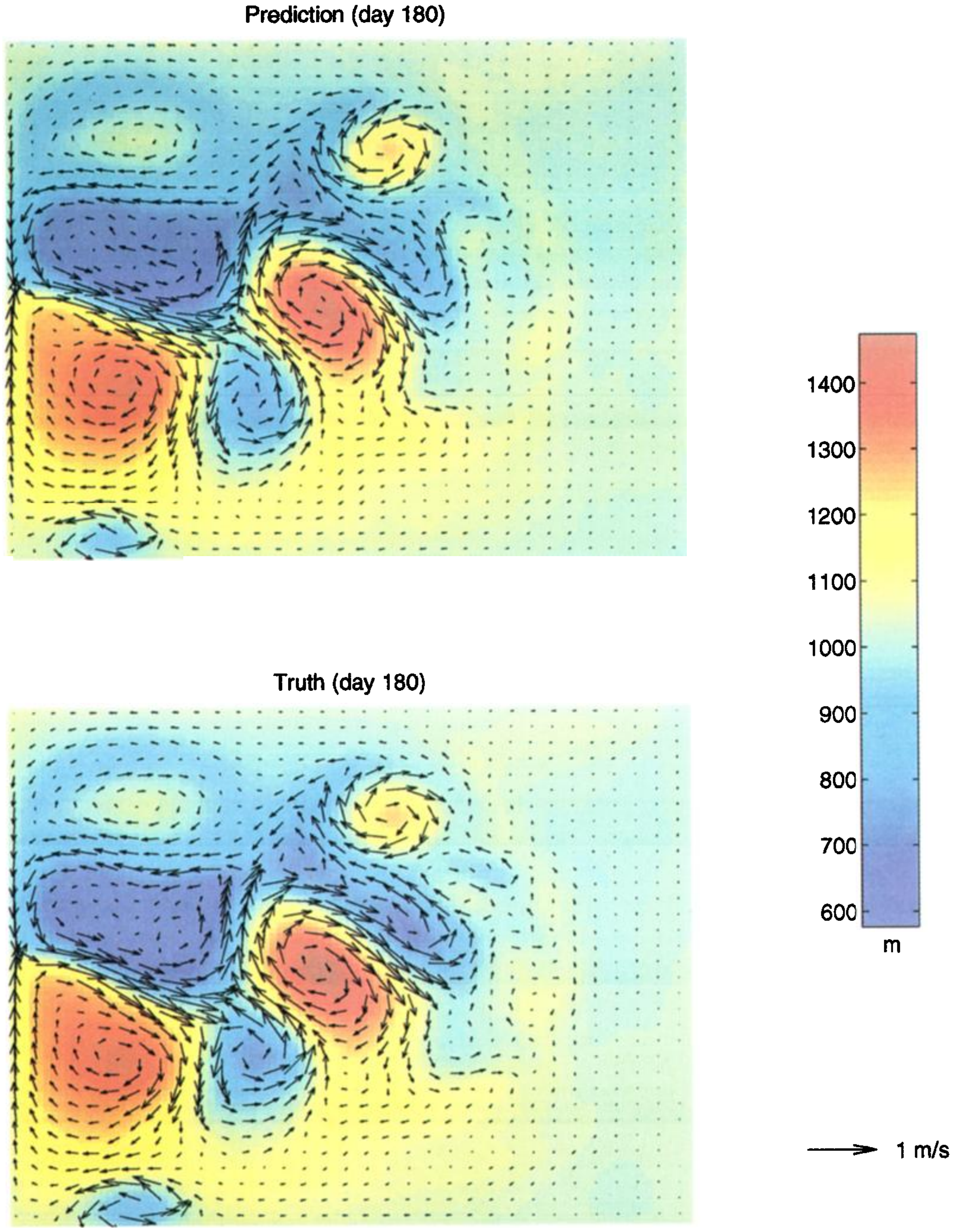


Plate 3. (top) Prediction fields after 180 days (12,960 time steps) of assimilating the h field from the (bottom) truth.

increasing time steps (except in degenerate cases such as $q = 0$). The effective filter parameters are therefore the scalars q^{-1} and r^{-1}, representing the respective weights of the model and data terms in the least squares formulation (3). Since only the relative weights affect the outcome, filter performance is actually controlled by a single scalar q/r.

4.1. Equations of Diffusion and Advection

4.1.1. Filtering problem. Simple numerical configurations are used to enable computation of the nonapproximated Kalman-filter estimates, which serve as a reference to evaluate the approximation schemes. Two equations of univariate diffusion and advection over a 1-D periodic spatial domain are thus considered and are given, for the dynamic variable $\eta(s, t)$, as

$$\frac{\partial}{\partial t}\eta = D\frac{\partial^2}{\partial s^2}\eta$$

$$\frac{\partial}{\partial t}\eta = -U\frac{\partial}{\partial s}\eta$$

which are discretized over space and time as

$$\mathbf{x}_k = \left(\mathbf{I} - D\Delta t\left[\frac{\partial^2}{\partial s^2}\right]\right)\mathbf{x}_{k-1} \tag{19}$$

$$\mathbf{x}_k = \left(\mathbf{I} + U\Delta t\left[\frac{\partial}{\partial s}\right]\right)\mathbf{x}_{k-1} \tag{20}$$

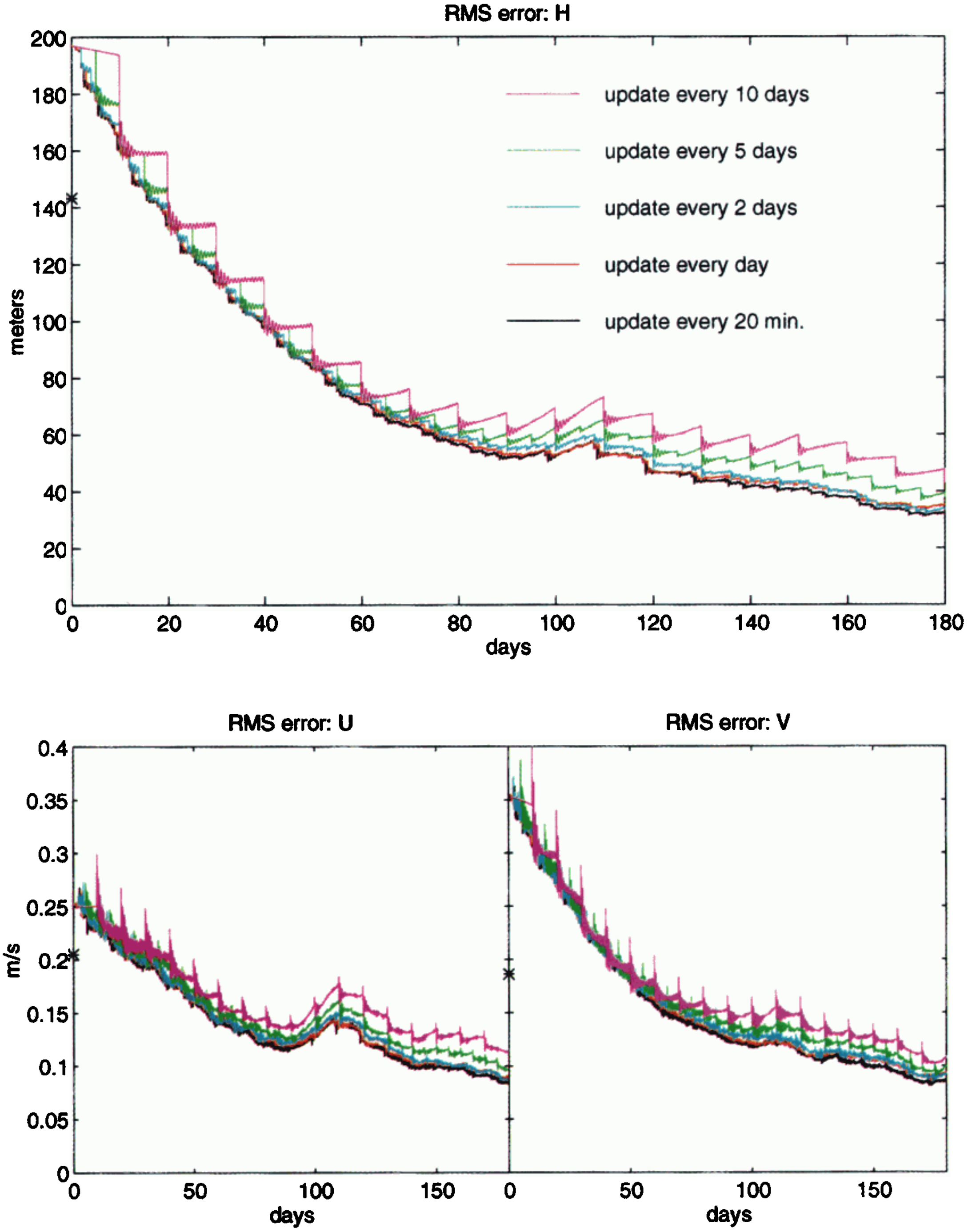


Plate 4. Time sequence of prediction error for the variables h, u, and v in the identical-twin experiment.

to yield the linear state transition matrix **F** for each system, where $[\partial^2/\partial s^2]$ and $[\partial/\partial s]$ denote periodic second- and first-order difference operators, respectively, and

$$\mathbf{x}_k \equiv [\eta(s_1, k\Delta t),\ \eta(s_2, k\Delta t),\ \cdots,\ \eta(s_N, k\Delta t)]^T.$$

The periodic spatial domain is discretized uniformly to $N = 128$ points. The diffusion coefficient and advection speed used in the experiments are

$$D = 0.2\,\frac{(\Delta s)^2}{\Delta t}$$

$$U = 0.2\,\frac{\Delta s}{\Delta t}$$

respectively, where Δs is the grid interval. Figure 4 (solid lines) shows the respective initial conditions for the diffusion and advection equations. Both initial fields are the results of truncated Fourier series, with which the analytic solutions can be obtained easily. Sparse observations of the state vector $\mathbf{x}_k$ are simulated by evaluating the analytic solutions at eight uniformly spaced locations (Figure 4, circles) for each k. These observed values are then corrupted by an additive noise to

