A Three-Dimensional Variational Data Assimilation Scheme for the Regional Ocean Modeling System

ZHIJIN LI AND YI CHAO

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

JAMES C. MCWILLIAMS AND KAYO IDE

Department of Atmospheric and Oceanic Sciences, Institute of Geophysics and Planetary Physics, University of California, Los Angeles, Los Angeles, California

ABSTRACT

A three-dimensional variational data assimilation (3DVAR) scheme has been developed within the framework of the Regional Ocean Modeling System (ROMS). This ROMS3DVAR enables the capability of predicting meso- to small-scale variations with temporal scales from hours to days in coastal oceans. To cope with particular difficulties that result from complex coastlines and bottom topography, unbalanced flows, and sparse observations, ROMS3DVAR includes novel strategies. These strategies include the implementation of three-dimensional anisotropic and inhomogeneous error correlations based on a Kronecker product, application of particular weak dynamic constraints, and implementation of efficient and reliable algorithms for minimizing the cost function. The formulation of ROMS3DVAR is presented here, and its implementation off the West Coast is currently under way.

1. Introduction

The coastal ocean has a number of unique physical and meteorological processes that promote high biological productivity, active sedimentary processes, dynamic chemical transformations, and intense air–sea interactions. In addition to being an area of great scientific interest, the coastal ocean is important for a number of economic reasons, including recreation, coastal development, mineral extraction, and fisheries. As the population and various activities in the coastal region increase, real-time monitoring and forecasting become increasingly important.

In response to its scientific and economic importance, there are ongoing efforts in the federal and state government to establish a network of coastal ocean observing systems that will suit both research and operational needs (National Research Council 2003). As a consequence, we have been seeing an explosive development of new observational tools and platforms. Development in satellite oceanography provides unprecedented views of the ocean with increasing resolutions and accuracy, coupled with capabilities of measuring new variables (e.g., sea surface salinity). We also envision a new generation of in situ observations, ranging from autonomous instruments and probes to cabled observatories. Shore-based high-frequency (HF) radars provide spatially and temporally continuous sea surface velocity maps. Coastal oceanography is in a rapid transition from an era of exploratory mapping and sampling to one of real-time observing systems.

Real-time observing systems make it possible to develop predictive numerical modeling systems that will routinely provide coastal oceanic "weather" predictions, as today's meteorological models routinely provide numerical weather predictions for the atmosphere. A predictive modeling system requires two basic components: an advanced numerical model and real-time data assimilation system (Kalnay 2003). During the last two decades, coastal ocean modeling has greatly advanced. Data assimilation, which integrates data from observing systems into models, has also been extensively investigated (e.g., Wunsch 1996; Bennett 1997).

Corresponding author address: Zhijin Li, Jet Propulsion Laboratory, M/S 300–323, 4800 Oak Grove Drive, Pasadena, CA 91109.
E-mail: zhijin.li@jpl.nasa.gov

From the beginning of the 1990s, major meteorological centers around the world have operationally implemented advanced data assimilation systems. These include three- and four-dimensional variational data assimilation (3DVAR and 4DVAR) methods (e.g., Parrish and Derber 1992; Cohn et al. 1998; Courtier et al. 1998; Rabier et al. 1998a; Moore et al. 2004; DiLorenzo et al. 2007) and the Kalman filter method (e.g., Fukumori et al. 1999; Lermusiaux and Robinson 1999). The progress made in coastal ocean science during the last two decades has laid a sound scientific foundation for the development of advanced real-time coastal ocean data assimilation systems similar to the ones meteorological centers have been using on a routine basis.

In this paper we present a theoretical framework of a 3DVAR algorithm within the framework of the Regional Ocean Modeling System (ROMS) (Shchepetkin and McWilliams 2003, 2005; Marchesiello et al. 2001, 2003).

We chose to implement 3DVAR for the following reasons: first, it has been widely adopted for operational use in meteorological centers and proved to be a reliable algorithm; second, it is computationally efficient enough to serve as part of an operational system that focuses on coastal oceans and has time scales of hours to days; third, it is flexible enough to incorporate nonstandard measurements; and, fourth, it is the soundest path to the ultimate development of more advanced data assimilation systems, such as the Kalman filter (e.g., Fisher et al. 2006; Li and Navon 2001) and 4DVAR (e.g., Daley and Barker 2001).

ROMS3DVAR was developed as an operational data assimilation system similar to those implemented in major meteorological centers and thus possesses the following capabilities: first, it generates initial conditions that ROMS uses to predict future coastal oceanic states; second, it provides basic analyzed fields for various users, including oceanic forecasters and the public; and, third, it produces datasets for various diagnostic analyses, such as those used for health monitoring. ROMS3DVAR also has the capability of being used as a numerical test bed for the evaluation and assessment of new types of measurements. ROMS3DVAR was developed to maximize flexibility and stability in assimilating different types of new measurements. For example, it allows simultaneous assimilation of measurements ranging from radial velocities from HF radars (Paduan and Graber 1997) to radiance from satellite-flown remote sensing instruments. Moreover, assimilation of chemical and biological quantities can be formulated as an extension of the physical quantities. For example, biological concentrations can be treated in the same way as salinity.

We here present the formulation of ROMS3DVAR. ROMS3DVAR has been implemented in the coastal regions off central California (http://ourocean.jpl.nasa.gov/MB06) and in the Southern California Bight (http://ourocean.jpl.nasa.gov/SCB) as an operational system. The implementation and basic assessment of the system are described in Li et al. (2008), and the evaluation of this system off central California is reported in Chao et al. (2008, manuscript submitted to *Deep-Sea Res.*).

This paper is organized as follows: sections 2 and 3 present the basic formulation of ROMS3DVAR. Section 3 details the particular formulation of the dynamic constraints. Section 4 is dedicated to the estimation and modeling of error covariance, especially the Kronecker product–based algorithms proposed for model error correlations. In section 5, the minimization method is addressed, and a sophisticated procedure for reconditioning cost functions is described. Section 6 gives a summary.

2. Three-dimensional variational data assimilation

In this section, we will describe the basic formulation of ROMS3DVAR. We will follow the notation conventions suggested by Ide et al. (1997). Vectors will be represented by boldface lowercase letters, and matrices by sans serif uppercase Roman and Greek letters. Also, the superscript "f" stands for forecast, the superscript "a" for analysis, the superscript "t" for truth, the superscript "T" for matrix transpose, and the symbol $\langle \cdot \rangle$ for mathematical expectation.

The formulation presented in this paper will involve only the physical components of ROMS for simplicity, though ROMS has chemical, biological, and other components. The physical variables in ROMS are sea surface height (SSH) ζ, horizontal velocity components u and v, temperature T, and salinity S.

The model state at a single time can be written as a vector, consisting of the values of ζ, u, v, T, and S at the model grid points. We use $\mathbf{x}_u$, $\mathbf{x}_v$, $\mathbf{x}_T$, $\mathbf{x}_s$, and $\mathbf{x}_\zeta$ for the model state along with N_u, N_v, N_T, N_S, and N_ζ for their dimensions as denoted by the corresponding subscript. For convenience, we organize them into three groups: $\mathbf{x}_\zeta$, $\mathbf{x}_{uv} = (\mathbf{x}_u^T, \mathbf{x}_v^T)^T$ for velocity with dimension $N_{uv} = N_u + N_v$, $\mathbf{x}_{TS} = (\mathbf{x}_T^T, \mathbf{x}_S^T)^T$ for T, and S with dimension $N_{TS} = N_T + N_S$. Altogether, the net state vector $\mathbf{x}$ of ROMS at a given time can be written as

$$\mathbf{x}^T = (\mathbf{x}_\zeta^T, \mathbf{x}_{uv}^T, \mathbf{x}_{TS}^T), \tag{1}$$

where $N = N_\zeta + N_u + N_v + N_T + N_S$ is the dimension of $\mathbf{x}$.

In data assimilation, the model state is combined

with observations. Generally the M-dimensional observation vector $\mathbf{y}$ is related to the model state $\mathbf{x}$ by

$$\mathbf{y} = \mathcal{H}(\mathbf{x}^t) + \mathbf{e}, \tag{2}$$

where $\mathbf{x}^t$ is the unknown true state to be estimated; $\mathcal{H}$ is the observational operator that maps the model state to the observation location and $\mathbf{e}$ is the observational error.

We consider three types of observations: M_ζ-dimensional SSH observation $\mathbf{y}_\zeta$, M_{uv}-dimensional velocity observation $\mathbf{y}_{uv}$, and M_{TS}-dimensional T and S observation $\mathbf{y}_{TS}$. Similar to a model state, we write the net observation as

$$\mathbf{y}^T = (\mathbf{y}_\zeta^T, \mathbf{y}_{uv}^T, \mathbf{y}_{TS}^T), \tag{3}$$

where $M_{\mathrm{obs}} = M_\zeta + M_{uv} + M_{TS}$ is the total number of observations. When the observational operator is linear, we have $\mathcal{H}(\mathbf{x}) = \mathbf{H}\mathbf{x}$, where

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_\zeta & \mathbf{0}_{M_\zeta \times N_{uv}} & \mathbf{0}_{M_\zeta \times N_{TS}} \\ \mathbf{0}_{M_{uv} \times N_\zeta} & \mathbf{H}_{uv} & \mathbf{0}_{M_{uv} \times N_{TS}} \\ \mathbf{0}_{M_{TS} \times N_\zeta} & \mathbf{0}_{M_{TS} \times N_{uv}} & \mathbf{H}_{TS} \end{pmatrix} \tag{4}$$

is the $M \times N$-dimensional observation matrix; $\mathbf{H}_\zeta$ is the $M_\zeta \times N_\zeta$-dimensional SSH observation matrix; $\mathbf{H}_{uv}$ is $M_{uv} \times N_{uv}$-dimensional velocity observation matrix; $\mathbf{H}_{TS}$ is $M_{TS} \times N_{TS}$-dimensional velocity observation matrix; and the $\mathbf{0}_{M_* \times N_*}$ stands for $M_* \times N_*$-dimensional zero matrix. When a type of observation is unavailable, corresponding rows are removed from $\mathbf{y}$ and $\mathbf{H}$ in (2) and (4). For simplicity in presentation, the dimension of vectors and matrices is noted only when necessary from here on.

The observational error $\mathbf{e}$ consists of two parts:

$$\mathbf{e} = \mathbf{e}_O + \mathbf{e}_F, \tag{5}$$

where $\mathbf{e}_O$ is the instrumental error and $\mathbf{e}_F$ is the representativeness error, respectively. While the source of $\mathbf{e}_O$ is obvious, the causes of $\mathbf{e}_F$ can be complex. The representativeness error can be described as that component present in the observation that cannot be represented by $\mathcal{H}(\mathbf{x})$ (Lorenc et al. 2000). For coastal ocean models there are mostly two processes that contribute to $\mathbf{e}_F$. First, a tidal model is implemented as an independent component and is not coupled with the ocean model, such as the one used in this study; and second, the model does not resolve subgrid-scale processes. The latter can be crucial in some coastal regions. For example, in the Southern California Bight there are active and energetic small-scale eddies with spatial scales ranging from a few hundred meters to a few kilometers (DiGiacomo and Holt 2001).

In the standard 3DVAR formulation, the cost function is defined by

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^f)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^f)$$
$$+ \frac{1}{2}[\mathcal{H}(\mathbf{x}) - \mathbf{y}]^T \mathbf{R}^{-1}[\mathcal{H}(\mathbf{x}) - \mathbf{y}], \tag{6}$$

where $\mathbf{x}^f$ is the forecast generated by the forward time integration of ROMS, and

$$\mathbf{B} = \langle(\mathbf{x}^f - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^t)^T\rangle \quad \text{and} \quad \mathbf{R} = \langle \mathbf{e}\mathbf{e}^T\rangle \tag{7}$$

are the $N \times N$ background (forecast) error covariance matrix for $\mathbf{x}^f$ and the $M \times M$ observation error covariance matrix for $\mathbf{y}$, respectively. The minimizer of $J(\mathbf{x})$ is the so-called analysis $\mathbf{x}^a$, which is the desired estimate for $\mathbf{x}^t$. This analysis is statistically optimal if $\mathbf{B}$ and $\mathbf{R}$ are accurately given. When the observation operator $\mathcal{H}$ is linear, $\mathbf{x}^a$ is also a minimum error variance estimate (e.g., Jazwinski 1970; Cohn 1997). Furthermore, if both forecast and observation errors are Gaussian, then it yields a maximum likelihood (Bayesian) estimate.

