

The M_2 Internal Tide off Oregon: Inferences from Data Assimilation

ALEXANDER L. KURAPOV, GARY D. EGBERT, J. S. ALLEN, ROBERT N. MILLER,
SVETLANA Y. EROFEEVA, AND P. M. KOSRO

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

(Manuscript received 8 May 2002, in final form 21 January 2003)

ABSTRACT

A linearized baroclinic, spectral-in-time tidal inverse model has been developed for assimilation of surface currents from coast-based high-frequency (HF) radars. Representer functions obtained as a part of the generalized inverse solution show that for superinertial flows information from the surface velocity measurements propagates to depth along wave characteristics, allowing internal tidal flows to be mapped throughout the water column. Application of the inverse model to a $38 \text{ km} \times 57 \text{ km}$ domain off the mid-Oregon coast, where data from two HF radar systems are available, provides a uniquely detailed picture of spatial and temporal variability of the M_2 internal tide in a coastal environment. Most baroclinic signal contained in the data comes from outside the computational domain, and so data assimilation (DA) is used to restore baroclinic currents at the open boundary (OB). Experiments with synthetic data demonstrate that the choice of the error covariance for the OB condition affects model performance. A covariance consistent with assumed dynamics is obtained by nesting, using representers computed in a larger domain. Harmonic analysis of currents from HF radars and an acoustic Doppler profiler (ADP) mooring off Oregon for May–July 1998 reveals substantial intermittence of the internal tide, both in amplitude and phase. Assimilation of the surface current measurements captures the temporal variability and improves the ADP/solution rms difference. Despite significant temporal variability, persistent features are found for the studied period; for instance, the dominant direction of baroclinic wave phase and energy propagation is always from the northwest. At the surface, baroclinic surface tidal currents (deviations from the depth-averaged current) can be 10 cm s^{-1} , 2 times as large as the depth-averaged current. Barotropic-to-baroclinic energy conversion is generally weak within the model domain over the shelf but reaches 5 mW m^{-2} at times over the slopes of Stonewall Bank.

1. Introduction

Internal tides are generated over sloping topography, where the vertical component of the barotropic tidal current forces oscillations in the density field (Wunsch 1975; Baines 1982). The M_2 tide off Oregon is superinertial such that internal wave energy propagates away from generation sites along wave characteristic surfaces, or beams. The slope of the wave characteristics can be estimated as

$$\tan\varphi = (\omega^2 - f^2)^{1/2}(N^2 - \omega^2)^{-1/2}, \quad (1)$$

where φ is the angle that the characteristics make with the horizontal, ω is the tidal frequency, f is the inertial frequency, and N is the buoyancy frequency. Internal waves incident upon supercritical bathymetry (where the bottom slope is steeper than the wave characteristics) reflect toward deeper water, while those incident upon subcritical bathymetry will propagate into shallower water. Where the bathymetric slope is close to $\tan\varphi$, strong

bottom intensification of the baroclinic currents occurs. The most plausible site for generating energetic internal tides in the coastal ocean is the continental shelf break, where over the distance of several kilometers bathymetry changes from supercritical (on the continental slope) to subcritical (on the shelf) (Figs. 1, 2).

The Oregon coastal shelf is relatively narrow (30–50 km wide) such that the internal tide propagating onshore remains energetic up to the coast, enhancing spatial and temporal variability in currents, mixing, and biological productivity. The total M_2 tidal current in this region may reach 15 cm s^{-1} or more, while maximum barotropic (depth-averaged) currents are typically about 5 cm s^{-1} or less (Hayes and Halpern 1976; Torgrimson and Hickey 1979; Erofeeva et al. 2003). In coastal waters, internal tides show spatial variability on the scale of several kilometers and temporal variability on the scale of days. The first-mode internal M_2 horizontal length scale, estimated as $2NH/\omega$, is 14 km in water of depth $H = 100 \text{ m}$, given a typical value of $N = 0.01 \text{ s}^{-1}$. Topographic variation may introduce additional length scales. One reason for high temporal variability may be the spring–neap cycle (Petruncio et al. 1998). Other causes of variation may be changes in the strat-