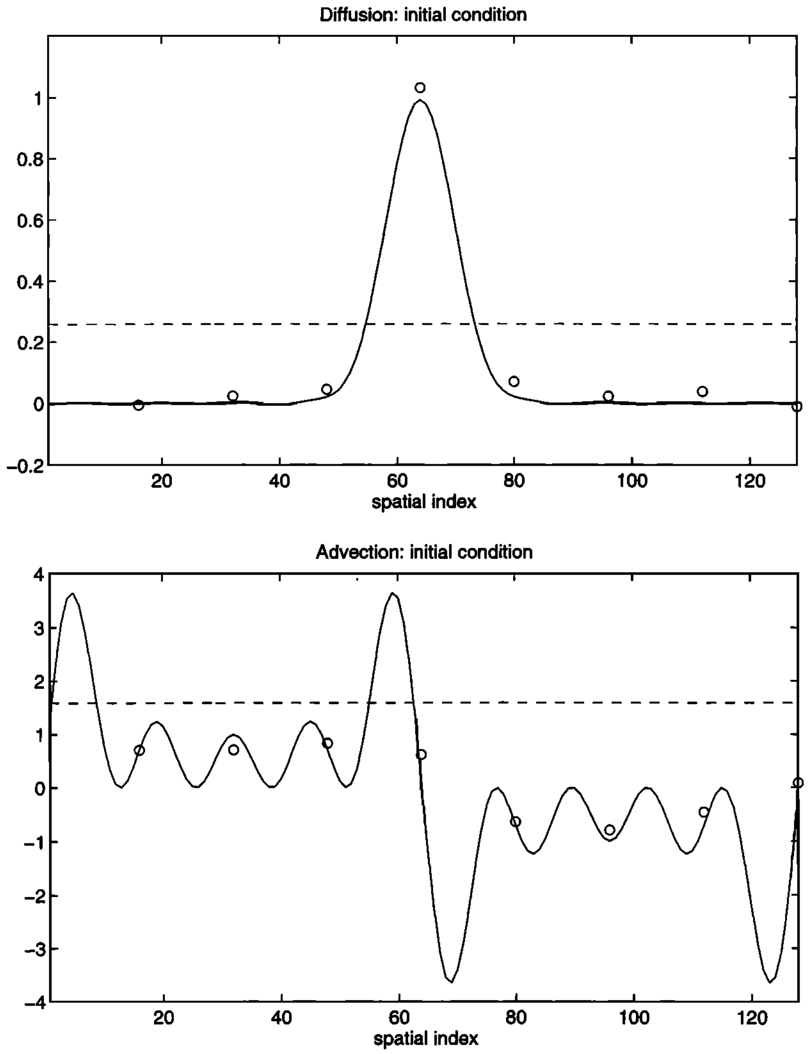


Figure 4. The initial states (solid lines) for the (top) diffusion and (bottom) advection experiments, their root-mean-square variability (dashed lines) used as the initial conditions in the flat cases, and the sample noisy observations (circles) at the fixed set of grid points.

simulate a set of noisy and sparse measurements $\mathbf{y}_k$. The observation noise has been numerically generated based on a Gaussian distribution with variance equal to 1% of the signal variability (10% of standard deviation) in each of the two dynamic recursions.

Two extreme cases of filter initialization are examined. In one case the initial estimate $\mathbf{x}_0^d$ is perfect (i.e., completely sampled and noiseless), as given by the solid curves in Figure 4. In the other case the initial estimate is just a flat line (Figure 4, dashed lines). These two types of initial conditions combined with the two dynamic equations provide four cases of filtering problems, hereafter referred to as "diffusion perfect," "diffusion flat," "advection perfect," and "advection flat." In all cases the recursions for the filter covariance (and information) matrix are initialized as $\mathbf{P}_0 = \mathbf{L}_0 = \mathbf{I}$.

The estimates $\mathbf{x}_k^a$ from three implementations of the Kalman filter, the "optimal" (nonapproximated), "MRF-approximated," and "wavelet-approximated" filters, are compared to the analytic solution in each of the four cases described above. The wavelet-approximated filter here has been implemented with the mode function $\tilde{\phi}^{3,3}$ (Appendix A, $\ell = 2$). The root-mean-square (RMS) error, using the analytic solution as the truth, over the last 20 of the 200 time steps is computed for each filter run as a measure of steady state performance. Each of the two filter parameters q and r is set to one of five values, 10^{-2}, 10^{-1}, 1, 10, 100, for a total of 25 parameter combinations encompassing a wide range of data assimilation scenarios. In particular, the ratio q/r ranges from 10^{-4} to 10^{4}.

4.1.2. Results and discussions. The three Kalman filter implementations (with and without approximations) displayed qualitatively identical behaviors under all numerical scenarios examined here. The long-term performance of the filters, as measured by the steady state RMS error, depends only on the parameter ratio q/r and not on the initial filter conditions, in general. In particular, the arbitrary and inaccurate "flat" initial estimates (Figure 4, dashed lines) can be compensated for by the noisy and sparse data $\mathbf{y}_k$ over time. Figure 5 shows that the final estimates (at $k = 200$) for the diffusion-flat and advection-flat cases are reasonably good reproductions of the analytic solution, considering the sparseness and inaccuracy (noise) of the data. All three filter implementations have produced very similar estimates, displaying no degradation in

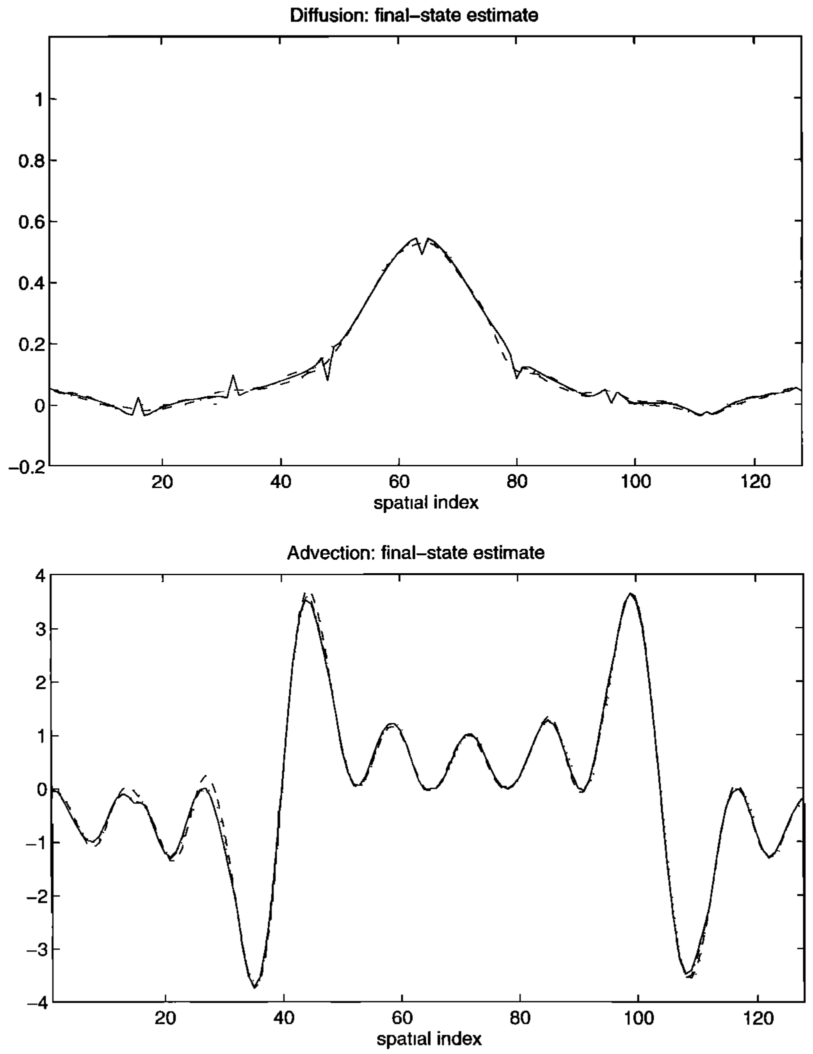


Figure 5. Examples of the final estimates (at time step $k = 200$ and with the filter parameters $q = r = 1$) for the optimal Kalman filter (solid lines), the wavelet-approximated filter (dashed lines), and the MRF-approximated filter (dash-dotted lines) in the (top) diffusion-flat and (bottom) advection-flat cases. The analytic solution for each case is given as the dotted lines. Note that the lines are difficult to distinguish owing to overlapping, indicating good reproductions of the analytic solutions by all three filters.

qualitative accuracy due to the approximations of the Kalman filter. In the diffusion-flat case both approximated filters have actually produced more desirable (regular) estimates owing to the imposed smoothness that compensates for our use of diagonal parameter matrices (e.g., $\mathbf{Q}_k$).

For each filter (optimal, MRF, or wavelet) and assimilation problem (one of the four cases of diffusion/advection–perfect/flat) a steady state RMS error value is computed for each of the 25 combinations of the filter parameters (q, r, and the initial variance). It is observed that, among the 25 parameter combinations for a given assimilation problem, the RMS values are nearly identical when the parameter ratio q/r has the same value. This "model-to-data" variance ratio defines the relative weights assigned to the model and data in the least squares formulation (3). A small value of the q/r ratio implies relatively high trust (weight) in the model, while a large value implies more trust in the data. Figure 6 shows a plot of the average RMS error (as a percentage of the signal variability in the initial states, Figure 4) as a function of q/r, varying over a wide range from 10^{-4} to 10^4, for each filter implementation and assimilation case. It can be observed in Figure 6 that, for each and every assimilation case, the error functions for all three filters have very similar characteristics, giving further evidence that the approximated filters (MRF and wavelet) are performing near optimally.

Although the model-to-data ratio q/r is effectively the only filter parameter in the problems examined here, the long-term filter performance is relatively insensitive to the value of q/r, except when q/r is extremely small (when the data are weighed negligibly compared with the model). This can be observed in each RMS error plot of Figure 6 as the almost-constant error value for $q/r > 0.01$. In particular, the RMS errors are, at most, 10% for $q/r > 0.01$, indicating that an estimation quality similar to one displayed in Figure 5 (corresponding to $q/r = 1$) can be expected over a wide range of values for q/r.