For operational atmospheric and oceanic data assimilation systems using large-dimensional models and a substantial number of observations, minimizing $J(\mathbf{x})$ is always a challenge. To improve computational efficiency and reliability, $J(\mathbf{x})$ defined by (6) is often transformed and/or approximated before a minimization algorithm is applied. The incremental formulation has been used as a basic approximation in operational 3DVAR and 4DVAR systems (Courtier et al. 1994). When a forecast is accurate, the analysis increment is a small derivation from the forecast. This grants a foundation for linearization around $\mathbf{x}^f$ in $J(\mathbf{x})$. It is a reasonable assumption for current coastal ocean models like ROMS.

In the incremental 3DVAR formulation, the cost function is defined by

$$J_{\mathrm{inc}}(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T \mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^T \mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}).$$
$$\tag{8}$$

The first and second terms are often referred to as the background and observational cost functions. Here $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^f$ is the increment, measuring the deviation of $\mathbf{x}$ from $\mathbf{x}^f$; $\mathbf{d} = \mathbf{y} - \mathcal{H}(\mathbf{x}^f)$ is the innovation, measuring the deviation of $\mathbf{y}$ from $\mathbf{x}^f$; and $\mathbf{H} = \partial\mathcal{H}/\partial\mathbf{x}$ is Jacobian matrix for nonlinear $\mathcal{H}$. The incremental approximation makes $J_{\mathrm{inc}}(\delta\mathbf{x})$ quadratic and positive definite in the control variables, $\delta\mathbf{x}$. The quadratic and positive definite properties guarantee that $J_{\mathrm{inc}}(\delta\mathbf{x})$ has a unique minimum; this uniqueness is also essential for the effi-

ciency and reliability of the minimization process. The final analysis of the incremental 3DVAR is

$$\mathbf{x}^a = \mathbf{x}^f + \delta\mathbf{x}^a, \tag{9}$$

where the analysis increment $\delta\mathbf{x}^a$ minimizes $J_{\text{inc}}(\delta\mathbf{x})$.

The adjustment of $J_{\text{inc}}(\delta\mathbf{x})$ for dynamical constraints and improved computational performance is addressed in sections 3 through 5.

3. Dynamical constraints

ROMS3DVAR is designed to assimilate different types of observations simultaneously and reliably. Different types of observations come with different errors and generally are dynamically inconsistent. One challenge is to keep the obtained analysis in reasonable dynamic balance after the different types of observations are assimilated. For example, in many cases the velocity is nearly in geostrophic balance, but the data assimilation algorithm described above has no capacity for keeping the velocity analysis close in geostrophic balance. The concept of dynamical constraints has been used extensively by meteorological data analysis since their introduction by Sasaki (1970).

In ROMS3DVAR we are concerned with two dynamic constraints. One is hydrostatic balance

$$\frac{\partial \delta p}{\partial z} = -g\delta\rho, \tag{10}$$

and the other is geostrophic balance

$$\mathbf{\Omega} \times \delta\boldsymbol{v} = -\rho^{-1}\boldsymbol{\nabla}\delta p, \tag{11}$$

where both are written in the continuous form using pressure p, density ρ, gravitational acceleration g, angular velocity of earth's rotation $\mathbf{\Omega}$, and horizontal velocity vector $\boldsymbol{v}$. In (10) and (11), the balance constraints are applied to the increment, rather than the total field. Applying the balance constraints to the increments is important for coastal regions. In regions with complex coastline and bottom topography, and especially for upwelling-related strong fronts and energetic eddies, the internal dynamic consistency of $\mathbf{x}^f$ is subtle and delicate. Such consistency should not be disrupted by applying relatively simple dynamic balance constraints like (10) and (11). When the dynamic balance constraint is applied to $\delta\mathbf{x}$, the analysis can still inherit the dynamic consistency of $\mathbf{x}^f$ when $\delta\mathbf{x}^a$ is not large.

Moreover, $\delta\mathbf{x}_\zeta$ and $\delta\mathbf{x}_{uv}$ in the coastal ocean may not necessarily be in exact dynamic balance. Rather, they have both balanced and unbalanced components (Li et al. 2008). To formulate an incremental 3DVAR algorithm that has the capacity to address balanced and

unbalanced components simultaneously, we follow the method proposed by Parrish and Derber (1992). Their method partitions the model state vector into slow (balanced) and fast (unbalanced) components, and replaces a portion of the control variables with the unbalanced components. Doing so imposes dynamic balance only partially and hence implements what are called weak constraints. This method has recently been applied in several different data assimilation systems (Lorenc 2003a; Barker et al. 2004).

Based on the hydrostatic balance (10), $\delta\mathbf{x}_\zeta$ can be approximately represented in discrete form as

$$\delta\mathbf{x}_\zeta = \mathbf{\Pi}\delta\mathbf{x}_{TS} + \delta\mathbf{x}_{\zeta'}, \tag{12}$$

where $\mathbf{\Pi}\delta\mathbf{x}_{TS}$ is the steric SSH that is in balance with the T and S increment (11), and $\delta\mathbf{x}_{\zeta'}$ is the nonsteric SSH increment due to water mass convergence in the column; see appendix A for details. In contrast to ocean data assimilation with large-scale models (e.g., Fukumori et al. 1999), $\delta\mathbf{x}_{\zeta'}$ is nonnegligible for coastal data assimilation as illustrated in Li et al. (2008).

Based on the geostrophic balance (11), $\delta\mathbf{x}_{uv}$ can be approximately represented in discrete form as

$$\delta\mathbf{x}_{uv} = \mathbf{\Gamma}_{TS}\delta\mathbf{x}_{TS} + \mathbf{\Gamma}_{\zeta'}\delta\mathbf{x}_{\zeta'} + \delta\mathbf{x}_{u''v''}, \tag{13}$$

where $\mathbf{\Gamma}_{TS}\delta\mathbf{x}_{TS} + \mathbf{\Gamma}_{\zeta'}\delta\mathbf{x}_{\zeta'}$ is the geostrophic velocity that is in balance with the pressure gradient (11) in a response to changes in steric and nonsteric SSH (12), and $\delta\mathbf{x}_{u''v''}$ is the ageostrophic velocity; see appendix A for details. While $\mathbf{\Gamma}_{TS}\delta\mathbf{x}_{TS}$ due to T/S profile changes may be dominant offshore, $\mathbf{\Gamma}_{\zeta'}\delta\mathbf{x}_{\zeta'}$ due to nonsteric SSH and ageostrophic $\delta\mathbf{x}_{u''v''}$ can be sizable nearshore.

As explained later, the streamfunction and velocity potential are more suitable than the velocity components for use as control variables in data assimilation (Daley 1991). As such, the incremental ageostrophic velocity ($\delta u''$, $\delta v''$) may be written using the incremental ageostrophic streamfunction $\delta\psi''$ and the velocity potential $\delta\chi''$ as

$$\delta u'' = -\frac{\partial\delta\psi''}{\partial y} + \frac{\partial\delta\chi''}{\partial x},$$

$$\delta v'' = \frac{\partial\delta\psi''}{\partial x} + \frac{\partial\delta\chi''}{\partial y}. \tag{14}$$

Using the discrete form of (14), (13) becomes

$$\delta\mathbf{x}_{uv} = \mathbf{\Gamma}_{\zeta'}\delta\mathbf{x}_{\zeta'} + \mathbf{\Gamma}_{TS}\delta\mathbf{x}_{TS} + \mathbf{\Phi}\delta\mathbf{x}_{\psi''\chi''}, \tag{15}$$

where $\delta\mathbf{x}_{\psi''\chi''}$ is the $N_{\psi\chi}$-dimensional vector, encompassing the discrete $\delta\psi''$ and $\delta\chi''$, and $\mathbf{\Phi}$ is a matrix for computing the velocity component from the streamfunction and velocity potential. A numerical method for computing streamfunction and velocity potential

from the velocity components for irregular domains like coastal oceans is available (Li et al. 2006).

To adjust $J_{\text{inc}}(\delta\mathbf{x})$ for dynamic constraints, we assume that errors in $\delta\mathbf{x}_{\zeta'}$, $\delta\mathbf{x}_{\psi''\chi''}$, and $\delta\mathbf{x}_{TS}$ are uncorrelated to each other in the present ROMS3DVAR. This assumption is examined in Li et al. (2008), although some adjustments may be needed. For example, it can be helpful to use spatially smoothed steric SSHs, rather than steric SSHs directly. With this assumption, a natural form of the control variable under the constraints chosen is then

$$\delta\mathbf{x}_c^{\text{T}} = (\delta\mathbf{x}_{\zeta'}^{\text{T}}, \delta\mathbf{x}_{\psi''\chi''}^{\text{T}}, \delta\mathbf{x}_{TS}^{\text{T}}), \quad (16)$$

whose linear relation to $\delta\mathbf{x}$ is given by (12) and (15). We further assume that observational errors $\mathbf{e}_\zeta$, $\mathbf{e}_{uv}$, and $\mathbf{e}_{TS}$ in $\mathbf{y}_\zeta$, $\mathbf{y}_{uv}$, and $\mathbf{y}_{TS}$ are also uncorrelated to each other as well as any errors in $\delta\mathbf{x}_c$. Accordingly, an incremental 3DVAR (8) can be formulated as the weak constraints:

$$J_c(\delta\mathbf{x}_c) = J_{\zeta'}^b(\delta\mathbf{x}_{\zeta'}) + J_{\psi''\chi''}^b(\delta\mathbf{x}_{\psi''\chi''}) + J_{TS}^b(\delta\mathbf{x}_{TS})$$
$$+ J_\zeta^o(\delta\mathbf{x}_{\zeta'}, \delta\mathbf{x}_{TS}) + J_{uv}^o(\delta\mathbf{x}_{\zeta'}, \delta\mathbf{x}_{\psi''\chi''}, \delta\mathbf{x}_{TS})$$
$$+ J_{TS}^o(\delta\mathbf{x}_{TS}). \quad (17)$$

The background cost function breaks up into three parts:

$$J_{\zeta'}^b = \frac{1}{2}\delta\mathbf{x}_{\zeta'}^{\text{T}}\mathbf{B}_{\zeta'}^{-1}\delta\mathbf{x}_{\zeta'},$$

$$J_{\psi''\chi''}^b = \frac{1}{2}\delta\mathbf{x}_{\psi''\chi''}^{\text{T}}\mathbf{B}_{\psi''\chi''}^{-1}\delta\mathbf{x}_{\psi''\chi''},$$

$$J_{TS}^b = \frac{1}{2}\delta\mathbf{x}_{TS}^{\text{T}}\mathbf{B}_{TS}^{-1}\delta\mathbf{x}_{TS}, \quad (18)$$

where $\mathbf{B}_{\zeta'}$, $\mathbf{B}_{\psi''\chi''}$, and $\mathbf{B}_{TS}$ are the background error covariance matrices corresponding to nonsteric SSH, ageostrophic streamfunction and velocity potential, and T and S. The observational cost function also breaks up into three parts:

$$J_{\zeta'}^o = \frac{1}{2}(\boldsymbol{\Pi}\delta\mathbf{x}_{TS} + \delta\mathbf{x}_{\zeta'} - \mathbf{d}_\zeta)^{\text{T}}\mathbf{R}_\zeta^{-1}(\boldsymbol{\Pi}\delta\mathbf{x}_{TS} + \delta\mathbf{x}_{\zeta'} - d_\zeta),$$

$$J_{uv}^o = \frac{1}{2}(\boldsymbol{\Gamma}_{\zeta'}\delta\mathbf{x}_{\zeta'} + \boldsymbol{\Gamma}_{TS}\delta\mathbf{x}_{TS} + \boldsymbol{\Phi}\delta\mathbf{x}_{\psi''\chi''} - \mathbf{d}_{uv})^{\text{T}}$$
$$\times \mathbf{R}_{uv}^{-1}(\boldsymbol{\Gamma}_{\zeta'}\delta\mathbf{x}_{\zeta'} + \boldsymbol{\Gamma}_{TS}\delta\mathbf{x}_{TS} + \boldsymbol{\Phi}\delta\mathbf{x}_{\psi''\chi''} - \mathbf{d}_{uv}),$$

$$J_{TS}^o = \frac{1}{2}(\delta\mathbf{x}_{TS} - \mathbf{d}_{TS})^{\text{T}}\mathbf{R}_{TS}^{-1}(\delta\mathbf{x}_{TS} - \mathbf{d}_{TS}), \quad (19)$$

where $\mathbf{R}_\zeta$, $\mathbf{R}_{uv}$, and $\mathbf{R}_{TS}$ are the error covariance matrices of $\mathbf{e}_\zeta$, $\mathbf{e}_{uv}$, and $\mathbf{e}_{TS}$, respectively. The resulting cost function (17) possesses several desirable properties in comparison to (8), as follows.