Corresponding author address: Dr. Alexander L. Kurapov, College of Oceanic and Atmospheric Sciences, Oregon State University, 104 Ocean Administration Bldg., Corvallis, OR 97331-5503.
E-mail: kurapov@coas.oregonstate.edu

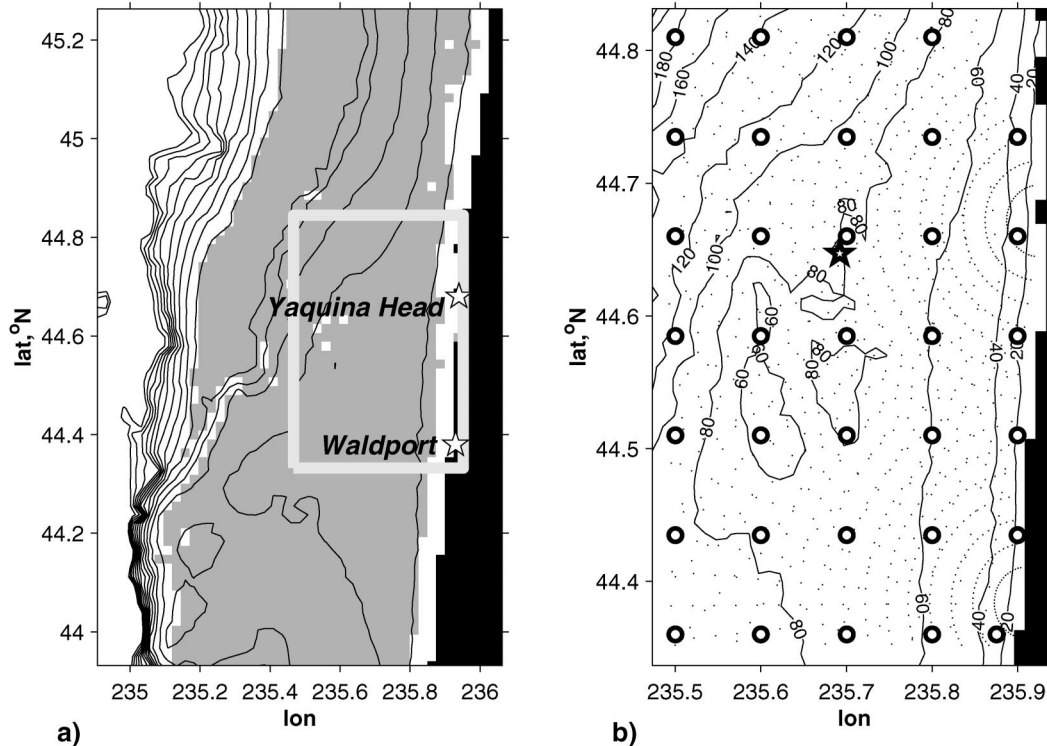


FIG. 1. Maps: (a) central part of the Oregon shelf. Contours are bathymetry every 50 m, but depths larger than 750 m are not contoured; shaded (white) areas are the regions where the slope is subcritical (supercritical) with respect to the M_2 internal tidal wave characteristics, corresponding to the horizontally uniform stratification shown in Fig. 2; stars denote locations of the two HF radars. (b) Computational domain for the represent analysis and inversion [its boundary is shown in (a) as the bold gray line]; contours are bathymetry, every 20 m; dots are the locations where HF surface current data are obtained; circles are the reduced basis reductor locations (see section 3c); the star is the ADP mooring location.

ification or in the low-frequency background currents, which may result from surface heat flux or wind forcing on the shelf. During upwelling the lifting of isopycnal surfaces near the coast affects the angle of wave char-