The RMS error plots (Figure 6) display distinct characteristics for each of the four cases of data assimilation scenarios. In the diffusion-perfect case, lower values of the model-to-data ratio q/r are associated with more accurate estimates (lower RMS values) because of fidelity in the initial condition and

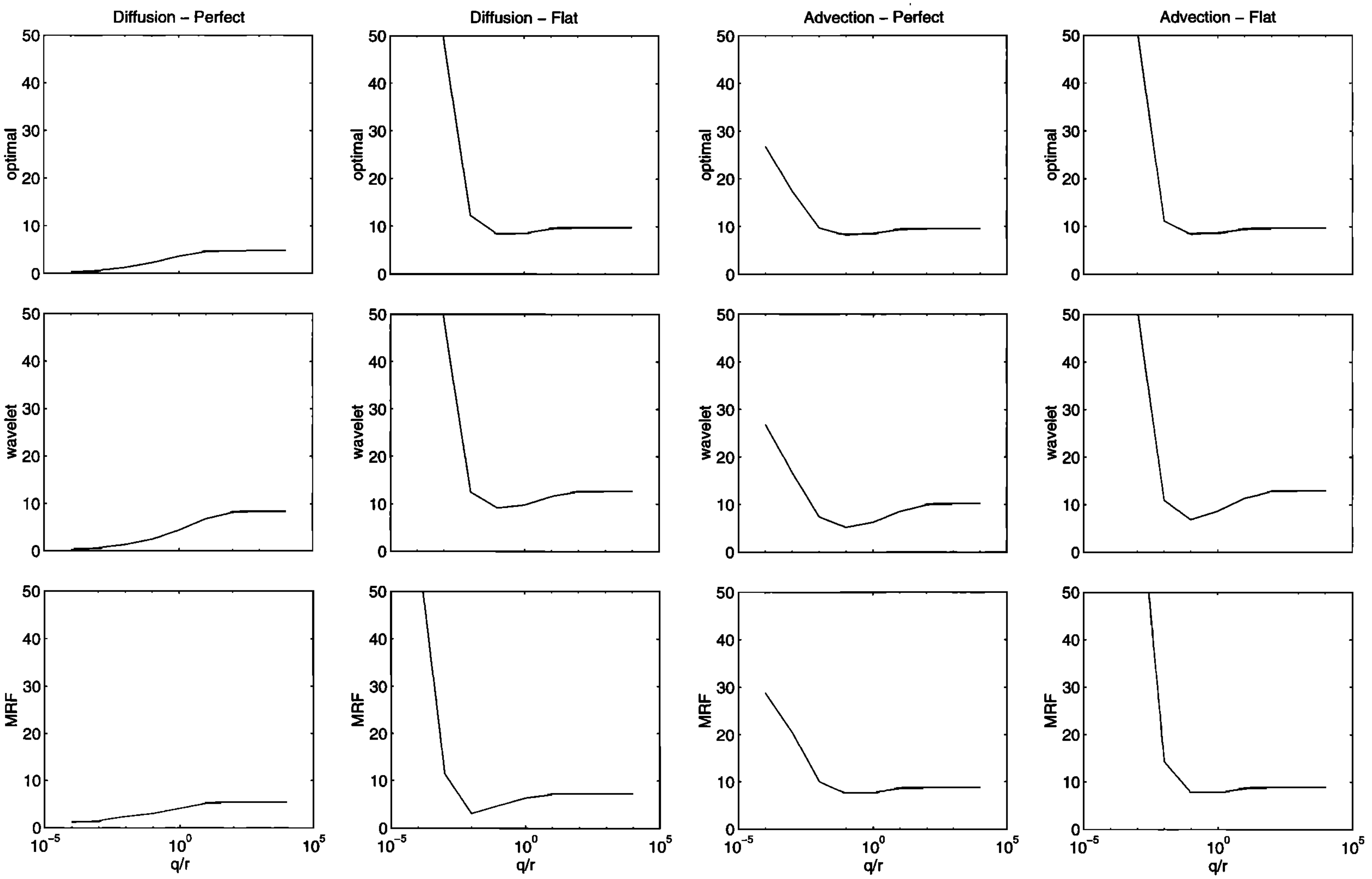


Figure 6. The root-mean-square (RMS) steady state estimation error plotted as a function of the model-to-data variance ratio q/r for each of the four diffusion/advection cases. The relative trust in data increases from left to right along the horizontal axis, which uses a logarithmic scaling.

model numerics. For higher q/r values, implying increasing trust in data, the estimation accuracy decreases owing to the measurement noise. On the other hand, in the diffusion-flat case, low q/r values lead to high RMS errors and even filter failures, as the effects of the inaccurate (flat) initial condition persists owing to the implied overreliance on the model. This suggests that the model-variance parameter q, controlling the retention time for information in the filter, needs to be reasonably high relative to the data variance r to ensure that the filter remains sensitive to the data [*Lewis*, 1986; *Miller et al.*, 1994]. The same RMS error characteristics can also be observed in the advection-flat case for the same reason. Unlike the diffusion-perfect case the advection-perfect case does not yield good estimates when the q/r ratio is small, implying inadequacy in the numerical advection scheme (an implementation of Lax-Wendroff scheme) for this particular configuration with relatively high advection speed and undulating initial condition. For a higher q/r the noisy data have actually compensated for the numerical biases introduced by the advection scheme. *Miller* [1986] reached a similar conclusion that Kalman filters can compensate for imperfect numerics and can even stabilize an unstable numerical scheme.

The advection-flat case is the most difficult of the four data assimilation problems examined here owing to the combination of wrong initial condition and high spatial variability. In this case the subspace projection approach (the wavelet-approximated filter) is found dependent on the choice of subspace basis functions. For example, a slightly different biorthogonal wavelet function $\tilde{\phi}^{4,4}$ (Figure 1, $\ell = 2$), as well as the cubic B spline and several other spline functions, has lead to an instable and inaccurate filter. The performance of subspace projection filter has been further examined using SVD and EOF subspaces (see section 3.1.1). The EOFs used here have been obtained from the empirical covariance matrix of signal variability in the analytic solution over 500 time steps. As depicted in Figure 7, all the singular values of the linear dynamics **F** have comparable magnitudes, as the advection dynamics have no mode selectivity. (Some singular values corresponding to high-wavenumber modes are clipped slightly owing to discretization effects.) The SVD subspaces are thus not expected to offer compact approximations in this case. For example, a subspace as large as $n = 48$ has been observed necessary to produce estimates that begin to resemble the analytic solution qualitatively. Unlike the singular values the EOF variances are dominated by the first 10 modes, as shown in Figure 7. The subspace using the corresponding 10 EOFs leads to a stable filter that produces estimates with a constant bias (Figure 7). This error has resulted because none of the 10 EOF modes is able to represent (and hence correct for) the constant bias in the flat initial condition (Figure 4). The fundamental difficulty is that the particular empirical statistics used to obtain the EOFs do not contain any constant bias owing to the method of ensemble simulation. Such deficiency in the EOF subspace may be difficult to assess in practice. In our idealized scenario, however, an ad hoc addition of an eleventh basis function of a constant (a normalized vector of all ones) would adequately augment the EOF subspace, leading to an accurate, stable, and efficient ($n = 11$) approximate filter for the advection-flat case. This illustrates the potential for high computational efficiency as well as strong dependency on subspace selection for the EOF-based method.

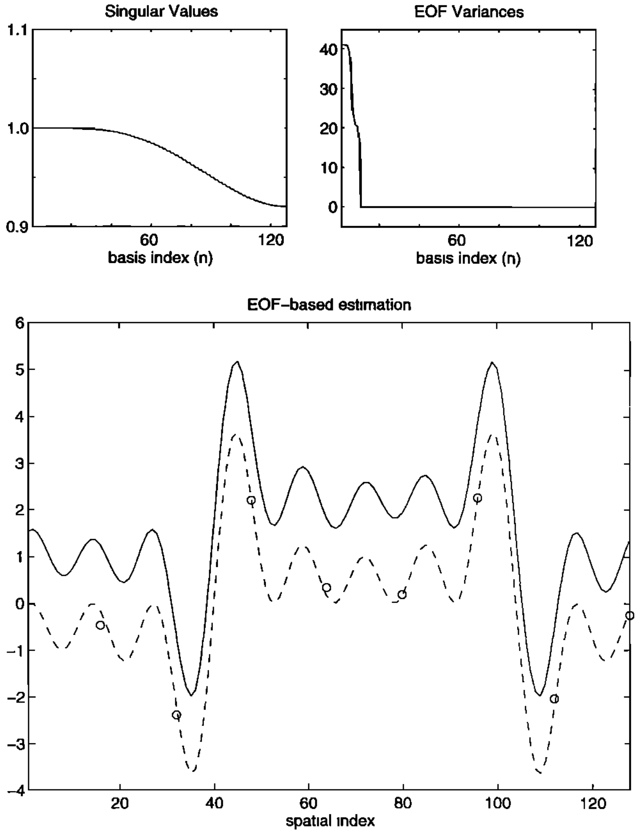


Figure 7. (top left) Singular values of the 128 × 128 state-transition matrix **F** in the advection case. (top right) Variances (eigenvalues) of an empirical variability matrix in the advection case. (bottom) Sample final estimate (solid line, at $k = 200$) for a subspace-projection filter using the subspace of the 10 dominant empirical orthogonal functions (EOFs) (eigenvectors corresponding to the large variances in Figure 7, top right). The corresponding analytic solution or the "truth" is shown by dashed line, while examples of the noisy observations (at $k = 200$) are shown by circles. Note that the set of 10 EOFs cannot correct for the constant bias, despite the information available in the observations.

In all approximate filtering approaches the absolute covariance values are too sensitive to operating conditions for a reliable approximation of the second-order statistics; however, the correlation structures (i.e., normalized covariance) are approximated well. Figure 3 depicts representative covariance matrices from optimal Kalman filter and covariance approximation errors by the MRF and wavelet filters. The MRF-based approach seems to approximate local (near diagonal) correlation better, while the wavelet-based (and other subspace projection) approach tends to capture the far-field correlation structure better than local correlation structure.