First, the sizes of $\mathbf{B}_\zeta$, $\mathbf{B}_{\psi''\chi''}$, and $\mathbf{B}_{TS}$ are considerably smaller than that of $\mathbf{B}$, which greatly reduces the difficulty in the estimation of error covariance and the computational load in minimizing $J_c(\delta\mathbf{x}_c)$, as will be discussed in section 4.

Second, $\delta\mathbf{x}_{\psi''\chi''}$ is used as a control variable, rather than $\delta\mathbf{x}_{u''v''}$, in $\delta\mathbf{x}_c$. It is known that the self-error covariances of $\delta\mathbf{x}_{u''}$ and $\delta\mathbf{x}_{v''}$ are always nonhomogeneous and anisotropic, and their cross-error covariance is nonzero. Practically, it is difficult to estimate and simplify the error covariance, and thus it is difficult to use $\delta\mathbf{x}_{u''}$ and $\delta\mathbf{x}_{v''}$ in data assimilation (Daley 1991). In contrast, the self-error covariance of streamfunction and velocity potential is much more homogeneous and isotropic, and their cross-error covariance tends to be negligible (estimation and simplification of forecast error covariances are discussed in depth in section 4).

Third, if there are no observations in velocity and sea surface height to be assimilated, the analysis of $\delta\mathbf{x}_{\psi''\chi''}$ and $\delta\mathbf{x}_{\zeta'}$ shows them to be zero vectors. This is due to the implicit assumption of the weak constraint formulation, that is, the errors in the unbalanced part of $\delta\mathbf{x}_{\psi''\chi''}$ and $\delta\mathbf{x}_{\zeta'}$ are not correlated with those of $\delta\mathbf{x}_{TS}$. Accordingly, the measurements in T, S, and SSH are not able to correct ageostrophic velocities and nonsteric SSH.

Fourth, the cost function $J_c(\delta\mathbf{x}_c)$ permits substantial flexibility in implementation. It is straightforward to add additional dynamic constraints. For example, setting the ageostrophic velocity potential to zero forces $\delta\mathbf{x}_{u''v''}$ to be nondivergent.

Finally, if the assumptions of uncorrelated errors for $\delta\mathbf{x}_c$ and $\mathbf{e}$ are violated, (17) needs to be extended (Daley 1991). When necessary, we can add more terms to (18), such as penalty or regularization terms as suggested by the theory for an ill-posed problem (Tikhonov and Arsenin 1977), which may improve the performance of optimization.

4. Construction of background error covariance

Construction of error covariances is central to the development of any 3DVAR system. The performance of ROMS3DVAR hinges on the accurate description of $\mathbf{B}$ and $\mathbf{R}$ in (18) and (19). The choice of $\mathbf{R}$ will not be discussed in this paper because it depends specifically on the properties of observations. Statistically, an accurate estimate of $\mathbf{B}$ is required to render the analysis $\mathbf{x}_c^a$ the maximum likelihood estimate. Practically, $\mathbf{B}$ plays the role of spreading out observational information contained in $\mathbf{d}$'s, which is especially crucial when observations are sporadic and irregularly distributed.

In general, there are two challenges in constructing

B. First, we do not know $\mathbf{x}^t$, and thus $\mathbf{x}^f - \mathbf{x}^t$ [see (7)]. Although the observations may be assumed to sample $\mathbf{x}^t$, the available observations are too few to reliably estimate **B**. To address this challenge, the National Meteorological Center (NMC; currently the Environmental Center for Environmental Prediction) method (Parrish and Derber 1992; Rabier et al. 1998b) is generally employed to generate data to represent forecast errors, and then the generated data are used for estimating model error correlations. It is also possible to follow the method proposed by Hollingsworth and Lönnberg (1986). In that method the forecast error is estimated as the difference between the model forecast and raw observations and then the correlated length is estimated. In ROMS3DVAR, we use the NMC method due to the lack of raw observations. Second, the size of **B** is too large to handle numerically. For a realistic coastal ocean model, the number of model grid points is generally on the order of $O(10^6)$. The number of elements in **B** is therefore $O(10^{12})$. It is impossible to perform explicit manipulation of matrices of this size even on supercomputers. This is true even for $\mathbf{B}_{\zeta'}$, $\mathbf{B}_{\psi''\chi''}$, and $\mathbf{B}_{TS}$, although they are much smaller than **B** for the standard (incremental) 3DVAR method. Hence, it is necessary to simplify them.

This section is devoted to addressing this second challenge in constructing $\mathbf{B}_{\zeta'}$, $\mathbf{B}_{\psi''\chi''}$ and $\mathbf{B}_{TS}$ that are computationally tractable yet physically appropriate for coastal data assimilation. For this purpose, we factorize them as

$$\mathbf{B}_{\zeta'} = \mathbf{\Sigma}_{\zeta'}\mathbf{C}_{\zeta'\zeta'}^{\xi\eta}\mathbf{\Sigma}_{\zeta'},$$

$$\mathbf{B}_{\psi''\chi''} = \mathbf{\Sigma}_{\psi''\chi''}\begin{pmatrix} \mathbf{C}_{\psi''\psi''}^{\xi\eta\kappa} & \mathbf{C}_{\psi''\chi''}^{\xi\eta\kappa} \\ \mathbf{C}_{\chi''\psi''}^{\xi\eta\kappa} & \mathbf{C}_{\chi''\chi''}^{\xi\eta\kappa} \end{pmatrix}\mathbf{\Sigma}_{\psi''\chi''},$$

$$\mathbf{B}_{TS} = \mathbf{\Sigma}_{TS}\begin{pmatrix} \mathbf{C}_{TT}^{\xi\eta\kappa} & \mathbf{C}_{TS}^{\xi\eta\kappa} \\ \mathbf{C}_{ST}^{\xi\eta\kappa} & \mathbf{C}_{SS}^{\xi\eta\kappa} \end{pmatrix}\mathbf{\Sigma}_{TS}. \tag{20}$$

Here $\mathbf{\Sigma}_{\zeta'}$, $\mathbf{\Sigma}_{\psi''\chi''}$, and $\mathbf{\Sigma}_{TS}$ are standard deviation matrices that are diagonal with the effective dimension being the same as the corresponding incremental vectors. Therefore, they can be used without simplification. Because of the assumption made for the errors among the five variables, there are five self-correlation matrices, $\mathbf{C}_{\zeta'\zeta'}^{\xi\eta}$, $\mathbf{C}_{\psi''\psi''}^{\xi\eta\kappa}$, $\mathbf{C}_{\chi''\chi''}^{\xi\eta\kappa}$, $\mathbf{C}_{TT}^{\xi\eta\kappa}$, and $\mathbf{C}_{SS}^{\xi\eta\kappa}$, and two cross-correlation matrices, $\mathbf{C}_{\psi''\chi''}^{\xi\eta\kappa} = (\mathbf{C}_{\chi''\psi''}^{\xi\eta\kappa})^{\mathrm{T}}$ and $\mathbf{C}_{TS}^{\xi\eta\kappa} = (\mathbf{C}_{ST}^{\xi\eta\kappa})^{\mathrm{T}}$. The superscripts of the correlation matrices indicate the spatial dimensionality of the correlation matrices, with ξ, η, and κ representing the cross-shore, alongshore, and vertical directions, respectively. The subscripts represent the two corresponding control variables. The cross-correlation matrices $\mathbf{C}_{\psi''\chi''}^{\xi\eta\kappa}$ and $\mathbf{C}_{TS}^{\xi\eta\kappa}$ should not be

confused with the correlation matrices of $\mathbf{C}_{\psi''\chi''}$ and $\mathbf{C}_{TS}$, which are the correlation matrices corresponding to $\mathbf{x}_{\psi''\chi''}$ and $\mathbf{x}_{TS}$, respectively. In contrast to the standard deviation matrices, self- and cross-correlation matrices may be dense and hence require simplification because of their prohibitive size.

a. Self-correlation matrices

1) APPROACH

For convenience, we introduce some new notations. Since ROMS employs curvilinear coordinates, we assume that the horizontal coordinates are in the cross-shore and alongshore directions. For simplicity, we consider a model grid box of dimensions $N^\xi \times N^\eta \times N^\kappa$. Extension to an irregular model grid is straightforward.

Since the same formulation is used for the construction of all five self-correlation matrices, we simply use the notation **C** in the following discussion. We further employ the following notations: $\mathbf{C}^\xi$ for any one-dimensional (1D) cross-shore correlation matrix with the dimension $(N^\xi)^2$; $\mathbf{C}^\eta$ for any 1D alongshore correlation matrix with dimension $(N^\eta)^2$; $\mathbf{C}^{\xi\eta}$ for any two-dimensional (2D) horizontal correlation matrix with dimension $(N^\xi N^\eta)^2$ and $\mathbf{C}^{\xi\kappa}$ for any 2D vertical cross-shore correlation matrix with dimension $(N^\xi N^\kappa)^2$; and $\mathbf{C}^{\xi\eta\kappa}$ for any three-dimensional (3D) correlation matrix with dimension $(N^\xi N^\eta N^\kappa)^2$. Generally, 1D correlations are manageable, and can be constructed with various methods, while simplification is necessary for the 2D and 3D correlations for a sophisticated model. As correlation matrices, any **C** must be positive definite.

A variety of methods for constructing **C** has been developed by the data assimilation community. Two basic classes are extensively used in operational data assimilation systems. The first class is to directly reduce the degrees of freedom of **C**. The simplest method in the first class is to assume **C** to be diagonal. This assumption is not acceptable for 3DVAR associated with a grid model, since this assumption implies that there are no horizontal correlations. It is useful only with spectral models as implemented for atmospheric data assimilation systems such as that at the European Centre for Medium-Range Weather Forecasts (ECMWF) (Courtier et al. 1998), since it is equivalent to the assumption of homogeneity and isotropy for **B** (Boer 1983). A more sophisticated method in this class is the so-called rank reduction. With this method, **C** is computed on a resolution lower than the one for the forecast model. This method has been applied to global ocean data assimilation (e.g., Fukumori et al. 1999).

The second class can be described as parameterization algorithms. One frequently used method in this
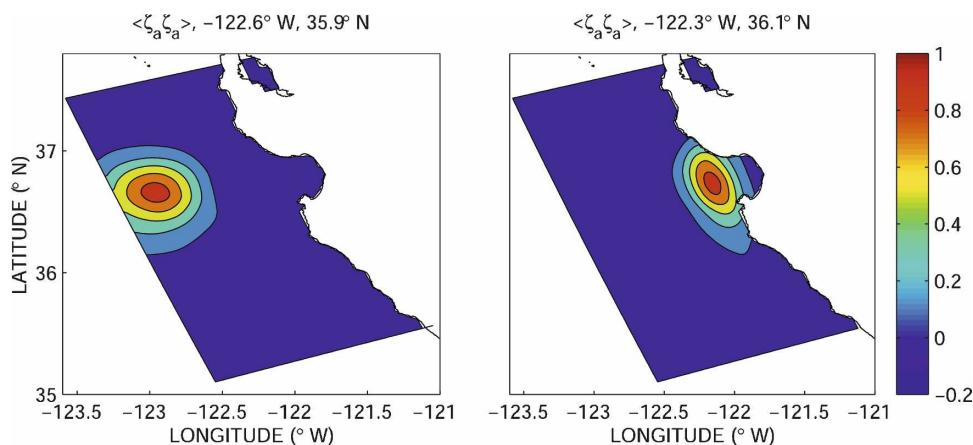
Fig. 1. Actual 2D point correlation of nonsteric SSH at (left) an offshore and (right) a nearshore model grid point. The contour interval is 0.2. The dataset used to compute the correlation is discussed in Li et al. (2008).

class is the computation of all the entries of **C** with a given formula along with a set of prescribed parameters. A classic example is to compute **C** by assuming that the correlation function is a Gaussian with a prescribed correlated radius. Other known methods include the use of second- or third-order autoregression (SOAR–TOAR), or particular polynomial functions (Gaspari and Cohn 1999). There are some merits to using such methods: they are relatively easy to implement; the prescribed parameters may be derived by using a limited number of observations (e.g., Dee 1995; Lönnberg and Hollingsworth 1986; Hollingsworth and Lönnberg 1986; Xu and Li 2001); and instead of saving all the entries of **C**, they can be instantly computed. The last merit is particularly important in increasing computational feasibility. The methods of this class have been implemented in operational global atmospheric data assimilation systems (e.g., Cohn et al. 1998; Daley and Barker 2001). The limitation of this class is that it can generate mostly only isotropic and homogeneous correlations. The recursive filter method also belongs to this class, but it has the capability of modeling some aspects of anisotropic and inhomogeneous correlations (Purser et al. 2003).