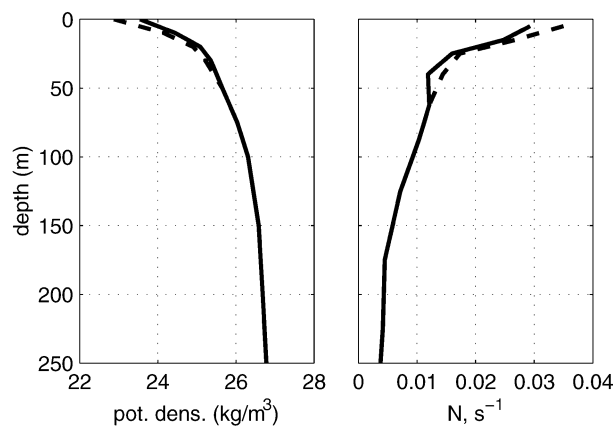


FIG. 2. Potential density $\bar{\rho}$ and corresponding buoyancy frequency N used in computations: solid line is the mean of summer observations, 1961–71, at a station 46 km offshore of Yaquina Head (Smith et al. 2000); the dashed line is a variant used to test solution sensitivity to stratification (see section 7a).

acteristics (Torgrimson and Hickey 1979). Intensive wind-induced vertical mixing near the coast may weaken the baroclinic tidal current (Hayes and Halpern 1976). As a result, there are significant seasonal variations, with much weaker internal tides in winter (Erofeeva et al. 2003).

Modeling internal tide variability on the shelf is difficult because of lack of information about surface heat flux, wind forcing, and stratification on the scales of interest. The requirement for fine spatial resolution restricts the size of the computational domain, and the need for open boundary flows contributes further to uncertainty. Observations of shelf currents may provide additional information about internal tides that can be used to constrain the model by means of data assimilation (DA). Since 1997, measurements of the surface currents on the mid-Oregon shelf have been available from two land-based high-frequency (HF) radars, over an area about $50 \text{ km} \times 50 \text{ km}$, with resolution 1–2 km in space, and 20–60 min in time (Kosro et al. 1997). In this paper, the focus is on semidiurnal tidal variations during May–July 1998, when the HF surface data were complemented by the currents from an acoustic Doppler profiler (ADP) mooring, providing information about

the flow variability with depth. Oke et al. (2002) used this dataset to show the value of assimilating HF radar measurements into a model of wind-forced coastal circulation. For that study the data were low-pass filtered to eliminate tidal and inertial oscillations, and the focus was on subinertial time scales.

Data assimilation is a way to estimate the state of the ocean that best fits both the dynamics and the data (Bennett 1992). Since the data provide extra information for the model, they may be utilized to correct the model inputs in some way: to correct open boundary values, estimate model parameters, and/or recover errors in the dynamical equations. Optimal DA schemes require estimates of the model solution error covariance (Kurapov et al. 2002). Ideally a DA system for the coastal ocean would model both tidal and wind-driven circulation and be based on implementation of the fully nonlinear dynamical equations. However, before such a complicated system can be built, it is appropriate to focus on some fundamental questions. Do measurements of surface currents contain information about internal tides at depth? If so, how can this surface information be projected downward in an optimal way? What are the principal features of the model solution error covariance in the superinertial band? These questions are addressed here with a model based on simplified dynamics for which a rigorous, variational, generalized inverse DA method (GIM) is readily implemented (Bennett 1992, 2002). The DA model presented below is based on dynamics that describe linear harmonic oscillations of the stratified ocean with respect to a state of rest, on realistic bottom topography. Both generation and propagation of internal waves are essentially linear mechanisms, so application of this simple model to realistic data is promising. Our model does not include advection by mean currents and accompanying spatial variability in the density field, which are limitations: wind-induced low-frequency baroclinic currents along the coast may be as high as 0.5 m s^{-1} , and their effect on the internal tide cannot be addressed with our model.

The main goals of this study are to 1) learn about the information content and utility of HF radar surface data for modeling the superinertial internal tide, 2) demonstrate the importance of a dynamically consistent inverse problem formulation at the open boundary, and 3) describe the spatial and temporal variability of the internal tide on the Oregon shelf at the dominant M_2 tidal frequency.