4.2. Linearized Shallow-Water Model

A multivariate dynamic system is used to observe the performance of the optimal, wavelet-approximated, and MRF-approximated Kalman filters. The data assimilation problem examined here is nearly identical (except for some differences in discretization and parameterization) to the configuration studied by *Ghil et al.* [1981]. The model here is given by a synoptic-scale atmospheric dynamics expressed by the linearized shallow-water equation

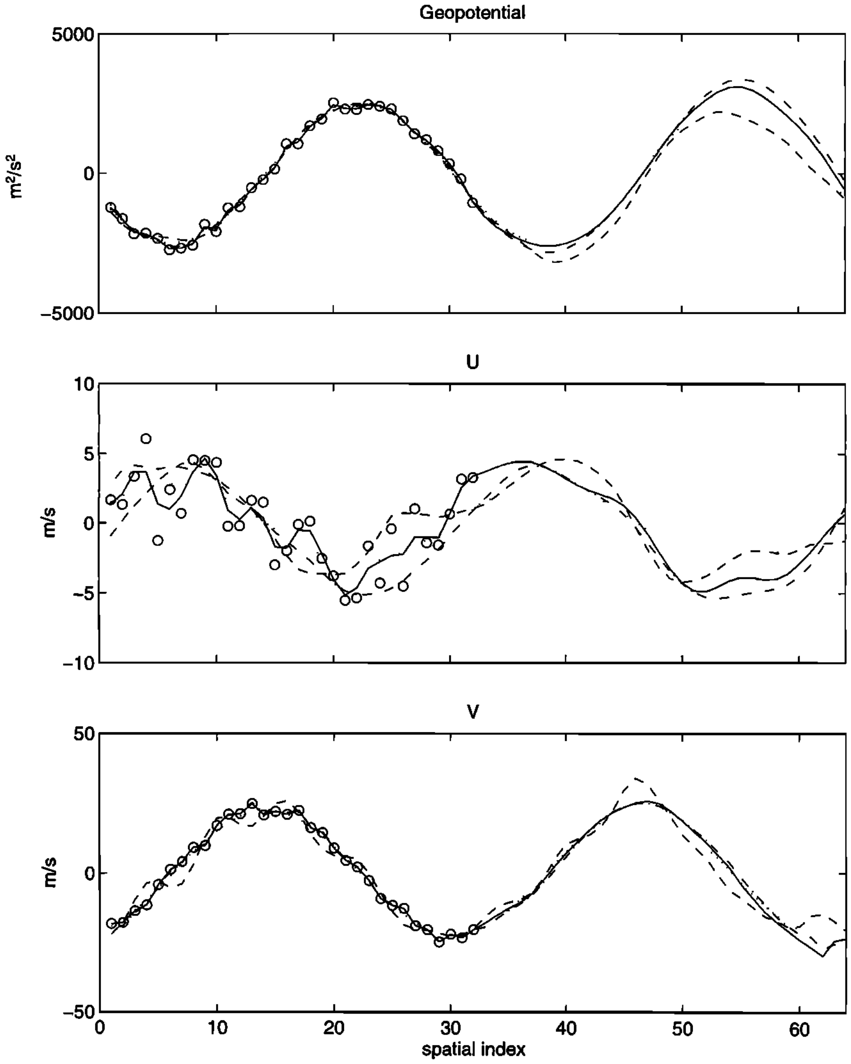


Figure 8. Examples of final estimates ($k = 1344$) computed with the optimal Kalman filter (solid lines), wavelet-approximated filter (dashed lines), and MRF-approximated filter (dash-dotted lines), along with the analytic solution truth (dotted lines) and noisy observations (circles) made every 48 time steps for (top) geopotential, (middle) zonal wind speed u, and (bottom) meridional wind speed v.

$$\frac{\partial}{\partial t}\begin{bmatrix} u \\ v \\ \phi \end{bmatrix} = \begin{bmatrix} -U\dfrac{\partial}{\partial s} & f & -\dfrac{\partial}{\partial s} \\ -f & -U\dfrac{\partial}{\partial s} & 0 \\ -\Phi\dfrac{\partial}{\partial s} & fU & -U\dfrac{\partial}{\partial s} \end{bmatrix}\begin{bmatrix} u \\ v \\ \phi \end{bmatrix} \tag{21}$$

over a periodic spatial domain along a fixed latitude, where the state (prognostic) variables are the variability in the zonal and meridional wind speeds (u, v) and the geopotential ϕ. The mean zonal flow, mean geopotential, and ambient Coriolis parameter are given as $U = 2$ m/s, $\Phi = 30000$ m^2/s^2, and $f = 0.0001$ s^{-1}, respectively. An analytic solution of this equation, along with other details of the experimental setup, are given by *Ghil et al.* [1981]. The spatial domain is discretized to $N = 64$ points. The spatial and temporal intervals are $\Delta s = 218.75$ km and $\Delta t = 15$ min, respectively. The measurements of the state variables u, v, ϕ are made by sampling the analytic solution only at the 32 grid points on the left side of the domain (Figure 8, reflecting land-based measurements), only twice daily (every 48 time steps). The measurements are artificially corrupted with simulated Gaussian random numbers with standard deviations of 200 m^2/s^2 for ϕ and 2 m/s for u and v.

The filter parameter matrices ($\mathbf{P}_0$, $\mathbf{Q}_k$, and $\mathbf{R}_k$) used here are diagonal (uncorrelated) as before, and their variances (diagonal elements) are homogeneous for each variable type (u, v, or ϕ). The variances in $\mathbf{R}_k$ are given by the actual measurement noise variances, as above. The variances in $\mathbf{P}_0$ and $\mathbf{Q}_k$ are set to be 1% (10% root-mean-square) of the spatial variability in the analytic solution for each of the three variables. The resulting values for the model-to-data ratio (q/r) are about 0.02 for u and 1 for v and ϕ.

Like the advection-flat case examined previously, the numerical model here is marginally stable owing to the large dynamic-range mismatch among the three prognostic variables (e.g., ϕ oscillates at an order of 1000 times larger amplitude than u) combined with the spatial differentiations in (21). Without data assimilation the time trajectories of the variables do indeed deviate from the analytic solutions systematically (Figure 9). Steady state estimates (at $k = 1344$ or 14 days)

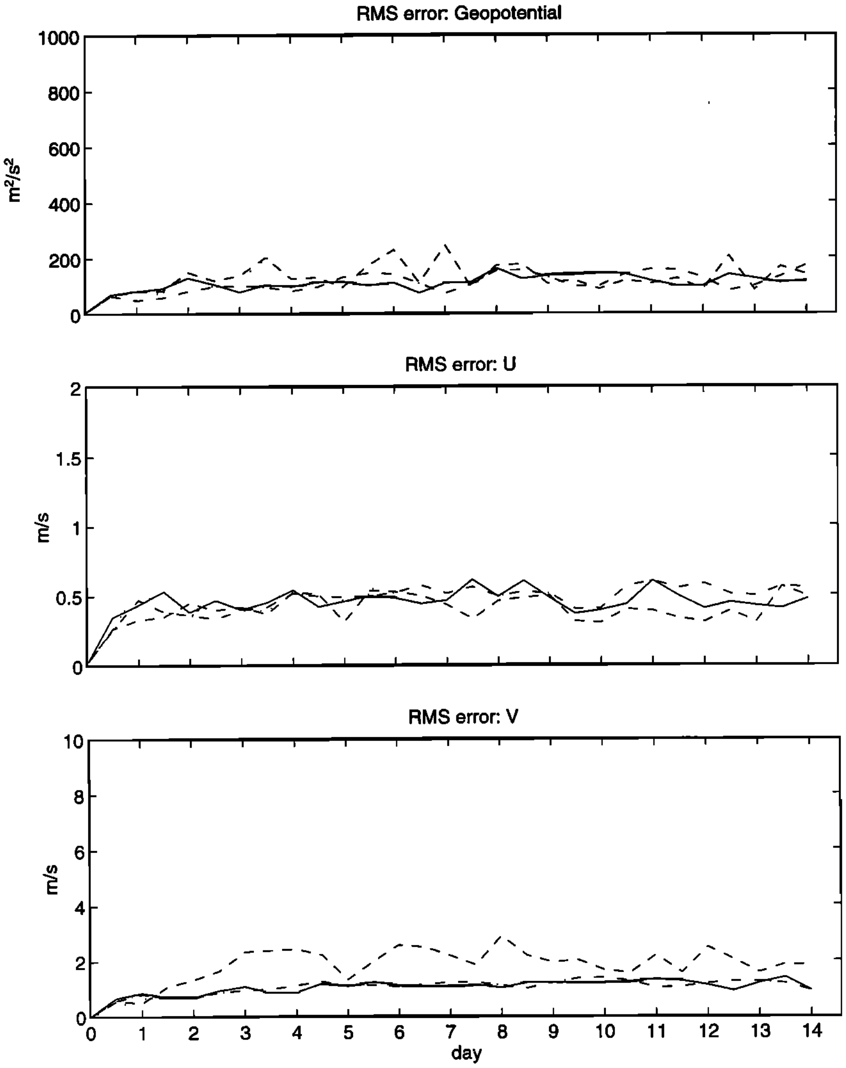


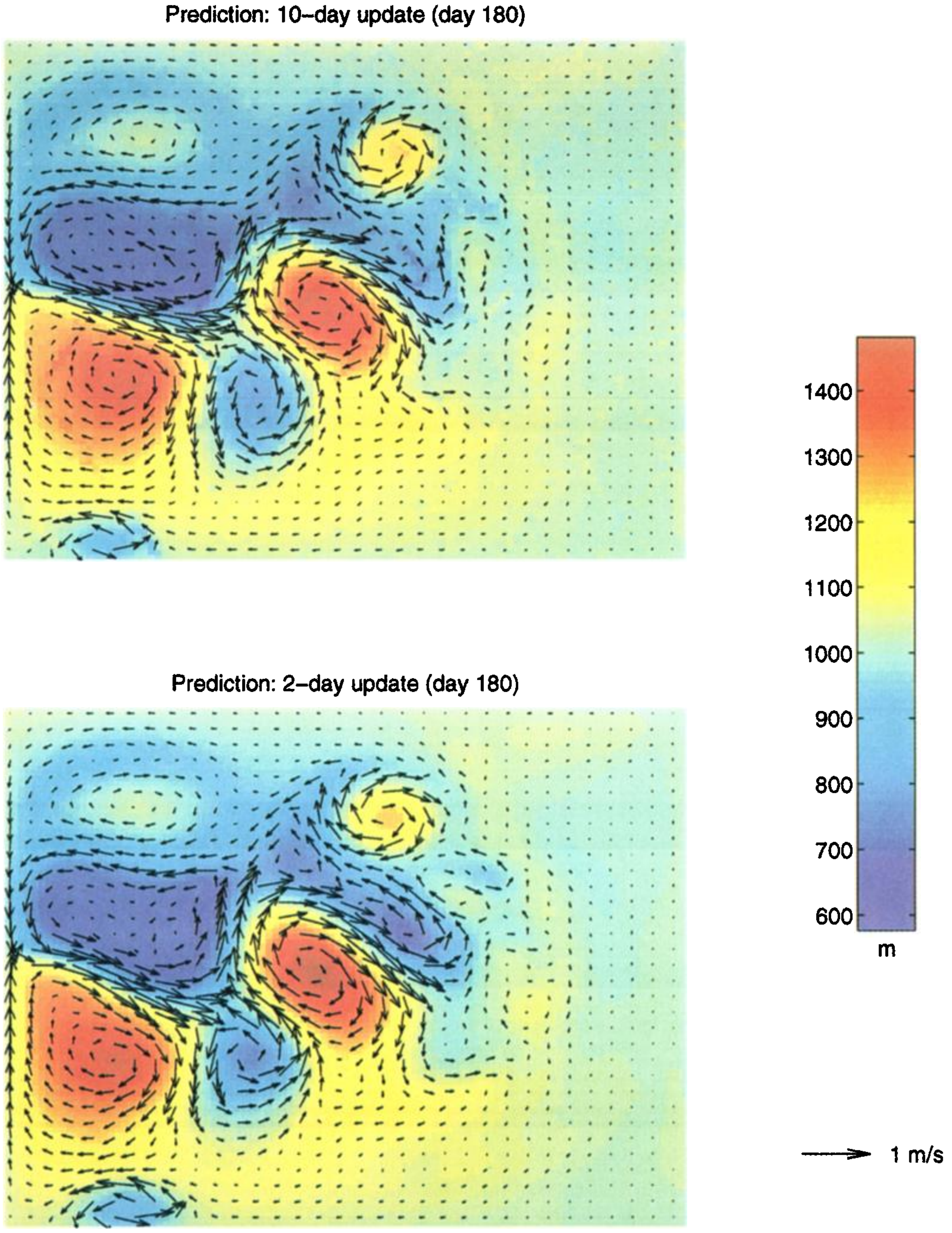
Figure 9. Time series of the root-mean-square errors between the analytic solution and the estimates by the optimal Kalman filter (solid lines), wavelet-approximated (dashed lines), MRF-approximated (dash-dotted lines), and model run without data assimilation (dotted lines) for (top) geopotential, (middle) zonal wind speed u, and (bottom) meridional wind speed v.

with data assimilation by the three filters are displayed in Figure 8, indicating again that the noisy and sparse data can compensate for the imperfect model numerics and that the approximate (wavelet and MRF) filters are just as effective for this purpose as the nonapproximated Kalman filter.