Following the reasoning of the parameterization used in the second class, we propose a new method. This new method has the capability of expressing the major anisotropic and inhomogeneous features of correlations for coastal oceans. In the coastal ocean the anisotropy and inhomogeneity of self-correlations are not negligible. The typical expression of this anisotropy and inhomogeneity is that self-correlations far offshore have larger spatial scales in the cross-shore direction than those nearshore, and correlations nearshore have larger spatial scales along the shore than cross-shore. Such

spatial features of correlations can be seen in Fig. 1, where the correlations are computed from model-generated data as described in Li et al. (2008). Further, the constructed error correlation matrices have other advantages that are discussed later. For example, they guarantee positive definiteness and can be numerically manipulated efficiently and reliably.

The formulation of this new method hinges on the Kronecker product. The Kronecker product is also known as the tensor product or direct product. For its general definition and properties, readers are referred to Graham (1981). Appendix B presents some useful properties of the Kronecker product in constructing error covariance matrices.

2) TWO-DIMENSIONAL ERROR CORRELATION $\mathbf{C}^{\xi\eta}$

With the Kronecker product we construct a 2D correlation matrix $\mathbf{C}^{\xi\eta}$ from two 1D correlation matrices, $\mathbf{C}^{\xi}$ and $\mathbf{C}^{\eta}$:

$$\mathbf{C}^{\xi\eta} = \mathbf{C}^{\eta} \otimes \mathbf{C}^{\xi}, \qquad (21)$$

where the definition and detailed properties of the Kronecker product can be found in appendix B. This is practically equivalent to choosing the elements of $\mathbf{C}^{\xi\eta}$ to be the correlations between the 2D grid points $(i_1 : j_1)$ and $(i_2 : j_2)$ given by

$$C^{\xi\eta}(i_1 : j_1, i_2 : j_2) = C^{\xi}(i_1, i_2)C^{\eta}(j_1, j_2), \qquad (22)$$

where $C^{\xi}(i_1, i_2)$ is the element of $\mathbf{C}^{\xi}$ corresponding to correlation between the 1D ξ grid points i_1 and i_2, and $C^{\eta}(j_1, j_2)$ follows the same convention for 1D η grid points. Quantitatively, such a $\mathbf{C}^{\xi\eta}$ can be an accurate approximation to the 2D correlation in many circumstances.
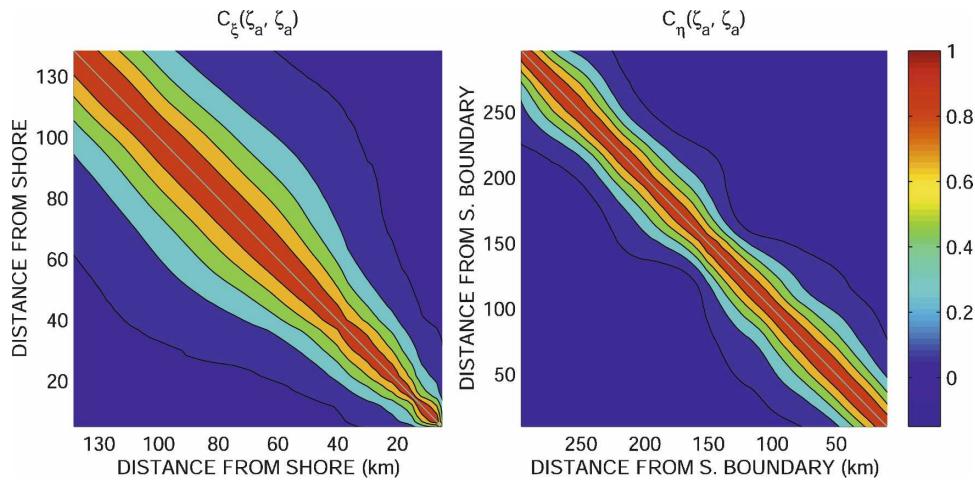
FIG. 2. The 1D correlations of the nonsteric SSH in the cross-shore and alongshore directions, respectively. The contour interval is 0.2.

One desirable characteristic of the Kronecker product formulation is that we can directly construct relatively sophisticated 2D anisotropic and inhomogeneous correlations using data. With the NMC method, we can construct the 2D correlations as follows: the element of the 1D correlation $\mathbf{C}^\xi$ is computed as the correlation between the two columns of the model grids, and the element of the 1D correlation $\mathbf{C}^\eta$ as the correlation between two rows of the model grids. The 2D correlation can then be constructed with the Kronecker product of $\mathbf{C}^\xi$ and $\mathbf{C}^\eta$. The correlation computed in this way will be detailed in Li et al. (2008). Here we present the nonsteric SSH correlation $\mathbf{C}^{\xi\eta}_{\zeta'\zeta'}$ as an illustration. The computed 1D correlation is shown in Fig. 2.

The cross-shore 1D correlation $\mathbf{C}^\xi_{\zeta'\zeta'}$ displays significant inhomogeneity. The decorrelation length scale varies from about 15 km nearshore to about 35 km offshore. The alongshore 1D correlation $\mathbf{C}^\eta_{\zeta'\zeta'}$ is basically homogeneous, with a decorrelation length scale of about 35 km. From the decorrelation length scale in these two 1D correlations, we learn that the 2D correlation is approximately isotropic offshore with a decorrelation length scale of about 35 km, while it is significantly anisotropic nearshore with an alongshore decorrelation length scale 2 times larger than the cross-shore one. Such inhomogeneity and anisotropy can be illustrated in the 2D correlation constructed with the Kronecker product as shown in Fig. 3, which reflects the
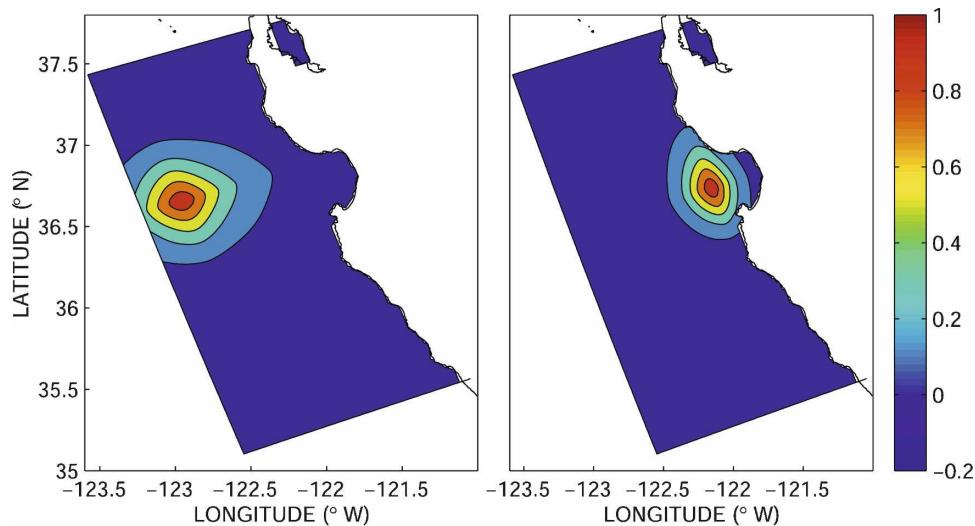

FIG. 3. 2D point correlation of nonsteric SSH constructed using the Kronecker product–based algorithm and the two 1D correlations shown in Fig. 2 at the two locations in Fig. 1.

inhomogeneity and anisotropy seen in Fig. 1, though the correlation computed from the Kronecker product is not exactly the same as the correlation computed from the data.

3) THREE-DIMENSIONAL ERROR CORRELATION $\mathbf{C}^{\xi\eta\kappa}$

Like horizontal 2D error correlation matrices, a 3D error correlation matrix can also be constructed using the Kronecker product formulation. In a simple form it may be constructed from the three 1D correlation matrices $\mathbf{C}^\xi$, $\mathbf{C}^\eta$, and $\mathbf{C}^\kappa$, that is, $\mathbf{C}^{\xi\eta\kappa} = \mathbf{C}^\xi \otimes \mathbf{C}^\eta \otimes \mathbf{C}^\kappa$, where $\mathbf{C}^\kappa$ is the 1D correlation matrix in the vertical direction. This formulation implies the separability of the correlations in the three spatial dimensions. For coastal oceans, horizontal separability may be acceptable. Also the separability between the alongshore and vertical directions can be used when the model domain does not span a long distance alongshore, and when the coastline is relatively straight such as the area off California. However, the correlation is usually not separable between the vertical and cross-shore directions, because the vertical correlation function depends on the distance away from the coastline cross-shore. This dependence can be significant, especially when strong upwelling is present. The upwelling raises the thermocline nearshore, and results in a shallower mixed layer nearshore than in areas farther offshore. The difference in the mixed layer vertical structure, in turn, causes the difference in the vertical correlation.

Such inseparability of the correlation between the vertical and cross-shore directions can be incorporated using the Kronecker product formulation. To do so, we first construct the 2D error correlation $\mathbf{C}^{\xi\kappa}$. The 3D correlation is then constructed with the Kronecker product

$$\mathbf{C}^{\xi\eta\kappa} = \mathbf{C}^{\xi\kappa} \otimes \mathbf{C}^\eta. \tag{23}$$

4) COMPUTATIONAL ISSUES

Having described the construction of the anisotropic, inhomogeneous, and inseparable 3D error correlation $\mathbf{C}^{\xi\eta\kappa}$ using the Kronecker product formulation, we now describe some computational issues in implementing $\mathbf{C}^{\xi\eta\kappa}$ in 3DVAR. The Kronecker product formulation requires limited memory. For a coastal ocean model the number of grid points in both the alongshore and cross-shore directions are on the order of $O(10^3)$, and the number of the vertical levels is on the order of $O(10^1)$. Therefore, the memory required to define a symmetric matrix $\mathbf{C}^{\xi\eta\kappa}$ is $(1/2)(N^\xi N^\kappa)^2 + (1/2)(N^\eta)^2$ and is of $O(10^8)$. This memory requirement is easily satisfied by

present day computers. Further, the positive definiteness of $\mathbf{C}^{\xi\eta\kappa}$ is guaranteed if $\mathbf{C}^{\xi\kappa}$ and $\mathbf{C}^\eta$ are individually positive definite. More mathematical details can be found in appendix B.

b. Cross-correlation matrices

As shown in (20), ROMS3DVAR uses two 3D cross-correlation matrices, $\mathbf{C}^{\xi\eta\kappa}_{\psi''\chi''}$ and $\mathbf{C}^{\xi\eta\kappa}_{TS}$. We parameterize $\mathbf{C}^{\xi\eta\kappa}_{TS}$ as the average amplitude of the cross correlations and the average decorrelation length scale of the corresponding self-correlations:

$$\mathbf{C}^{\xi\eta\kappa}_{TS} = \mathbf{C}^{\xi\eta\kappa}_{mTS} \odot \mathbf{C}^{\xi\eta\kappa}_{mTS}, \tag{24}$$

where $\odot$ stands for the Hadamard (or Schur) product. The amplitude matrix $\mathbf{D}^{\xi\eta\kappa}_{mTS}$ is obtained by the local cross-correlation vector $\mathbf{r}_{TS}$ arranged in the same order as $\delta\mathbf{x}_{TS}$, and $\mathbf{D}^{\xi\eta\kappa}_{rTS}$ whose column vectors are all $\mathbf{r}_{TS}$:

$$\mathbf{D}^{\xi\eta\kappa}_{mTS} = \frac{1}{2}[\mathbf{D}^{\xi\eta\kappa}_{rTS} + (\mathbf{D}^{\xi\eta\kappa}_{rTS})^{\mathrm{T}}]. \tag{25}$$

By this construction, the diagonal elements of $\mathbf{D}^{\xi\eta\kappa}_{mTS}$ are the local cross correlations, and the off-diagonal elements are the average of local cross correlations at the two distinct locations. The average of the two self-correlation matrices

$$\mathbf{C}^{\xi\eta\kappa}_{mTS} = \frac{1}{2}(\mathbf{C}^{\xi\eta\kappa}_{TT} + \mathbf{C}^{\xi\eta\kappa}_{SS}) \tag{26}$$

sets the decorrelation length scale of the cross correlation.