Section 2 provides an analysis of the ADP and HF radar data. The inverse model is described in section 3. In sections 4 and 5, the sensitivity of the inverse solution to the open-boundary error covariance is studied in a series of experiments with synthetic data. Section 6 discusses the effect of assumed dynamics on representer functions that show the zones of influence of the data and give the error covariance of the model solution. Then, in section 7, realistic surface currents are assimilated and the series of inverse solutions for May–July

1998 is validated against the ADP, followed by analysis of the internal tide on the Oregon shelf. Section 8 contains a summary.

2. Data analysis

For the period of May–July 1998, time series of surface currents on the mid-Oregon shelf are available from two HF radars (SeaSonde instrument systems, manufactured by CODAR Ocean Sensors), installed at Yaquina Head and Waldport (Fig. 1), as well as currents from an ADP mooring installed 25 km off Yaquina Head. The ADP, anchored at a depth of 80 m, provides time series of velocity measurements in the water column at depths 12–68 m every 4 m. The ADP data were harmonically analyzed in a number of overlapping 2-week time windows such that variations associated with the spring–neap tidal cycle are largely filtered out. Complex harmonic constants for eastward (u) and northward (v) components of velocity are computed in windows centered on days 131–191, spaced 4 days apart. In the following these time-varying estimates are referred to by the center day in the window. Substantial intermittence of the internal tide is seen in the tidal ellipses of the ADP currents (Fig. 3a). The first baroclinic mode remains dominant, but with amplitudes and phases varying with time. During days 131–155 the estimated ADP M_2 baroclinic velocities are relatively low, with maximum currents about 5 cm s^{-1} both near the bottom and the surface. Starting around day 159 the internal tide at the ADP location becomes much larger, reaching 9 cm s^{-1} near the surface and 7 cm s^{-1} near the bottom. During days 159–187, the tide becomes surface intensified, as near-bottom currents gradually decrease to 4 cm s^{-1} . The barotropic (depth averaged) current, with amplitude about 4 cm s^{-1} , shows significantly lower variability than the deviations from the depth average (Erofeeva et al. 2003). Some variability in these estimates of barotropic tidal currents may be due to data noise, the fact that the depth-averaged current is not exactly the barotropic current (separation of the flow into barotropic and baroclinic components on the sloping bathymetry is not obvious), and remnants of the spring–neap cycle not entirely filtered as a result of data processing in short time windows.

A land-based HF radar infers the surface current from measurements of the radar backscatter from ocean waves, considered with other information about the waves (Kosro et al. 1997). Measured data were processed to radial vectors every 10 min, which were then averaged over 1 h, providing maps of the radial component of the current with resolution 2 km radially and 5° azimuthally. The M_2 harmonic tidal constants for the radial components of surface velocity from the HF radars were computed in a manner similar to that used for the ADP data. To reduce error, harmonic constants are estimated by fitting measurements at four adjacent locations of the data array for each HF radar system.