The time series of RMS errors, displayed in Figure 9, show that the filter errors saturate quickly, to values less than the corresponding measurement errors (i.e., less than the standard deviations given above). The RMS error plots are similar for the MRF and optimal filters, implying near-optimal performance of the MRF-based approximation. The wavelet filter, however, has produced estimates with systematically higher RMS errors, as evident from the time series (Figure 9) for v and, less obviously, ϕ. The biorthogonal basis function $\tilde{\Phi}^{4,4}$ ($\ell = 2$) has been used to construct the subspace for the wavelet filter examined here. The basis function $\tilde{\Phi}^{3,3}$, which has been used successfully in the diffusion/advection cases discussed previously, is found to cause unstabilities in the multivariate estimates examined here. These again demonstrate sensitivity of the subspace projection approach to the selection of the subspace modes.

4.3. Eddy-Resolving Shallow-Water Model

The MRF-approximated filter is now applied to a wind-forced, nonlinear, reduced-gravity, shallow-water model configured in a rectangular domain. The numerical implementation of the dynamic model is a single-layer version (without thermodynamics) of the Miami Isopycnal Coordinate Ocean Model (MICOM) [*Bleck et al.*, 1992; *Bleck and Chassignet*, 1994]. With a horizontal grid spacing of 20 km and grid size of 100 × 100, this particular model configuration has been used to simulate mesoscale features like rings associated with the free-jet portion of the western boundary currents [*Chassignet et al.*, 1990]. Since the multivariate prognostic state consists of the horizontal current velocity (u, v) and layer thickness h, the state dimension is about $N = 30{,}000$. This dimension is relatively small for a typical ocean general circulation model but is already too large to be practical for the nonapproximated Kalman filter. The MRF-approximated filter uses the second-order MRF configuration shown in Figures 2b and 2d, with which the information matrix is represented by approximately



Prediction: 10–day update (day 180)

Prediction: 2–day update (day 180)

Plate 5. Day-180 prediction fields as in Plate 3, except using the data-update intervals of (top) 10 and (bottom) 2 days.

$39N$ parameters. (This number is approximate due to the staggered (C grid) discretization used by MICOM. Also, $39N = 13 \times 3 \times N$, where 13 is the number of parameters per grid point in the second-order MRF (Figure 2b) and 3 is the number of variable types.) A similar level of compactness in covariance representation can be achieved with the wavelet-approximated filter (using the length scale $\ell = 2$); however, the wavelet approach is not examined here owing to the sensitivity to basis selection as observed in previous experiments.

4.3.1. Twin experiment with the TOPEX/POSEIDON sampling pattern. We have conducted a set of "identical-twin" experiments, in which one of two model runs with dynamically independent initial conditions is treated as the "truth," while the other run, called "prediction" hereafter, assimilates simulated measurements from this truth. The tenth and eleventh year estimates of a spin-up run are used to initialize the prediction and truth, respectively. The initial conditions for the truth and prediction are shown in Plate 1. To simulate (under the framework of the simplified model configuration) satellite-based altimeter measurements, the layer thickness h (which is directly proportional to the sea-surface height in the reduced-gravity model) have been sampled from the truth along a sequence of the TOPEX-POSEIDON sampling swaths with a 10-day repeat cycle, as shown in Figure 10. Each satellite sampling point has been collocated to the nearest model grid point. Since the RMS difference in h values between adjacent grid points in the truth field is about 7.2 m and the full-field variability of the same field is 77 m, the collocations have intro-

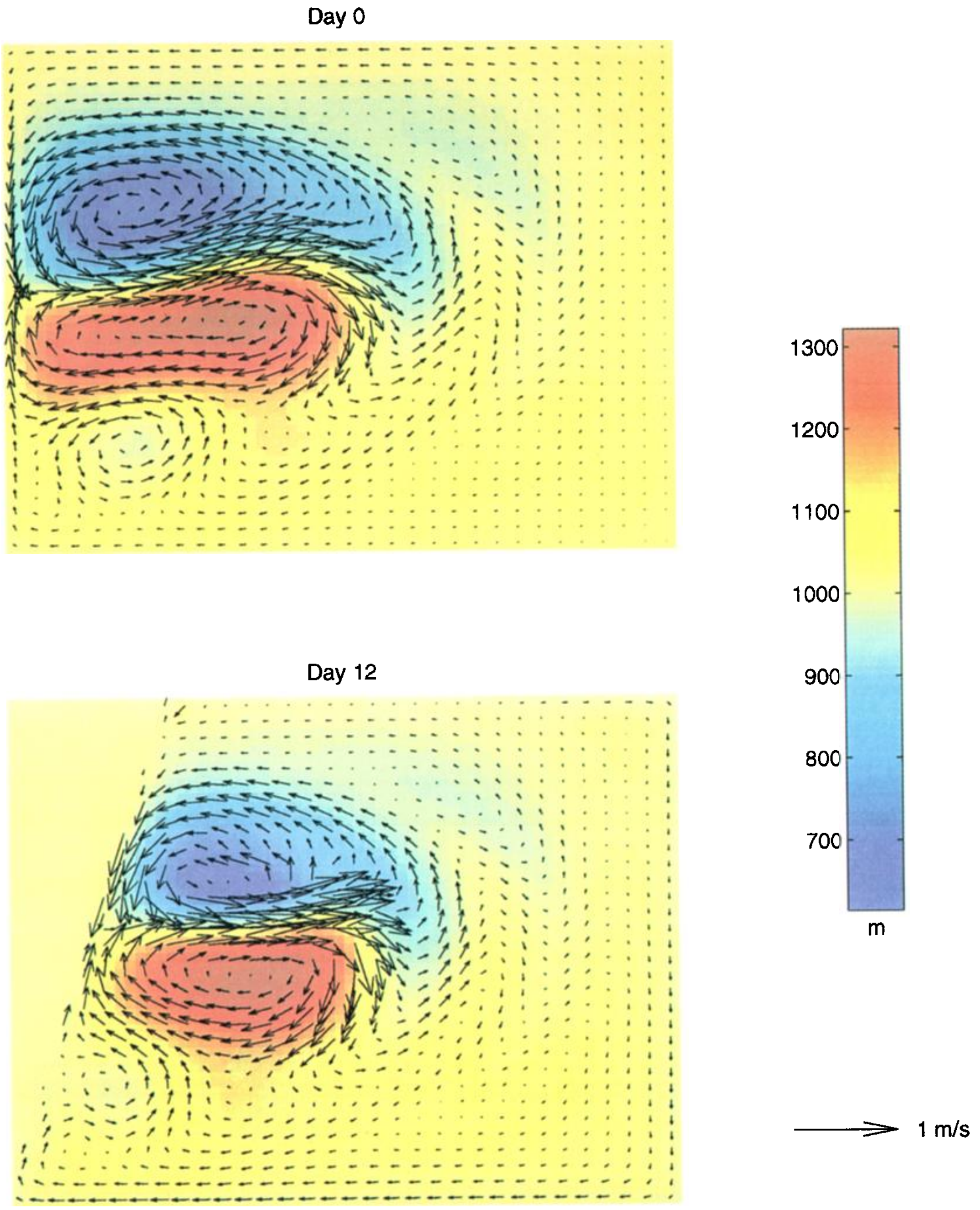


Plate 6. (top) Initial and (bottom) final states of a generalized data assimilation problem in which the "data" are the weak constrains simulating "land" along the left edge of the domain (visible as the triangular shape in Plate 6, bottom). Every two velocity vectors, in each of the horizontal and vertical directions, are shown.

duced about 10% (≈7.2/77) signal uncertainty, on average. The layer-thickness samples have thus been artificially corrupted with a 10%-RMS additive Gaussian noise to simulate the collocation effects. This noise level is also in rough agreement with the data noise (4/40 cm) expected across mesoscale features in areas such as the free-jet portion of the Gulf Stream. We note, however, that realistic TOPEX/POSEIDON data noise can be higher (4/9 cm) in less topologically active (low signal) ocean basins and that the along-track correlation in data noise is ignored while lumping these measurement and collocation errors together.