This approximation of $\mathbf{C}^{\xi\eta\kappa}_{TS}$ constitutes an extension of the method used for modifying S with T observations in a few studies (e.g., Vossepoel et al. 1999; Troccoli and Haines 1999), where the off-diagonal elements of $\mathbf{C}^{\xi\eta\kappa}_{TS}$ are approximated to be zero. Such an approximation implies that $\delta\mathbf{x}_T$ and $\delta\mathbf{x}_S$ have a nonzero cross correlation only at the same location. In the approximation given by (24), the cross correlation of $\delta\mathbf{x}_T$ and $\delta\mathbf{x}_S$ is allowed among distinct locations.

Although $\mathbf{C}^{\xi\eta\kappa}_{\psi''\chi''}$ between the ageostrophic streamfunction and the ageostrophic velocity potential can be obtained using the same procedure as $\mathbf{C}^{\xi\eta\kappa}_{TS}$, we choose a zero cross correlation for $\mathbf{C}^{\xi\eta\kappa}_{\psi''\chi''}$ in ROMS3DVAR. This is made possible by the use of $\delta\mathbf{x}_{\psi''}$ and $\delta\mathbf{x}_{\chi''}$ as control variables, rather than $\delta\mathbf{x}_{u''}$ and $\delta\mathbf{x}_{v''}$. The zero cross-correlation $\mathbf{C}^{\xi\eta\kappa}_{\psi''\chi''}$ leads to nonzero cross-correlation $\mathbf{C}^{\xi\eta\kappa}_{u''v''}$ between the ageostrophic horizontal velocities that are adequately represented in various circumstances (Daley 1991, chapter 5.2).

It is noteworthy that the use of cross correlation between $\delta\mathbf{x}_T$ and $\delta\mathbf{x}_S$ can be very complicated in data assimilation, particularly for coastal oceans. One major

concern is that the cross correlation may not be as steady in coastal oceans as it is in the tropics and hence a flow-dependent T–S correlation may be required. Also, the approximation in the cross correlation may introduce nonpositiveness in the covariance matrix $\mathbf{B}_{TS}$. It may require tweaking $\mathbf{D}_{mTS}^{\xi\eta\kappa}$ to guarantee the positive definiteness of $\mathbf{B}_{TS}$. More discussions of the cross correlations can be found in Li et al. (2008).

To construct computationally feasible $\mathbf{B}_{\psi''\chi''}$ and $\mathbf{B}_{TS}$, we further apply Cholesky factorizations to the five self-correlation matrices obtained in section 4a,

$$\mathbf{C}_{\zeta'}^{\xi\eta} = \mathbf{L}_{\zeta'}^{\xi\eta}(\mathbf{L}_{\zeta'}^{\xi\eta})^{\mathrm{T}},$$

$$\mathbf{C}_{\psi''\psi''}^{\xi\eta\kappa} = \mathbf{L}_{\psi''\psi''}^{\xi\eta\kappa}(\mathbf{L}_{\psi''\psi''}^{\xi\eta\kappa})^{\mathrm{T}} \quad \mathbf{C}_{\chi''\chi''}^{\xi\eta\kappa} = \mathbf{L}_{\chi''\chi''}^{\xi\eta\kappa}(\mathbf{L}_{\chi''\chi''}^{\xi\eta\kappa})^{\mathrm{T}},$$

$$\mathbf{C}_{TT}^{\xi\eta\kappa} = \mathbf{L}_{TT}^{\xi\eta\kappa}(\mathbf{L}_{TT}^{\xi\eta\kappa})^{\mathrm{T}} \quad \mathbf{C}_{SS}^{\xi\eta\kappa} = \mathbf{L}_{SS}^{\xi\eta\kappa}(\mathbf{L}_{SS}^{\xi\eta\kappa})^{\mathrm{T}}, \qquad (27)$$

where $\mathbf{L}_{\zeta'}^{\xi\eta}$, $\mathbf{L}_{\psi''\psi''}^{\xi\eta\kappa}$, $\mathbf{L}_{\chi''\chi''}^{\xi\eta\kappa}$, $\mathbf{L}_{TT}^{\xi\eta\kappa}$, and $\mathbf{L}_{SS}^{\xi\eta\kappa}$ are lower triangular matrices and are also referred to as Cholesky triangles. Since these self-error correlation matrices are constructed with the Kronecker product formulations, the Cholesky factorizations (27) are practically feasible as shown in appendix B.

Overall, (20) with (27) and (24) gives

$$\mathbf{B}_{\zeta'} = \mathbf{\Sigma}_{\zeta'}\mathbf{L}_{\zeta'\zeta'}\mathbf{L}_{\zeta'\zeta'}^{\mathrm{T}}\mathbf{\Sigma}_{\zeta'},$$

$$\mathbf{B}_{\psi''\chi''} = \mathbf{\Sigma}_{\psi''\chi''}\mathbf{L}_{\psi''\chi''}\mathbf{W}_{\psi''\chi''}\mathbf{L}_{\psi''\chi''}^{\mathrm{T}}\mathbf{\Sigma}_{\psi''\chi''},$$

$$\mathbf{B}_{TS} = \mathbf{\Sigma}_{TS}\mathbf{L}_{TS}\mathbf{W}_{TS}\mathbf{L}_{TS}^{\mathrm{T}}\mathbf{\Sigma}_{TS}, \qquad (28)$$

where $L_{\zeta'\zeta'} = L_{\zeta'\zeta'}^{\xi\eta}$ and

$$\mathbf{D}_{\psi''\chi''}^{\xi\eta\kappa} = (\mathbf{L}_{\psi''\psi''}^{\xi\eta\kappa})^{-1}\mathbf{C}_{\psi''\chi''}^{\xi\eta\kappa}[(\mathbf{L}_{\chi''\chi''}^{\xi\eta\kappa})^{-1}]^{\mathrm{T}},$$

$$\mathbf{D}_{TS}^{\xi\eta\kappa} = (\mathbf{L}_{TT}^{\xi\eta\kappa})^{-1}\mathbf{C}_{TS}^{\xi\eta\kappa}[(\mathbf{L}_{SS}^{\xi\eta\kappa})^{-1}]^{\mathrm{T}},$$

$$\mathbf{L}_{\psi''\chi''} = \begin{pmatrix} \mathbf{L}_{\psi''\psi''}^{\xi\eta\kappa} & \mathbf{0}_{N_{\psi''}\times N_{\chi''}} \\ \mathbf{0}_{N_{\chi''}\times N_{\psi''}} & \mathbf{L}_{\chi''\chi''}^{\xi\eta\kappa} \end{pmatrix},$$

$$\mathbf{W}_{\psi''\chi''} = \begin{pmatrix} \mathbf{I}_{N_{\psi''}\times N_{\psi''}} & \mathbf{D}_{\psi''\chi''}^{\xi\eta\kappa} \\ (\mathbf{D}_{\psi''\chi''}^{\xi\eta\kappa})^{\mathrm{T}} & \mathbf{I}_{N_{\chi''}\times N_{\chi''}} \end{pmatrix},$$

$$\mathbf{L}_{TS} = \begin{pmatrix} \mathbf{L}_{TT}^{\xi\eta\kappa} & \mathbf{0}_{N_T\times N_S} \\ \mathbf{0}_{N_S\times N_T} & \mathbf{L}_{SS}^{\xi\eta\kappa} \end{pmatrix},$$

$$\mathbf{W}_{TS} = \begin{pmatrix} \mathbf{I}_{N_T\times N_T} & \mathbf{D}_{TS}^{\xi\eta\kappa} \\ (\mathbf{D}_{TS}^{\xi\eta\kappa})^{\mathrm{T}} & \mathbf{I}_{N_S\times N_S} \end{pmatrix}. \qquad (29)$$

5. Minimization procedures

After all the matrices are given, the next major computational task is to minimize the cost functions (17). Because of the large dimension, this minimization must be done using particular algorithms. In the past two decades, minimization algorithms for large-scale problems have been developed by the computational community (e.g., Nocedal and Wright 1999). Two types of minimization algorithm have been applied in data assimilation. The first type is preconditioned conjugate gradient algorithms, and the second type is limited memory quasi-Newton algorithms. These two methods minimize the cost function iteratively using the information on the gradient of the cost function. ROMS3DVAR uses a limited memory quasi-Newton algorithm (Liu and Nocedal 1989) that can update the preconditioning during the minimization procedure.

In the implementation of these iterative algorithms, two major steps need to be carefully considered. First, a preconditioning is applied to improve the convergence of the iteration. Second, the cost functions and their gradients must be accurately and efficiently computed. These two issues are addressed in this section.

a. Preconditioning of cost functions

When minimizing the $J_c(\delta\mathbf{x}_c)$ defined by (17), a well-known numerical difficulty may be encountered in computing the inverse of the large-dimensional $\mathbf{B}_{\zeta'}$, $\mathbf{B}_{\psi''}$, and $\mathbf{B}_{TS}$ in J_{TS}^b, J_ζ^b, and J_{uv}^b. In contrast, the observational cost functions J_{TS}^o, J_ζ^o, and J_{uv}^o are generally less troublesome because $\mathbf{R}_{TS}$, $\mathbf{R}_\zeta$, and $\mathbf{R}_{uv}$ are mostly diagonal with nearly uniform values in the diagonal elements. These matrices' behavior is more like the identity matrix during the minimization procedure. The other primary numerical difficulty associated with minimization is that (17) might be ill-conditioned. When it is ill-conditioned, the minimization procedure may converge at a slow rate or not at all (e.g., Nocedal and Wright 1999). To improve the efficiency and reliability of the minimization procedure, preconditioning is often employed in 3DVAR (e.g., Courtier et al. 1994). It mainly focuses on the transformation of $\mathbf{x}_c$ to improve the minimization of J_{TS}^b, J_ζ^b, and J_{uv}^b.