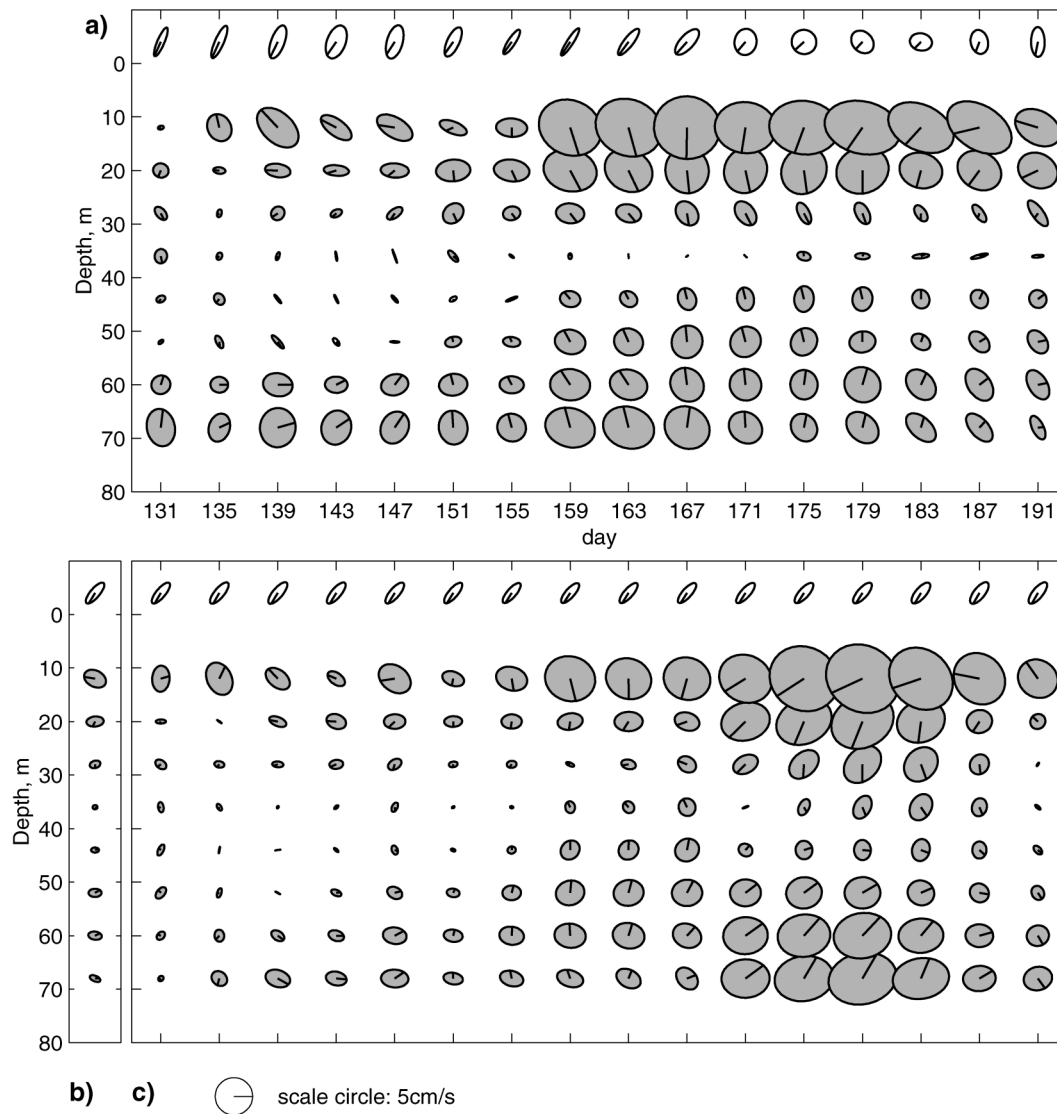


FIG. 3. Horizontal M_2 tidal current ellipses at the ADP data locations, decomposed into the depth average (shown above the zero depth) and deviations over the water column: (a) harmonically analyzed ADP data, (b) prior solution, and (c) inverse solution obtained by assimilating HF radar radials west of longitude 235.8° . Here and in Figs. 10 and 13 the filled ellipses correspond to CW vector rotation; the line from the ellipse center shows the velocity direction at zero phase. In the plot, northward velocity is directed up, and eastward velocity to the right.

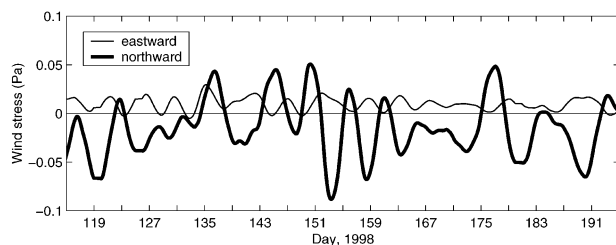


FIG. 4. Low-pass filtered wind stress, from wind measurements at Newport, Oregon (note that wind stress is shown here only for reference, no wind forcing is applied to the model).