At a time step Δt of 20 min, the twin experiment has been performed for 360 simulated days, during which the prediction computed by the MRF-approximated filter has remained stable throughout. As before, diagonal parameter matrices are used for the filter, with the model-to-data ratio $q/r = 10$ for the variable h. Plate 2 shows an instance of prediction at day 30 and the corresponding truth, displaying a transition period during which the MRF filter attempts to track the truth, given the sparse and noisy observations of the h field. Plate 3 shows that by day 180 the prediction has been able to reproduce essentially all the mesoscale features in the truth. The RMS difference (prediction error) between the truth and prediction (Plate 4, black lines) displays an exponential decay in time, with an e-folding scale of about 30 days. These results indicate that a half-year to year sequence of the TOPEX/POSEIDON

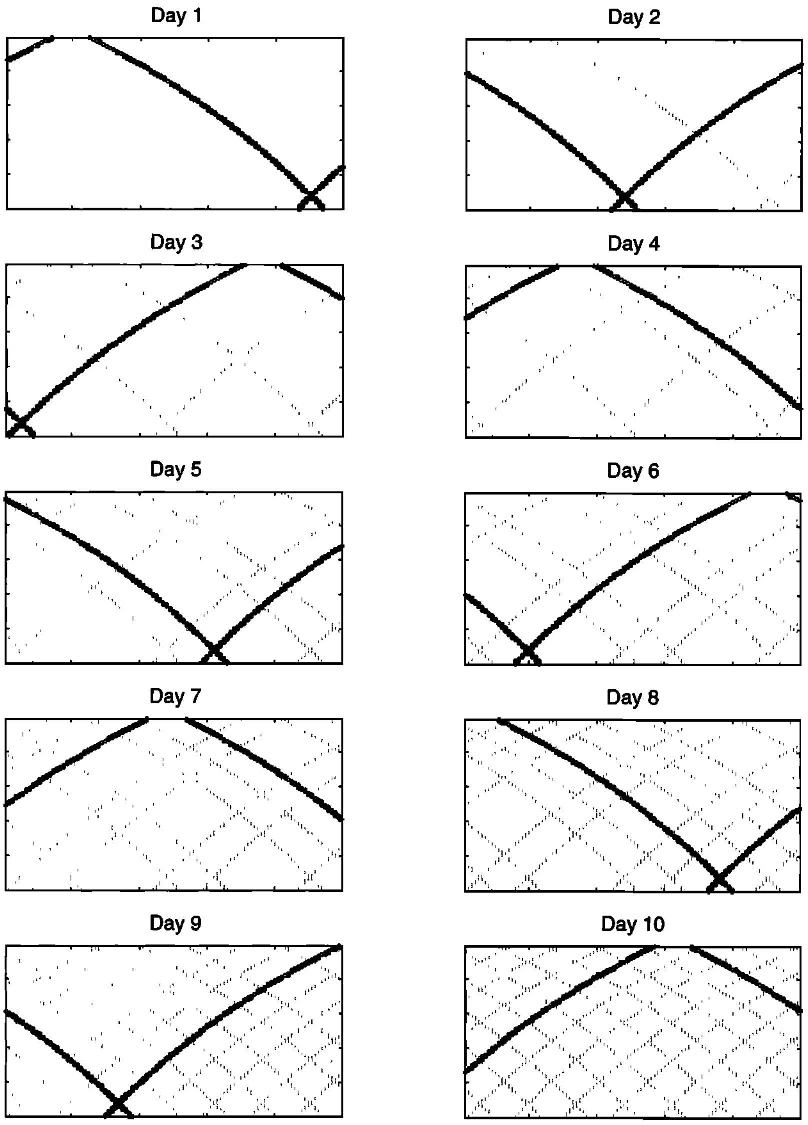


Figure 10. TOPEX-POSEIDON satellite tracks over the 100 × 100 grid (at 20 km grid size) used for the Miami Isopycnal Coordinate Ocean Model (MICOM) shallow-water model. Circles represent the observed grid points for a given day. Stippled lines denote all the tracks from the previous days.

altimetry data can be used to estimate the entire prognostic state of a reduced-gravity, shallow-water dynamics for reproduction of mesoscale and possibly larger-scale features.

4.3.2. Data-update frequency. The duration of each TOPEX/POSEIDON satellite swath over the model domain is well within the model time step of 20 min. The data updates are thus performed only two or three times per day, corresponding to the number of swaths per day (Figure 10). For computational economy, as well as for possible dynamic benefits of increased data density, we have considered performing data update (5) at an interval significantly larger than Δt, by accumulating the measurements during that interval and treating the accumulated data set as "synoptic." An update interval equaling the repeat cycle of 10 days is particularly attractive, not only for the near-regular coverage during the assumed synoptic interval (Figure 10) but also for possible computational saving due to a steady state filter approximation [e.g., *Fukumori and Malanotte-Rizzoli*, 1995] enabled by the time independence in data pattern over this interval. On the other hand, large update intervals can undersample dynamic mesoscale features such as advecting rings [*Chassignet et al.*, 1992], which desirably should not be compromised for computational convenience.

Data-update intervals of 1, 2, 5, and 10 days have been used to repeat the identical-twin experiments. Plate 4 shows the resulting RMS prediction errors in time for each update interval. In general, the RMS error at a given time increases monotonically with the length of the update intervals, indicating that systematic biases have been introduced to the estimates owing to delays in data update. For the longer intervals of 5 and 10 days, more pronounced oscillations in RMS error are observed after each data update owing to larger data densities. These oscillations are especially noticeable in the error plots for the velocity components (u, v). The steady state estimates are, nevertheless, stable for each update-interval, as exemplified by Plate 5, displaying the day-180 predictions for the 2-day and 10-day update intervals, where the mesoscale features have been reproduced well in both cases. The differences between the two prediction in Plate 5 are in the details of the features, such as the locations and intensities of the rings. Such differences in qualitative features are significantly smaller between the cases of 2-day and 20-min intervals (Plate 3, top), in accord with the RMS error values (Plate 4). These imply that the data can be updated at an interval larger than the model time step without degrading feature reproduction and filter stability, if the increase in the interval is moderate, i.e., up to 2 days in this particular twin experiment.

4.3.3. Assimilation of generalized data. The generic observation equation (2) in the Kalman filter leads to the weak constraint term $\|\mathbf{y}_k^d - \mathbf{H}_k\mathbf{x}_k\|^2$ in the least squares formulation (3). Algebraically, it makes no difference whether or not this data term contains physical measurements. Any extraneous constraint on the prognostic variables at a given time, e.g., a discretized diagnostic equation, can be expressed as a quadratic norm and treated as a data term. With this perspective of data in Kalman filter, a data assimilation technology can, in principle, be applied for numerical incorporation of such external constraints as open boundary conditions, air-sea interfaces in coupled models, and multiscale (e.g., coastal and open ocean) coupling.

Plate 6 illustrates that an arbitrary diagnostic constraint can be assimilated as a conceptually generalized observation using the MRF-approximated filter. The constraint in this case is intended to form an intrusive boundary, referred to as "coastline" hereafter, which can be seen in Plate 6, bottom, as the slanted right side of the vectorless triangular region, referred to as "land," near the left edge of the domain. The equations of the constraint,

$$4u - v = 0 \tag{22}$$

$$h = \bar{h} \tag{23}$$

where $\bar{h} = 1000\ m$ is the mean layer-thickness, are imposed on the prognostic variables in the "land" area. The first equation (22) is designed to constrain the ocean current to flow parallel to the coastline. Note that the left-hand side of (22) is an inner product between the current vector (u, v) and the normal vector $(4, -1)$ that defines the orientation of the coastline. The second equation, (23), is designed to null the fluid flow over land by smoothing. The algebraic constraints (22) and (23) can be organized as $\mathbf{H}_k\mathbf{x}_k = \mathbf{y}_k$ and then be incorporated into the shallow-water model dynamics using a Kalman filter. The initial and final estimates produced by the MRF filter, applied over 12 days (864 time steps) to a 64 × 64 grid version of the model equation used in the twin experiment described previously, are shown in Plate 6. The final estimate at day 12 displays intended patterns of velocity and layer-thickness fields. Note that the model performs dynamic evolution of the state variables over the land throughout as if it were "ocean"; the appearance of the triangular land region is strictly the effect of data assimilation. The numerical treatment of the coastline boundary condition examined here is applicable (in term of the algorithmic structure) to time-dependent constraints like an open-boundary condition as well. A realistic simulation of basin-scale ocean often requires a continuous incorporation of both measurements and extraneous boundary conditions into the model equations. The illustration here implies the possibility that a data assimilation scheme like the MRF filter can be used as a single consistent framework for numerical treatment of both types of extraneous information.

5. Discussion and Summary

Success in data assimilation depends heavily on characterization and parameterization of the prediction/analysis error covariances. A highly compact parameterization is not only a computational necessity but also inevitable by virtue, as error characterizations available to data assimilation tend to be only qualitative and/or empirical. For example, if a physical or systematic description of prediction errors were available, it is much more likely that this knowledge is used to improve the prognostic equations (4) rather than the data assimilation procedure. The two numerical approaches examined in this paper, subspace projection and MRF-based regression, can effectively represent the $N \times N$ error covariance matrix using only n^2 and nN parameters, respectively, where $n < N$. For subspace projection, n can have a wide range of values depending on the subspace dimension. In particular, an EOF subspace tends to have a very small n, e.g., $n \approx 10$ for cases in section 4.1. The scale-dependent subspace of wavelets usually requires a larger n (n/N is constrained by the scale constant), while the spectral (wavenumber) mode subspace is still larger. For the MRF-based approach, n is determined by the order of approximation, e.g., $n = 13$ for the second-order structure (Figure 2b) used in our numerical experiments, all of which have been performed on a desktop workstation. Our calculations with simple but geophysically relevant models of fluid flow have demonstrated that the second-order MRF parameterization (inhomogeneous and anisotropic) is both efficient and effective in producing near-optimal estimates of the prognostic variables.

The sequential assimilation schemes approximating the Kalman filter based on the two parameterization approaches have consistently produced estimates (analyses) nearly identical to those of nonapproximated Kalman filter, under a wide range of dynamic equations and filter parameters examined in this paper. The numerical experiments, conducted with uncorrelated parameter matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, have presented a stringent test of regularity in the covariance structure and hence stability of the filter. It is then noted that stability of the subspace projection approach can be dependent on choice of the modal functions. Both the empirical (e.g., EOFs) and prescribed (e.g., wavelets) modes have shown such dependency, while the MRF approach has been stable under all numerical situations examined here.

The performance of the approximate filters, if stable, is generally insensitive to the relative weights (i.e., the ratio q/r) given to the model and data, especially when the weight for the data is at least that of the model ($q/r \geq 1$). This parameter insensitivity is encouraging with respect to robustness of the filter outputs, such as analyses of satellite altimetric data. Assimilation of simulated TOPEX/POSEIDON data into a multilayer version of MICOM has resulted in asymptotically diminishing RMS tracking errors (to be presented), similar to those in the single-layer case discussed in section 4.3. For a much more sparsely sampled data set, assimilation outputs can become sensitive to the filter parameters, particularly to the correlation structure of model error in $\mathbf{Q}_k$, as noted by *Bennett and Budgell* [1987] and others. This implies that the correlation scale of the error covariances depends partly on the resolving power of the observation network at hand.