With the Cholesky factorizations of $\mathbf{B}_{\zeta'}$, $\mathbf{B}_{\psi''}$, and $\mathbf{B}_{TS}$ as in (28), we implement the transformation from $\delta\mathbf{x}_c$ to $\delta\mathbf{z}_c = (\delta\mathbf{z}_{\zeta'}^{\mathrm{T}}, \delta\mathbf{z}_{\psi''\chi''}^{\mathrm{T}}, \delta\mathbf{z}_{TS}^{\mathrm{T}})^{\mathrm{T}}$ defined by

$$\delta\mathbf{x}_{\zeta'} = \mathbf{\Sigma}_{\zeta'}\mathbf{L}_{\zeta'}\delta\mathbf{z}_{\zeta'},$$

$$\delta\mathbf{x}_{\psi''\chi''} = \mathbf{\Sigma}_{\psi''\chi''}\mathbf{L}_{\psi''\chi''}\mathbf{W}_{\psi''\chi''}\delta\mathbf{z}_{\psi''\chi''},$$

$$\delta\mathbf{x}_{TS} = \mathbf{\Sigma}_{TS}\mathbf{L}_{TS}\mathbf{W}_{TS}\delta\mathbf{z}_{TS}. \qquad (30)$$

This yields (17), the cost function of present ROMS3DVAR with $\delta\mathbf{z}_c$ as the control variables

$$J_c(\delta\mathbf{z}_c) = J_{\zeta'}^b(\delta\mathbf{z}_{\zeta'}) + J_{\psi''\chi''}^b(\delta\mathbf{z}_{\psi''\chi''}) + J_{TS}^b(\delta\mathbf{z}_{TS})$$

$$+ J_\zeta^b(\delta\mathbf{z}_{\zeta'}, \delta\mathbf{z}_{TS}) + J_{uv}^b(\delta\mathbf{z}_{\zeta'}, \delta\mathbf{z}_{\psi''\chi''}, \delta\mathbf{z}_{TS})$$

$$+ J_{TS}^o(\delta\mathbf{z}_{TS}), \qquad (31)$$

where

$$J_{\zeta'}^b = \frac{1}{2}\,\delta\mathbf{z}_{\zeta'}^{\mathrm{T}}\delta\mathbf{z}_{\zeta'},$$

$$J_{\psi'\chi''}^b = \frac{1}{2}\,\delta\mathbf{z}_{\psi'\chi''}^{\mathrm{T}}\mathbf{W}_{\psi'\chi''}\delta\mathbf{z}_{\psi'\chi''},$$

$$J_{TS}^b = \frac{1}{2}\,\delta\mathbf{z}_{TS}^{\mathrm{T}}\mathbf{W}_{TS}\delta\mathbf{z}_{TS} \tag{32}$$

are the preconditioned background cost function (18) and

$$J_{\zeta'}^o = \frac{1}{2}(\mathbf{H}_\zeta\delta\mathbf{x}_\zeta^\dagger - \mathbf{d}_\zeta)^{\mathrm{T}}\mathbf{R}_\zeta^{-1}(\mathbf{H}_\zeta\delta\mathbf{x}_\zeta^\dagger - \mathbf{d}_\zeta),$$

$$J_{uv}^o = \frac{1}{2}(\mathbf{H}_{uv}\delta\mathbf{x}_{uv}^\dagger - \mathbf{d}_{uv})^{\mathrm{T}}\mathbf{R}_{uv}^{-1}(\mathbf{H}_{uv}\delta\mathbf{x}_{uv}^\dagger - \mathbf{d}_{uv}),$$

$$J_{TS}^o = \frac{1}{2}(\mathbf{H}_{uv}\delta\mathbf{x}_{TS}^\dagger - \mathbf{d}_{TS})^{\mathrm{T}}\mathbf{R}_{TS}^{-1}(\mathbf{H}_{uv}\delta\mathbf{x}_{TS}^\dagger - \mathbf{d}_{TS}) \tag{33}$$

are observational cost functions (19) that are not preconditioned with respect to $\delta\mathbf{z}_c$. Here

$$\delta\mathbf{x}_\zeta^\dagger = \boldsymbol{\Sigma}_{\zeta'}\mathbf{L}_{\zeta'}\delta\mathbf{z}_{\zeta'} + \boldsymbol{\Pi}_{TS}\boldsymbol{\Sigma}_{TS}\mathbf{L}_{TS}\mathbf{W}_{TS}\delta\mathbf{z}_{TS},$$

$$\delta\mathbf{x}_{uv}^\dagger = \boldsymbol{\Gamma}_{\zeta'}\boldsymbol{\Sigma}_{\zeta'}\mathbf{L}_{\zeta'}\delta\mathbf{z}_{\zeta'} + \boldsymbol{\Gamma}_{TS}\boldsymbol{\Sigma}_{TS}\mathbf{L}_{TS}\mathbf{W}_{TS}\delta\mathbf{z}_{TS}$$
$$+ \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\psi''\chi''}\mathbf{L}_{\psi''\chi''}\mathbf{W}_{\psi''\chi''}\delta\mathbf{z}_{\psi''\chi''},$$

$$\delta\mathbf{x}_{TS}^\dagger = \boldsymbol{\Sigma}_{TS}\mathbf{L}_{TS}\mathbf{W}_{TS}\delta\mathbf{z}_{TS} \tag{34}$$

correspond to the increment $\delta\mathbf{x}$ in the full model space (8) but are computed from $\delta\mathbf{z}_c$ using preconditioning (30) as well as the dynamical constraints (12) and (13).

With this transformation (30), the condition of J^b is improved for efficient and reliable minimization: $J_{\zeta'}^b$ is perfectly conditioned; and $J_{\psi'\chi''}^b$ and J_{TS}^b are well conditioned, too, because $\mathbf{W}_{\psi'\chi''}$ and $\mathbf{W}_{TS}$ are almost identity matrices because of the small values of the diagonals in $\mathbf{D}(\delta\mathbf{x}_{\psi''}, \delta\mathbf{x}_{\chi''})$ and $\mathbf{D}(\delta\mathbf{x}_T, \delta\mathbf{x}_S)$. Further, inversion of error covariance matrices is no longer necessary. The observational cost functions are the same as in (19), but $\delta\mathbf{x}_c$ is now computed from $\delta\mathbf{z}_c$ by linear transformations given by (30).

Once $\delta\mathbf{z}_c^a$, which minimizes (31), is obtained using the limited memory quasi-Newton algorithm, the analysis increment $\delta\mathbf{x}_c^a = \delta\mathbf{x}_c^\dagger$ is obtained by setting $\delta\mathbf{z}_c = \delta\mathbf{z}_c^a$ in (34). The final analysis $\mathbf{x}^a$ is then obtained using (9).

b. Remarks on programming procedures and computing

The major computational overhead in minimizing the cost function is due to the matrix–vector products used to compute $\delta\mathbf{x}_c^\dagger$ given by (34). The Cholesky factoriza-

tion reduces the computational overhead by approximately one-half. Although we also employ the property of localized correlations, those matrix–vector products still use more than 80% of the total CPU time for each minimization iteration. Fortunately, these matrix–vector products are not difficult to parallelize in a load-balanced way with both OpenMP and MPI parallelism.

Another computational issue arises in computing the gradient of $J_c(\delta\mathbf{z}_c)$ with respect to $\delta\mathbf{z}_{\zeta'}$, $\delta\mathbf{z}_{\psi''\chi''}$, and $\delta\mathbf{z}_{TS}$ for application of the quasi-Newton minimization algorithm. By differentiation their gradients may be written as

$$\boldsymbol{\nabla}_{\delta\mathbf{z}_{\zeta'}}J_\zeta^o = \mathbf{L}_{\zeta'}^{\mathrm{T}}\boldsymbol{\Sigma}_{\zeta'}\mathbf{H}_\zeta^{\mathrm{T}}\mathbf{R}_\zeta^{-1}(\mathbf{H}_\zeta\delta\mathbf{x}_\zeta^\dagger - \mathbf{d}_\zeta),$$

$$\boldsymbol{\nabla}_{\delta\mathbf{z}_{TS}}J_\zeta^o = \mathbf{W}_{TS}^{\mathrm{T}}\mathbf{L}_{TS}^{\mathrm{T}}\boldsymbol{\Sigma}_{TS}\boldsymbol{\Pi}_{TS}^{\mathrm{T}}\mathbf{H}_\zeta^{\mathrm{T}}\mathbf{R}_\zeta^{-1}(\mathbf{H}_\zeta\delta\mathbf{x}_\zeta^\dagger - \mathbf{d}_\zeta),$$

$$\boldsymbol{\nabla}_{\delta\mathbf{z}_{\zeta'}}J_{uv}^o = \mathbf{L}_{\zeta'}^{\mathrm{T}}\boldsymbol{\Sigma}_{\zeta'}\boldsymbol{\Gamma}_{\zeta'}^{\mathrm{T}}\mathbf{H}_{uv}^{\mathrm{T}}\mathbf{R}_{uv}^{-1}(\mathbf{H}_{uv}\delta\mathbf{x}_{uv}^\dagger - \mathbf{d}_{uv}),$$

$$\boldsymbol{\nabla}_{\delta\mathbf{z}_{TS}}J_{uv}^o = \mathbf{W}_{TS}^{\mathrm{T}}\mathbf{L}_{TS}^{\mathrm{T}}\boldsymbol{\Sigma}_{TS}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{H}_{uv}^{\mathrm{T}}\mathbf{R}_{uv}^{-1}(\mathbf{H}_{uv}\delta\mathbf{x}_{uv}^\dagger - \mathbf{d}_{uv}),$$

$$\boldsymbol{\nabla}_{\delta\mathbf{z}_{\psi''\chi''}}J_{uv}^o = \mathbf{W}_{\psi''\chi''}^{\mathrm{T}}\mathbf{L}_{\psi''\chi''}^{\mathrm{T}}\boldsymbol{\Sigma}_{\psi''\chi''}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{H}_{uv}^{\mathrm{T}}\mathbf{R}_{uv}^{-1}(\mathbf{H}_{uv}\delta\mathbf{x}_{uv}^\dagger - \mathbf{d}_{uv}),$$

$$\boldsymbol{\nabla}_{\delta\mathbf{z}_{TS}}J_{TS}^o = \mathbf{W}_{TS}^{\mathrm{T}}\mathbf{L}_{TS}^{\mathrm{T}}\boldsymbol{\Sigma}_{TS}\mathbf{H}_{TS}^{\mathrm{T}}\mathbf{R}_{TS}^{-1}(\mathbf{H}_{uv}\delta\mathbf{x}_{TS}^\dagger - \mathbf{d}_{TS}). \tag{35}$$

The difficulty in evaluating these gradients arises primarily from taking the transpose of six matrices, that is, $\boldsymbol{\Pi}_{TS}^{\mathrm{T}}$, $\boldsymbol{\Gamma}^{\mathrm{T}}$, $\boldsymbol{\Phi}^{\mathrm{T}}$, $\mathbf{H}_\zeta^{\mathrm{T}}$, $\mathbf{H}_{uv}^{\mathrm{T}}$, and $\mathbf{H}_{TS}^{\mathrm{T}}$. These are also called the adjoint of $\boldsymbol{\Pi}_{TS}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$, $\mathbf{H}_\zeta$, $\mathbf{H}_{uv}$, and $\mathbf{H}_{TS}$. While we have seen the complexity of $\boldsymbol{\Pi}_{TS}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Phi}$, the observational operators of $\mathbf{H}_\zeta$, $\mathbf{H}_{uv}$, and $\mathbf{H}_{TS}$ can also be complicated when nonconventional observations are assimilated. For example, $\mathbf{H}_{uv}$ becomes complicated when HF radar radial velocity measurements are directly assimilated. To handle such complexity we develop the adjoint code of those six matrices, with the coding method taken directly from the computer codes that are extensively used in oceanic and atmospheric data assimilation systems (e.g., Navon et al. 1992; Stammer et al. 2002).

6. Summary

A 3DVAR data assimilation system has been developed for operational coastal oceanography within the framework of ROMS. The theoretical formulation was described here. In recognition of the dynamic complexities of coastal oceans, we comprehensively and carefully incorporated dynamic balance constraints, constructed forecast error covariances, and preconditioned the cost function.

The dynamic constraints are incorporated to keep the analysis in the desired dynamic balance. In

ROMS3DVAR, both strong and weak dynamic constraints are formulated in the framework of the maximum-likelihood estimation (Jazwinski 1970; Cohn 1997). The formulation is different from Sasaki's (Sasaki 1970), but it follows the method of partitioning for the balanced and unbalanced parts (Parrish and Derber 1992). Applying the partition to the increments, rather than the total field, ROMS3DVAR maintains the internal dynamic consistency with the forecast $\mathbf{x}^f$, which can be subtle and delicate in a coastal ocean model.

The dynamic constraints are incorporated not only for keeping the analysis near the dynamic balance, but also for improving the cross correlations between the control variables. The cross correlations between velocity components and T/S are mainly the consequence of geostrophic and hydrostatic balance. In ROMS3DVAR the incremental steric SSH and geostrophic velocity components are determined by the incremental temperature and salinity. The control variables are then transformed into incremental nonsteric SSH, ageostrophic streamfunction and velocity potential, and temperature and salinity. In this way the cross correlations among nonsteric SSH, ageostrophic streamfunction and velocity potential, and temperature and salinity are neglected in ROMS3DVAR.

A Kronecker product formulation has been proposed to construct forecast error correlations. The most important capability of this formulation is its ability to represent 3D inhomogeneous and anisotropic correlations. The inhomogeneity and anisotropy in the forecast error covariance are significant and ubiquitous in coastal oceans. The significance and ubiquity is due to the control of the dynamics of coastal oceans by two physical boundaries: the sea surface and the coastline. The atmosphere drives the ocean at its surface; therefore, more dramatic changes in the ocean occur near the surface than in the deep water. The difference is dramatically manifested by the existence of the thermocline. Such a vertical difference in the dynamics necessarily leads to a vertical difference in forecast error correlations. As shown earlier, the horizontal correlations are influenced by the coastline. The structure of the forecast error correlation is thus a function of the distance from the coastline. The Kronecker product formulation has the capability to model such inhomogeneity and anisotropy.

The construction of the 3D correlations is particularly important for oceanic data assimilations. Many oceanic observations are limited to the surface. The success of a data assimilation system relies on the spread of the observed information from the surface to the deep water. The 3D correlation constructed with (23) dictates how information is spread vertically.

The Kronecker product formulation also offers computational advantages. One advantage is that it reduces the memory requirement for the error correlation. As such, the memory required for a typical coastal ocean model becomes manageable on present day supercomputers. The most important computational advantage of the Kronecker product formulation is that the Cholesky factorizations of the 3D correlation matrices can be implemented directly. In fact, ROMS3DVAR relies greatly on the Cholesky factorizations of the 3D correlation matrices. The preconditioning of the cost function requires Cholesky factorizations of the correlation matrices. The Cholesky factorization is also used to guarantee that the correlation matrices are positive definite. These applications of the Cholesky factorizations play a crucial role in the efficiency and reliability of ROMS3DVAR. Because of these computational advantages, this 3DVAR system lays a path to more advanced data assimilation schemes, such as a hybrid ensemble and 3DVAR scheme (e.g., Hamill and Snyder 2000; Lorenc 2003b; Buehner 2005; Wang et al. 2007).

We have so far only described the theoretical framework of ROMS3DVAR. Extensive assessment and evaluation of this data assimilation system are necessary and is the focus of Li et al. (2008), in which we describe the constructed forecast error variance and correlations, conduct a number of idealized experiments to further illustrate the algorithm, perform various experiments under various operational conditions to examine the validity of the assumptions used, and explore some operational issues. Chao et al. (2008, manuscript submitted to *Deep-Sea Res.*) summarizes the real-time performance of ROMS3DVAR when it was employed in generating real-time analyses during the Autonomous Ocean Sampling Network (AOSN) experiment in August 2003 (http://www.mbari.org/aosn/). The ROMS3DVAR system implemented off the central California coast has been run operationally in real time since August 2006. Similar ROMS3DVAR real-time forecasting systems are also currently being developed for the Southern California Bight and the Prince Williams Sound, Alaska.

APPENDIX A

Seawater Density, Geostrophic Velocity, and Sea Surface Level

a. Seawater density

In ocean models, the seawater density anomaly, which is the deviation from the Boussinesq approximation of the mean density, is computed via the equation of state (EOS). In ROMS the nonlinear EOS by Jackett and McDougall (1995) is used. To be consistent with ROMS, ROMS3DVAR employs the same EOS.

b. Sea surface level

The hydrostatic equation is written as

$$\frac{\partial p}{\partial z} = -g(\rho_0 + \rho), \tag{A1}$$

where ρ_0 is the Boussinesq approximation of the mean density, ρ the seawater density anomaly, g gravitational acceleration, and p seawater pressure at the depth h.

The vertical integration of (Al) from depth h to ζ yields

$$p = \int_{-h}^{\zeta} g(\rho_0 + \rho)\, dz, \tag{A2}$$

where ζ is the sea surface level. The incremental form of (A2) can be approximately written as

$$\delta p = \int_{-h}^{\zeta^f} g\delta\rho\, dz + g(\rho_0 + \rho_s^f)\delta\zeta, \tag{A3}$$

where ρ_s^f is the forecast surface density anomaly, diagnostically computed from the temperature and salinity forecast. Let the depth of the bottom be h_b, and the bottom pressure p_{h_b}. We can obtain from (A3)

$$\delta\zeta = -\frac{1}{\rho_0 + \rho_s^f} \int_{-h_b}^{\zeta^f} \delta\rho\, dz + \frac{\delta p_{h_b}}{\rho_0 + \rho_s^f}. \tag{A4}$$

When only considering the steric change in SSH, that is, the incremental density does not increase or decrease the total mass of the water column from the bottom to surface, $\delta p_{h_b} = 0$. As such, we call the first part of the right-hand side in (A4) the steric compo-

nent, while the second part is the nonsteric component. The nonsteric component is generally neglected in data assimilation for large-scale models (e.g., Alves et al. 2001).

Since $\delta\rho$ is computed by the linearized EOS (Jackett and McDougall 1995), it is determined using the temperature and salinity increment. Let $\delta\zeta' = \delta p_{h_b}/\rho_0 + \rho_s^f$, and then the discrete form of (A4) can be written as (12).

c. Geostrophic velocity

We write the geostrophic balance Eq. (11) in the form of the geostrophic velocity components

$$-fv = -\frac{1}{\rho_0 + \rho}\frac{\partial p}{\partial x}, \tag{A5}$$

$$fu = -\frac{1}{\rho_0 + \rho}\frac{\partial p}{\partial y}. \tag{A6}$$

In the incremental form we have the approximation of (A5) and (A6)

$$-f\delta v = -\frac{1}{\rho_0 + \rho^f}\frac{\partial \delta p}{\partial x}, \tag{A7}$$

$$f\delta u = -\frac{1}{\rho_0 + \rho^f}\frac{\partial \delta p}{\partial y}. \tag{A8}$$

Substituting (A3) into (A7) and (A8), we get

$$-f\delta v = -\frac{g}{\rho_0 + \rho^f}\frac{\partial}{\partial x}\int_{-h}^{\zeta^f} \delta\rho\, dz - \frac{g(\rho_0 + \rho_s^f)}{\rho_0 + \rho^f}\frac{\partial \delta\zeta}{\partial x}, \tag{A9}$$

$$f\delta u = -\frac{g}{\rho_0 + \rho^f}\frac{\partial}{\partial y}\int_{-h}^{\zeta^f} \delta\rho\, dz - \frac{g(\rho_0 + \rho_s^f)}{\rho_0 + \rho^f}\frac{\partial \delta\zeta}{\partial y}. \tag{A10}$$

Since the increment $\delta\rho$ is determined by function of the temperature and salinity increment, the discrete form of (A9) and (Al0) can be written as (13).

APPENDIX B

Properties of the Kronecker Product

a. Basics

The Kronecker product is a binary matrix operator that maps two arbitrarily dimensioned matrices into a larger matrix with special block structure. Given the $m \times n$ matrix **B** and the $p \times q$ matrix **C**,
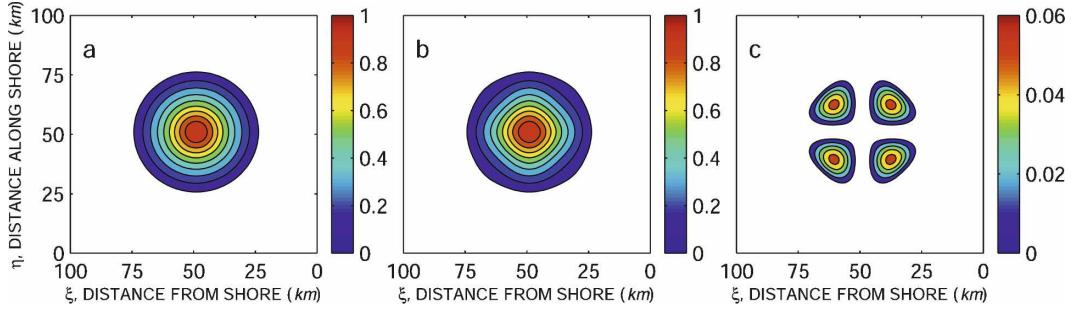
FIG. B1. 2D point correlation at the location (50, 50) based on the fifth-order polynomial proposed by Gaspari and Cohn [1999, their Eq. (4.11)]: (a) actual correlation, (b) its approximation using the Kronecker product–based algorithm, and (c) the difference between the actual correlation and its approximation.

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ & \dots & \dots & \\ & \dots & \dots & \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1q} \\ c_{21} & c_{22} & \dots & c_{2q} \\ & \dots & \dots & \\ & \dots & \dots & \\ c_{p1} & c_{p2} & \dots & c_{pq} \end{pmatrix}. \qquad (B1)$$

Their Kronecker product is denoted by $\mathbf{B} \otimes \mathbf{C}$ and is the $mp \times nq$ matrix with the block structure

$$\mathbf{B} \otimes \mathbf{C} = \begin{pmatrix} b_{11}\mathbf{C} & b_{12}\mathbf{C} & \dots & b_{1n}\mathbf{C} \\ b_{21}\mathbf{C} & b_{22}\mathbf{C} & \dots & b_{2n}\mathbf{C} \\ & \dots & \dots & \\ & \dots & \dots & \\ b_{m1}\mathbf{C} & b_{m2}\mathbf{C} & \dots & b_{mn}\mathbf{C} \end{pmatrix}. \qquad (B2)$$

For example, given the 2×2 matrix $\mathbf{B}$ and the 2×3 matrix $\mathbf{C}$,

$$\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \qquad (B3)$$

the Kronecker product $\mathbf{B} \otimes \mathbf{C}$ is the 4×6 matrix

$$\mathbf{B} \otimes \mathbf{C} = \begin{pmatrix} 1 & 2 & 3 & 2 & 4 & 6 \\ 4 & 5 & 6 & 8 & 10 & 12 \\ 0 & 0 & 0 & -1 & -2 & -3 \\ 0 & 0 & 0 & -4 & -5 & -6 \end{pmatrix}. \qquad (B4)$$

The Kronecker product operator has a number of properties (Graham 1981). In the following it is assumed that $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are real-valued matrices.

Some identities hold only for appropriately dimensioned matrices.

1) The Kronecker product is a bilinear operator. Given a constant α,

$$\mathbf{A} \otimes (\alpha\mathbf{B}) = \alpha(\mathbf{A} \otimes \mathbf{B}). \qquad (B5)$$

2) The Kronecker product is distributive over addition:

$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C}). \qquad (B6)$$

3) The Kronecker product is associative:

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}. \qquad (B7)$$

4) Transposition is distributive over the Kronecker product (does not invert the order):

$$(\mathbf{A} \otimes \mathbf{B})^{T} = \mathbf{A}^{T} \otimes \mathbf{B}^{T}. \qquad (B8)$$

5) Matrix multiplication, when dimensions are appropriate:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD}). \qquad (B9)$$

6) When $\mathbf{A}$ and $\mathbf{B}$ are square and full rank,

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}. \qquad (B10)$$

b. Inversion, Cholesky factorization, and positive definiteness

Error correlation matrices constructed by the Kronecker product can be inverted and factorized easily and efficiently. The cost function (17) includes the inverse of $\mathbf{B}_{\zeta'}$, $\mathbf{B}_{\psi''\chi''}$, and $\mathbf{B}_{TS}$. The minimization of (17) generally does not require an explicit computation of the inverse by preconditioning. Nevertheless, if necessary in some circumstances, the inverse can be computed,

$$(\mathbf{C}^{\xi\eta\kappa})^{-1} = (\mathbf{C}^{\xi\kappa})^{-1} \otimes (\mathbf{C}^{\eta})^{-1}. \qquad (B11)$$

However, Cholesky factorization is often required for preconditioning the cost function for minimization (sec-

tion 5). For the Cholesky factorization of $\mathbf{C}^{\xi\eta\kappa}$, we need only to perform the Cholesky factorization $\mathbf{C}^{\xi\kappa}$ and $\mathbf{C}^{\eta}$, since we have

$$\mathbf{C}^{\xi\eta\kappa} = \mathbf{G}^{\xi\kappa}(\mathbf{G}^{\xi\kappa})^{\mathrm{T}} \otimes \mathbf{G}^{\eta}(\mathbf{G}^{\eta})^{\mathrm{T}}$$

$$= (\mathbf{G}^{\xi\kappa} \otimes \mathbf{G}^{\eta})(\mathbf{G}^{\xi\kappa} \otimes \mathbf{G}^{\eta})^{\mathrm{T}}, \quad (\mathrm{B}12)$$

where

$$\mathbf{C}^{\xi\kappa} = \mathbf{G}^{\xi\kappa}(\mathbf{G}^{\xi\kappa})^{\mathrm{T}} \quad \text{and} \quad \mathbf{C}^{\eta} = \mathbf{G}^{\eta}(\mathbf{G}^{\eta})^{\mathrm{T}} \quad (\mathrm{B}13)$$

are the Cholesky factorizations $\mathbf{C}^{\xi\kappa}$ and $\mathbf{C}^{\eta}$. Efficiency is gained because the dimensions of $\mathbf{C}^{\xi\kappa}$ and $\mathbf{C}^{\eta}$ are much smaller than $\mathbf{C}^{\xi\eta\kappa}$.