The particular method of MRF parameterization has additional possibilities that remain to be investigated. First, the MRF framework has been used for statistical characterization of geometrical patterns like discontinuity contours and estimation of them using a stochastic relaxation procedure like simulated annealing [*Geman and Geman*, 1984]. This may lead to an assimilation procedure for contour data such as fronts and rings detected on sea-surface temperature maps (satellite based). Second, the MRF parameters should be able to be dynamically updated by simulation of error ensemble [*Evensen*, 1994], as an alternative to the tangent linear approximation scheme used in this paper. This alternative appears to be useful for a prediction of MRF over multiple time steps. Finally, the use of diagnostic equations as "generalized data" (section 4.3.3) needs to be examined for practical cases such as "assimilation" of open boundary conditions. In summary, the Kalman filter generally regards the prognostic equations (4) as a first-order autoregression in time, on which the covariance prediction equation (6) is based. It is then conceptually consistent that the covariance structure be parameterized in space by a regression procedure as well, and this has been achieved successfully here with an MRF. Spatial regression with an MRF is implicative of a differential/difference diagnostic equation, such as the geostrophic balance, which is a classic constraint for the error fields in data analysis/assimilation.

Appendix A: Wavelet Subspaces

Consider a discrete one-dimensional periodic domain consisting of N points. Compactly supported basis functions in a multiresolution framework can be given as the so-called scaling functions of orthonormal and biorthogonal wavelet transforms [*Daubechies*, 1992]. In particular, the symmetric biorthogonal basis functions given by *Daubechies* [1992, section 8.3] have compactly supported inverses as well. For example, the two scaling functions, referred to as $\tilde{\phi}^{3,3}$ and $\tilde{\phi}^{4,4}$, used in this paper are given by

$$\tilde{\phi}^{3,3} = \begin{bmatrix} \vdots \\ 0 \\ 0.1768 \\ 0.5303 \\ 0.5303 \\ 0.1768 \\ 0 \\ \vdots \end{bmatrix}, \qquad \tilde{\phi}^{4,4} = \begin{bmatrix} \vdots \\ 0 \\ -0.0645 \\ -0.0407 \\ 0.4181 \\ 0.7885 \\ 0.4181 \\ -0.0407 \\ -0.0645 \\ 0 \\ \vdots \end{bmatrix} \tag{24}$$

The basis vectors $\mathbf{b}_t$ in (9) are periodically shifted versions of these scaling functions. In the standard multiresolution analysis with diadic scaling, the scaling functions are shifted by two spatial indices. For example, using the scaling function $\tilde{\phi}^{3,3}$, the subspace $\mathbf{B}$ (whose columns are $\mathbf{b}_t$) would look like

$$\mathbf{B} = \begin{bmatrix} & & \vdots & & \\ & 0 & 0 & 0 & \\ & 0.1768 & 0 & 0 & \\ & 0.5303 & 0 & 0 & \\ & 0.5303 & 0.1768 & 0 & \\ \cdots & 0.1768 & 0.5303 & 0 & \cdots \\ & 0 & 0.5303 & 0.1768 & \\ & 0 & 0.1768 & 0.5303 & \\ & 0 & 0 & 0.5303 & \\ & 0 & 0 & 0.1768 & \\ & & \vdots & & \end{bmatrix} \tag{25}$$

and have a dimension of $N \times N/2$.

Further (diadic) reductions in the subspace dimension can be achieved by cascading. By letting $\mathbf{B}^{(m)}$ denote the $2m \times m$ subspace matrix with the exact same local structure as given in (25), an ℓth scale-level subspace is obtained as

$$\mathbf{B} = \mathbf{B}^{(N/2)}\mathbf{B}^{(N/4)} \ldots \mathbf{B}^{(N/2^{\ell})} \tag{26}$$

resulting in a $N \times N/2^{\ell}$ subspace $\mathbf{B}$ for Kalman filter approximation.

An advantage of this particular biorthogonal wavelet transformation is that the basis functions for the pseudoinverse subspace $\mathbf{T}$ (wavelet decomposition) are locally supported, promoting computational efficiency. In particular, the columns of $\mathbf{T}^T$ can be obtained in exactly the same manner as columns of the corresponding $\mathbf{B}$, starting from the scaling functions $\phi^{3,3}$ and $\phi^{4,4}$,

$$\phi^{3,3} = \begin{bmatrix} \vdots \\ 0 \\ 0.0663 \\ -0.1989 \\ -0.1547 \\ 0.9944 \\ 0.9944 \\ -0.1547 \\ -0.1989 \\ 0.0663 \\ 0 \\ \vdots \end{bmatrix}, \qquad \phi^{4,4} = \begin{bmatrix} \vdots \\ 0 \\ 0.0378 \\ -0.0238 \\ -0.1106 \\ 0.3774 \\ 0.8527 \\ 0.3774 \\ -0.1106 \\ -0.0238 \\ 0.0378 \\ 0 \\ \vdots \end{bmatrix} \tag{27}$$

which are the counterparts of $\tilde{\phi}^{3,3}$ and $\tilde{\phi}^{4,4}$, respectively. Figure 1 shows the basis functions associated with the four scaling functions $\tilde{\phi}^{3,3}$, $\tilde{\phi}^{4,4}$, $\phi^{3,3}$, and $\phi^{4,4}$, for the first three scale levels.

Appendix B: Time Recursion of Information Matrix

The Kalman filter that results from time recursion of the information matrix $\mathbf{L}_k$ instead of the covariance $\mathbf{P}_k$ is often referred to as the information filter [*Lewis*, 1986]. The MRF-conforming recursion (see (16)–(18)) is an approximation of such filtering equations. The nonapproximated version of the equations can be retrieved by replacing (16) with

$$\mathbf{\Theta}_k = (\mathbf{F}_k^T\mathbf{M}_k\mathbf{F}_k + \mathbf{L}_{k-1}^a)^{-1}\mathbf{F}_k^T\mathbf{M}_k \tag{28}$$

A systematic way to numerically approximate this is to perform the matrix inverse on the right-hand side with the Jacobi iterations [*Chin et al.*, 1992, 1994], where (16) represents the simplest of such approximations. Since (17) and (18) are sums of quadratic terms, $\mathbf{L}_k$ can remain as positive semidefinite throughout the recursion for any matrix $\mathbf{\Theta}_k$, i.e., for an arbitrary approximation of (28). If $\mathbf{\Theta}_k$ has a full rank for each k, then $\mathbf{L}_k$ would retain positive definiteness.

This appendix presents a derivation of the exact recursion for the information matrix; $\mathbf{L}_k$ is not approximated by truncation here. While the usual form of the information filter involves explicit computation of inverse (time reverse) dynamics $\mathbf{F}_k^{-1}$, the recursion steps (28), (17), and (18) avoid such an inversion. Duality with the standard recursion of the covariance matrix $\mathbf{P}_k$ can also be recognized, when comparing (28) with (7), (17) with (8), and (18) with (6). The original recursion (6)–(8) of the covariance matrix in Kalman filter corresponds to the following dynamic system for the error process $\mathbf{e}_k$:

$$\mathbf{e}_k = \mathbf{F}_k\mathbf{e}_{k-1} + \mathbf{w}_k \qquad (29)$$

$$0 = \mathbf{H}_k\mathbf{e}_k + \mathbf{v}_k \qquad (30)$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are zero-mean processes representing the model and data error processes, respectively, with respective covariance matrices of $\mathbf{Q}_k$ and $\mathbf{R}_k$. Our interest here is to derive an equivalent time recursion for the information matrix $\mathbf{L}_k$, defined as the square $\mathbf{L}_k \equiv \mathbf{\Gamma}_k^T\mathbf{\Gamma}_k$ of the regression operator $\mathbf{\Gamma}_k$ satisfying (15).

B1. Information Matrices for Model and Data Errors

The information matrices associated with the model and data error processes $\mathbf{w}_k$ and $\mathbf{v}_k$ can be obtained simply as the inverses of the corresponding covariance matrices, $\mathbf{M}_k \equiv \mathbf{Q}_k^{-1}$ and $\mathbf{N}_k \equiv \mathbf{R}_k^{-1}$, yielding the weighting matrices in the least squares formulation (3). For these error processes the regression operator can usually be translated from linear diagnostic balance equations that are given or assumed. The corresponding information matrix is then obtained by squaring. For example, *Jiang and Ghil* [1993] (among others) have assumed that the model errors/residuals (u', v', h') in a two-dimensional, shallow-water dynamic system satisfy a geostrophic balance, which can be written (assuming a proportionality between the dynamic pressure and layer depth) as

$$\begin{bmatrix} 1 & 0 & \frac{c}{f}\frac{\partial}{\partial y} \\ 0 & 1 & -\frac{c}{f}\frac{\partial}{\partial x} \end{bmatrix}\begin{bmatrix} u' \\ v' \\ h' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (31)$$

for a proportionality constant c and Coriolis parameter f. When this equality is accepted with some form of uncertainty (e.g., replacing the right-hand side with a zero-mean uncorrelated process with a variance of w^{-1}), the left matrix operator represents a regression operator for the error vector $[u', v', h']^T$. The corresponding information matrix $\mathbf{M}_{\text{geostrophy}}$ would be a finite difference version of

$$\mathbf{M}_{\text{geostrophy}} = w\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -\frac{c}{y}\frac{\partial}{\partial y} & \frac{c}{f}\frac{\partial}{\partial x} \end{bmatrix}\begin{bmatrix} 1 & 0 & \frac{c}{f}\frac{\partial}{\partial y} \\ 0 & 1 & -\frac{c}{f}\frac{\partial}{\partial x} \end{bmatrix}$$

$$= w\begin{bmatrix} 1 & 0 & \frac{c}{f}\frac{\partial}{\partial y} \\ 0 & 1 & -\frac{c}{f}\frac{\partial}{\partial x} \\ -\frac{c}{f}\frac{\partial}{\partial y} & \frac{c}{f}\frac{\partial}{\partial x} & -\frac{c^2}{f^2}\nabla \end{bmatrix} \qquad (32)$$

where ∇ is the Laplacian operator. The difference operators tend to make the information matrix like $\mathbf{M}_{\text{geostrophy}}$ singular, preventing an explicit expression of the corresponding covariance matrix $\mathbf{Q}_{\text{geostrophy}}$.

B2. Forecasting of Information Matrix

We consider computing the information matrix $\mathbf{L}_k^f$ associated with the forecast error, given the nearest-past analysis error information $\mathbf{L}_{k-1}^a$ and the forecast equation (29). The key algebraic steps involve application of the standard maximum likelihood (ML) estimation formulas and block matrix inversion techniques.