Further, the 2D and 3D correlations are guaranteed to be positive definite if constructed using the Kronecker product. The positive definiteness can be demonstrated by

$$\mathbf{C}^{\eta} \otimes \mathbf{C}^{\xi\kappa} = (\mathbf{V}^{\eta} \otimes \mathbf{V}^{\xi})(\mathbf{\Sigma}^{\eta} \otimes \mathbf{\Sigma}^{\xi\kappa})(\mathbf{V}^{\eta} \otimes \mathbf{V}^{\xi\kappa})^{\mathrm{T}},$$

$$(\mathrm{B}14)$$

where

$$\mathbf{C}^{\eta} = \mathbf{V}^{\eta}\mathbf{\Sigma}^{\eta}(\mathbf{V}^{\eta})^{\mathrm{T}}, \quad (\mathrm{B}15)$$

$$\mathbf{C}^{\xi\kappa} = \mathbf{V}^{\xi\kappa}\mathbf{\Sigma}^{\xi\kappa}(\mathbf{V}^{\xi\kappa})^{\mathrm{T}}. \quad (\mathrm{B}16)$$

Also, $\mathbf{\Sigma}^{\eta}$ and $\mathbf{\Sigma}^{\xi\kappa}$ are diagonal matrices whose diagonal elements are the singular values; $\mathbf{V}^{\eta}$ and $\mathbf{V}^{\xi\kappa}$ are the matrices whose columns are the singular vectors; and $\mathbf{C}^{\xi\eta\kappa}$ is positive definite only if $\mathbf{C}^{\eta}$ and $\mathbf{C}^{\xi\kappa}$ are positive definite. The formulation (B14) can also be used to perform the singular vector decomposition efficiently.

c. Demonstration

The Kronecker product–based algorithm gives the exact isotropic and homogeneous 2D correlation if the actual 2D correlation, $\mathbf{C}^{\xi\eta}$, and two 1D correlations, $\mathbf{C}^{\xi}$ and $\mathbf{C}$, satisfy the relation (22), as is the case for Gaussian correlations. The algorithm can give a fairly accurate 2D correlation when the correlations are based on SOAR–TOAR or polynomial functions. As an illustration, Fig. B1 presents the comparison between the actual 2D correlation based on the polynomial function proposed by Gaspari and Cohn (1999) and its approximation based on the Kronecker product-based algorithm. The difference in the correlation is less than 5%, although the algorithm introduces some anisotropy.

One desirable characteristic of the Kronecker product formulation (21) is the ability to represent anisotropic and inhomogeneous correlations. As an illustrative example, we construct 2D anisotropic and inhomogeneous correlations using the 1D cross-shore and alongshore correlations with the Gaussian function. Using
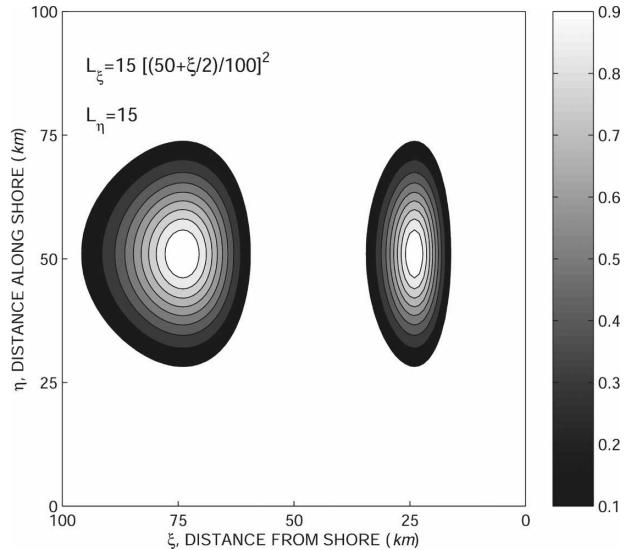

FIG. B2. 2D point correlation constructed using the Kronecker product–based algorithm and the 1D Gaussian correlations for (left) an offshore location at (70, 50) and (right) a nearshore location (30, 50). The cross-shore length scale L^{ξ} increases as a function of the distance from the coastline (the eastern boundary), while the alongshore length scale $L^{\eta} = 15$ km is constant.

the Gaussian function, the 1D cross-shore correlation can be constructed by

$$C^{\xi}(\xi_1, \xi_2) = \exp\left[-\frac{(\xi_2 - \xi_1)^2}{2(L^{\xi 1})^2} \right], \quad \xi_2 \geq \xi_1,$$

$$C^{\xi}(\xi_1, \xi_2) = C^{\xi}(\xi_2, \xi_1), \quad \xi_2 < \xi_1, \quad (\mathrm{B}17)$$

and the 1D alongshore correlation constructed by

$$C^{\eta}(\eta_1, \eta_2) = \exp\left[-\frac{(\eta_2 - \eta_1)^2}{2(L^{\eta 1})^2} \right], \quad \eta_2 \geq \eta_1,$$

$$C^{\eta}(\eta_1, \eta_2) = C^{\eta}(\eta_2, \eta_1), \quad \eta_2 < \eta, \quad (\mathrm{B}18)$$

where L^{ξ} and L^{η} are the cross-shore and alongshore length scales, respectively, and $\xi = 0$ corresponds to the eastern coastline. We let $L^{\xi} = 15(50 + \xi/2)^2$, while keeping $L^{\eta} = 15$ km fixed. As such, L^{ξ} increases with the distance from the shore. The constructed correlation shows the anisotropic and inhomogeneous features as in Fig. B2. In Fig. B2 we can see some similarities to the correlation computed from the data shown in Fig. 1.

REFERENCES

Alves, J. O. S., K. Haines, and D. L. T. Anderson, 2001: Sea level assimilation experiments in the tropical Pacific. *J. Phys. Oceanogr.*, **31**, 305–323.

Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N.

Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.,* **132,** 897–914.

Bennett, A. F., 1997: *Inverse Methods in Physical Oceanography.* Cambridge University Press, 346 pp.

Boer, G. J., 1983: Homogeneous and isotropic turbulence on the sphere. *J. Atmos. Sci.,* **40,** 154–163.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.,* **131,** 1013–1043.

Cohn, S. E., 1997: Estimation theory for data assimilation problems: Basic conceptual framework and some open questions. *J. Meteor. Soc. Japan,* **75,** 257–288.

——, D. M. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO Physical-Space Statistical Analysis System. *Mon. Wea. Rev.,* **126,** 2913–2926.

Courtier, P., J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120,** 1367–1387.

——, and Coauthors, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.,* **124,** 1783–1807.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

——, and E. Barker, 2001: NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.,* **129,** 869–883.

Dee, D. P., 1995: Online estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123,** 1128–1145.

DiGiacomo, P. M., and B. Holt, 2001: Satellite observations of small coastal ocean eddies in the Southern California Bight. *J. Geophys. Res.,* **106,** 22 521–22 543.

Di Lorenzo, E., A. M. Moore, H. G. Arango, B. D. Cornuelle, A. J. Miller, B. Powell, B. S. Chua, and A. F. Bennett, 2007: Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): Development and application for a baroclinic coastal upwelling system. *Ocean Modell.,* **16,** 160–187.

Fisher, M., M. Leutbecher, and G. A. Kelly, 2006: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.,* **131,** 3235–3246.

Fukumori, I., R. Raghunath, R. L. Fu, and Y. Chao, 1999: Assimilation of TOPEX/Poseidon altimeter data into a global ocean circulation model: How good are the results? *J. Geophys. Res.,* **104,** 25 647–25 665.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723–757.

Graham, A., 1981: *Kronecker Product and Matrix Calculus with Applications.* Ellis Horwood Ltd., 160 pp.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905–2919.

Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast error as determined from radiosonde data. Part I: The wind field. *Tellus,* **38A,** 111–136.

Ide, K., P. Courtier, M. Ghil, and A. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan,* **75,** 181–189.

Jackett, D. R., and T. J. McDougall, 1995: Minimal adjustment of hydrostatic profiles to achieve static stability. *J. Atmos. Oceanic Technol.,* **12,** 381–389.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 376 pp.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 341 pp.

Lermusiaux, P. F. J., and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.,* **127,** 1385–1407.

Li, Z., and I. M. Navon, 2001: Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Quart. J. Roy. Meteor. Soc.,* **127,** 661–683.

——, Y. Chao, and J. C. McWilliams, 2006: Computation of the streamfunction and velocity potential for limited and irregular domains. *Mon. Wea. Rev.,* **134,** 3384–3394.

——, ——, ——, and K. Ide, 2008: A three-dimensional variational data assimilation scheme for the Regional Ocean Modeling System: Implementation and basic experiments. *J. Geophys. Res.,* **113,** C05002, doi:10.1029/2006JC004042.

Liu, D. C., and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Math. Program.,* **45,** 503–528.

Lönnberg, P., and A. Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus,* **38A,** 137–161.

Lorenc, A. C., 2003a: Modelling of error covariances by 4D-Var data assimilation. *Quart. J. Roy. Meteor. Soc.,* **129,** 3167–3182.

——, 2003b: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.,* **129,** 3183–3203.

——, and Coauthors, 2000: The Met Office global 3-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **26,** 2991–3012.

Marchesiello, P., J. C. McWilliams, and A. F. Shchepetkin, 2001: Open boundary conditions for long-term integration of regional ocean models. *Ocean Modell.,* **3,** 1–20.

——, ——, and A. Shchepetkin, 2003: Equilibrium structure and dynamics of the California Current System. *J. Phys. Oceanogr.,* **33,** 753–783.

Moore, A. M., H. G. Arango, E. DiLorenzo, B. D. Cornuelle, A. J. Miller, and D. J. Neilson, 2004: A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. *Ocean Modell.,* **7,** 227–258.

National Research Council, 2003: *Enabling Ocean Research in the 21st Century: Implementation of a Network of Ocean Observatories.* The National Academies Press, 220 pp.

Navon, I. M., X. Zou, J. Derber, and J. Sela, 1992: Variational data assimilation with an adiabatic version of NMC spectral model. *Mon. Wea. Rev.,* **120,** 1433–1446.

Nocedal, J., and S. J. Wright, 1999: *Numerical Optimization.* Springer-Verlag, 636 pp.

Paduan, J. D., and H. C. Graber, 1997: Introduction to high-frequency radar: Reality and myth. *Oceanography,* **10,** 36–39.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral-interpolation analysis system. *Mon. Wea. Rev.,* **120,** 1747–1763.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomoge-

neous and anisotropic general covariances. *Mon. Wea. Rev.,* **131,** 1536–1548.

Rabier, F., H. Jarvinnen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 1998a: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.,* **126,** 1143–1170.

——, A. McNally, E. Andersson, P. Courtier, P. Unden, J. Eyre, A. Hollingsworth, and F. Bouttier, 1998b: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Quart. J. Roy. Meteor. Soc.,* **124,** 1809–1829.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.,* **98,** 875–883.

Shchepetkin, A. F., and J. C. McWilliams, 2003: A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate. *J. Geophys. Res.,* **108,** 3090, doi:10.1029/2001JC001047.

——, and ——, 2005: The Regional Ocean Modeling System (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modell.,* **9,** 347–404.

Stammer, D., and Coauthors, 2002: Global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model. *J. Geophys. Res.,* **107,** 3118, doi:10.1029/2001JC000888.

Tikhonov, A. N., and V. Arsenin, 1977: *Solutions of Ill-Posed Problems.* Halsted Press, 258 pp.

Troccoli, A., and K. Haines, 1999: Use of the temperature–salinity relation in a data assimilation context. *J. Atmos. Oceanic Technol.,* **16,** 2011–2025.

Vossepoel, F. C., R. W. Reynolds, and L. Miller, 1999: Use of sea level observations to estimate salinity variability in the tropical Pacific. *J. Atmos. Oceanic Technol.,* **16,** 1401–1415.

Wang, X., T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007: A comparison of hybrid ensemble transform Kalman filter–optimum interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.,* **135,** 1055–1076.

Wunsch, C., 1996: *The Ocean Circulation Inverse Problem.* Cambridge University Press, 458 pp.

Xu, Q., and W. Li, 2001: Estimation of three-dimensional error covariances. Part II: Analysis of wind innovation vectors. *Mon. Wea. Rev.,* **129,** 2939–2954.