Assume for the moment that the regression operator $\mathbf{\Gamma}_{k-1}^a$ is available, so that $\mathbf{\Gamma}_{k-1}^a\mathbf{e}_{k-1} = \boldsymbol{\delta}$. Augmenting with (29) will yield an ML formulation for a joint estimation of $\mathbf{e}_{k-1}$ and $\mathbf{e}_k$ as

$$\begin{bmatrix} \mathbf{\Gamma}_{k-1}^a & 0 \\ -\mathbf{F}_k & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{e}_{k-1} \\ \mathbf{e}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{\delta} \\ \mathbf{w}_k \end{bmatrix} \qquad (33)$$

Because $\mathbf{e}_{k-1}$ is uncorrelated with $\mathbf{w}_k$ (a fundamental assumption in the standard Kalman filter), this ML problem is equivalent to the optimal single-step forecasting for the estimation error. The standard formula for the $2N \times 2N$ posterior estimation error covariance $\mathcal{P}$ for the ML problem is [e.g., *Lewis*, 1986]

$$\mathcal{P} = \left(\begin{bmatrix} \mathbf{\Gamma}_{k-1}^a & 0 \\ -\mathbf{F}_k & \mathbf{I} \end{bmatrix}^T\begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{Q}_k^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{\Gamma}_{k-1}^a & 0 \\ -\mathbf{F}_k & \mathbf{I} \end{bmatrix}\right)^{-1}$$

$$= \begin{bmatrix} \mathbf{F}_k^T\mathbf{M}_k\mathbf{F}_k + \mathbf{L}_{k-1}^a & -\mathbf{F}_k^T\mathbf{M}_k \\ -\mathbf{M}_k\mathbf{F}_k & \mathbf{M}_k \end{bmatrix}^{-1} \qquad (34)$$

in which the square of the regression operator $\mathbf{\Gamma}_{k-1}^a$ has been replaced by the information matrix $\mathbf{L}_{k-1}^a$ in the last step. By partitioning this covariance matrix by $N \times N$ blocks $\mathbf{P}_{ij}$ as

$$\mathcal{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}, \qquad (35)$$

we can identify from (33) that the posterior covariance for $\mathbf{e}_k$ is $\mathbf{P}_{22}$. The desired information matrix can thus be obtained as

$$\mathbf{L}_k^f = \mathbf{P}_{22}^{-1}$$

$$= \mathbf{M}_k - \mathbf{M}_k\mathbf{F}_k(\mathbf{F}_k^T\mathbf{M}_k\mathbf{F}_k + \mathbf{L}_{k-1}^a)^{-1}\mathbf{F}_k^T\mathbf{M}_k \qquad (36)$$

The last step is due to a block matrix inversion formula, i.e., denoting (34) concisely as

$$\mathcal{P} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix}^{-1},$$

it can be verified by substitution that $\mathbf{P}_{11} = \mathbf{L}_{11}^{-1} + \mathbf{L}_{11}^{-1}\mathbf{L}_{12}\mathbf{P}_{22}\mathbf{L}_{21}\mathbf{L}_{11}^{-1}$, $\mathbf{P}_{22} = (\mathbf{L}_{22} - \mathbf{L}_{21}\mathbf{L}_{11}^{-1}\mathbf{L}_{12})^{-1}$, $\mathbf{P}_{12} = -\mathbf{L}_{11}^{-1}\mathbf{L}_{12}\mathbf{P}_{22}$, and $\mathbf{P}_{21} = -\mathbf{P}_{22}\mathbf{L}_{21}\mathbf{L}_{11}^{-1}$. Equation (36) can be shown to be equivalent to the recursion steps (28) and (17), which can be verified by elimination of terms after substituting (28) into (17).

B3. Data Update of Information Matrix

To verify the data-update step (18), we assume initially that the regression operator $\mathbf{\Gamma}_k^f$ is available, so that $\mathbf{\Gamma}_k^f\mathbf{e}_k = \boldsymbol{\delta}$. The following ML estimation problem can then be formulated by augmentation with the observation equation (30):

$$\begin{bmatrix} \mathbf{\Gamma}_k^f \\ \mathbf{H}_k \end{bmatrix}\mathbf{e}_k = \begin{bmatrix} \boldsymbol{\delta} \\ -\mathbf{v}_k \end{bmatrix} \qquad (37)$$

whose posterior covariance matrix is the desired analysis error covariance. Using the standard ML formula as before, we then have

$$\mathbf{L}_k^a \equiv (\mathbf{P}_k^a)^{-1}$$

$$= \begin{bmatrix} \mathbf{\Gamma}_k^f \\ \mathbf{H}_k \end{bmatrix}^T\begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{R}_k^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{\Gamma}_k^f \\ \mathbf{H}_k \end{bmatrix}$$

$$= \mathbf{L}_k^f + \mathbf{H}_k^T\mathbf{N}_k\mathbf{H}_k \qquad (38)$$

which is (18).

Acknowledgments. We thank the Office of Naval Research for funding this study through grants N00014-95-1-0257 (T.M.C. and A.J.M.) and N00014-93-1-0404 (E.P.C.), as well as NASA through grant NAGW-4329. We are grateful to Ichiro Fukumori (Jet Propulsion Laboratory) for the TOPEX/POSEIDON satellite track data. We also appreciate helpful comments on the manuscript by Bob Miller, Dimitris Menemenlis, and an anonymous reviewer.

References

Anderson, B. D. O., and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, N. J., 1979.

Bennett, A. F., *Inverse Methods in Physical Oceanography*, Cambridge Univ. Press, New York, 1992.

Bennett, A. F., and W. P. Budgell, Ocean data assimilation and the Kalman filter: Spatial regularity, *J. Phys. Oceanogr.*, *17*, 1583–1601, 1987.

Besag, J., Spatial interaction and the statistical analysis of lattice systems, *J. R. Stat. Soc. Ser. B*, *36*, 192–236, 1974.

Bleck, R., and E. Chassignet, Simulating the oceanic circulation with isopycnic-coordinate models, in *The Oceans: Physical-Chemical Dynamics and Human Impact*, edited by S. K. Majumbar and W. E. Miller, pp. 17–39, Pa. Acad. of Sci., Easton, 1994.

Bleck, R., C. Rooth, D. Hu, and L. T. Smith, Ventilation and mode water formation in a wind- and thermohaline-driven isopycnic coordinate model of the North Atlantic, *J. Phys. Oceanogr.*, *22*, 1486–1505, 1992.

Bleck, R., S. Dean, M. O'Keefe, and A. Sawdey, A comparison of data-parallel and message-passing versions of the Miami Isopycnic Coordinate Ocean Model (MICOM), *Parallel Comput.*, *21*, 1695–1720, 1995.

Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi, Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter, *J. Geophys. Res.*, *101*, 22,599–22,617, 1996.

Chassignet, E. P., D. B. Olson, and D. B. Boudra, Motion and evolution of oceanic rings in a numerical model and in observations, *J. Geophys. Res.*, *95*, 22,121–22,140, 1990.

Chassignet, E. P., W R. Holland, and A. Capotondi, Impact of the altimeter orbit on the reproduction of oceanic rings: Application to a regional model of the Gulf Stream, *Oceanol. Acta*, *15*, 479–490, 1992.

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

Chin, T. M., W. C. Karl, and A. S. Willsky, Probabilistic and sequential computation of optical flow using temporal coherence, *IEEE Trans. Image Process.*, *3*, 773–788, 1994.

Chin, T. M., W. C. Karl, and A. S. Willsky, A distributed and iterative method for square root filtering in space-time estimation, *Automatica*, *31*, 67–82, 1995.

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

Cohn, S. E., and D. F. Parrish, The behavior of forecast error covariances for a Kalman filter in two dimensions, *Mon. Weather Rev.*, *119*, 1757–1785, 1991.

Daley, R., *Atmospheric Data Analysis*, Cambridge Univ. Press, New York, 1991.

Daubechies, I., *Ten Lectures on Wavelets*, Soc. for Ind. and Appl. Math., Philadelphia, Pa., 1992.

Dee, D. P., Simplification of the Kalman filter for meteorological data assimilation, *Q. J. R. Meteorol. Soc.*, *117*, 365–384, 1991.

Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162, 1994.

Fukumori, I., and P. Malanotte-Rizzoli, An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model, *J. Geophys. Res.*, *100*, 6777–6793, 1995.

Geman, S., and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intel.*, *PAMI-6*, 721–741, 1984.

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

Ghil, M., S. Cohn, J. Travantzis, K. Bube, and E. Issacson, Application of estimation theory to numerical weather prediction, in *Dynamic Meteorology: Data Assimilation Methods*, edited by L. Bengtsson, M. Ghil, and E. Kallen, pp. 139–284. Springer-Verlag, New York, 1981.

Habibi, A., Two-dimensional Bayesian estimate of images, *Proc. IEEE*, *60*, 878–883, 1972.

Jiang, S., and M. Ghil, Dynamical properties of error statistics in a shallow-water model, *J. Phys. Oceanogr.*, *23*, 2541–2566, 1993.

Kindermann, R., and J. L. Snell, *Markov Random Fields and Their Applications*, Am. Math. Soc., Providence, R. I., 1980.

Lewis, F. L., *Optimal Estimation*, John Wiley, New York, 1986.

Mariano, A. J., and O. Brown, Efficient objective analysis of dynamically heterogeneous and nonstationary fields via the parameter matrix, *Deep Sea Res., Part A*, *39*, 1255–1271, 1992.

Menemenlis, D., and C. Wunsch, Linearization of an oceanic general circulation model for data assimilation and climate studies, *J. Atmos. Oceanic Technol.*, *14*, 1420–1443, 1997.

Miller, R. N., Toward the application of the Kalman filter to regional open ocean modeling, *J. Phys. Oceanogr.*, *16*, 72–86, 1986.

Miller, R. N., and M. A. Cane, A Kalman filter analysis of sea level height in the tropical Pacific, *J. Phys. Oceanogr.*, *19*, 773–790, 1989.

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

Parrish, D. F., and S. E. Cohn, A Kalman filter for a two-dimensional shallow-water model: Formulation and preliminary experiments, *Tech. Rep. Off. Note 304*, Natl. Cent. for Environ. Predict., Washington, D. C., 1985.

Pham, D. T., J. Verron, and M. C. Roubaud, A singular evolutive extended Kalman filter for data assimilation in oceanography, *J. Mar. Syst.*, *16*, 323–340, 1998.

Sasaki, Y., Some basic formalism in numerical variational analysis, *Mon. Weather Rev.*, *98*, 875–883, 1970.

Talagrand, O., Assimilation of observations, an introduction, *J. Meteorol. Soc. Jpn.*, *75*, 191–209, 1997.

Thacker, W. C., and R. B. Long, Fitting dynamics to data, *J. Geophys. Res.*, *93*, 1227–1240, 1988.

Wilks, D. S., *Statistical Methods in the Atmospheric Sciences*, Academic, San Diego, Calif., 1995.

Wunsch, C., *The Ocean Circulation Inverse Problem*, Cambridge Univ. Press, New York, 1996.

E. P. Chassignet, T. M. Chin, and A. J. Mariano, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149. (amariano@rsmas.miami.edu)

(Received November 10, 1997; revised November 10, 1998; accepted November 19, 1998.)