Only harmonic constants with estimated mean least square (MLS) error $< 5 \text{ cm s}^{-1}$ are assimilated. In the domain shown in Fig. 1b, the mean and median of MLS errors vary insignificantly with time, remaining close to 2 cm s^{-1} .

During the study period the winds are mostly upwelling favorable (Fig. 4). This is expected to cause vertical mixing in a surface boundary layer and lifting of isopycnal surfaces closer to the coast. However, we lack specific empirical information about the density structure for this period. In the computations of the internal tide the ambient potential density $\bar{\rho}(z)$ is assumed to be horizontally uniform and temporally constant (Fig. 2). This approximation may be a serious source of error

in our model, especially later in the study period since during upwelling the pycnocline structure is gradually built up, and during short relaxation events (no wind) the pycnocline does not return to horizontal. However, we adopt this approximation since the inverse model is run in a small domain, where the misspecified open boundary baroclinic flow is hypothesized to be a larger source of error.

3. Inverse model

a. Model equations

The model describes linear harmonic oscillations of the stratified ocean with respect to the state of rest. Boussinesq and hydrostatic approximations are made. State variables vary with time as $\exp(i\omega t)$. The model is written in horizontal Cartesian coordinates with x to the east, y to the north. Linearized surface boundary conditions are applied at the mean sea surface level ($z = 0$). The vertical coordinate is stretched into $\sigma = z/H(x, y)$, and so the model equations are (cf. Blumberg and Mellor 1987; Mellor 1998):

$$i\omega H\mathbf{u} + \mathbf{f} \times H\mathbf{u} + gH\nabla\eta + \frac{gH^2}{\rho_o} \int_{\sigma}^0 \left(\nabla\rho - \sigma' \frac{\nabla H}{H} \frac{\partial\rho}{\partial\sigma'} \right) d\sigma' - \frac{\partial}{\partial\sigma} \left(\frac{K_M}{H} \frac{\partial\mathbf{u}}{\partial\sigma} \right) = 0, \quad (2)$$

$$\frac{\partial w}{\partial\sigma} + \nabla \cdot (H\mathbf{u}) = 0, \quad (3)$$

$$i\omega H\rho + \frac{\partial}{\partial\sigma}(\bar{\rho}w) + \nabla \cdot (\bar{\rho}H\mathbf{u}) = 0. \quad (4)$$

In (2)–(4), all the dynamic variables represent complex amplitudes: $\mathbf{u}(x, y, \sigma) = \{u, v\}$ is the horizontal velocity vector, $\eta(x, y)$ is the surface elevation, $\rho(x, y, \sigma)$ is the potential density perturbation about a horizontally uniform mean state $\bar{\rho}(z) = \bar{\rho}(x, y, \sigma)$, and $w(x, y, \sigma)$ is the velocity component normal to the σ -surfaces ($w = w^{\text{cart}} - \sigma\mathbf{u} \cdot \nabla H$, where w^{cart} is the Cartesian vertical velocity component in the unstretched coordinates); the parameters to be specified are gravity g , the Coriolis parameter $f = \text{const}$, the vertical eddy viscosity $K_M(x, y, \sigma)$, bathymetry $H(x, y)$ and the reference density $\rho_o = \text{const}$; ∇ and $\nabla \cdot$ are the “horizontal” gradient and divergence operators on a σ surface.

Vertical dissipation in (2) is the only sink of energy in the model and it is a very simple and admittedly crude parameterization of turbulence associated with background flows. A horizontal diffusion term is not included in (2), partly for a technical reason, to allow the solution of the momentum equations in terms of velocity locally for each vertical column (see appendix A). In fact, for coastal applications, horizontal dissipation

is a minor contributor to the momentum balance. Since we do not solve the problem by time stepping, horizontal dissipation is not needed for computational stability. Vertical diffusion of ρ is not included in (4). This simplifies the energy balance and is consistent with the assumption that vertical diffusion does not have effect on the background $\bar{\rho}(z)$.

Surface and bottom boundary conditions, at $\sigma = 0$ and $\sigma = -1$, are

$$\frac{\partial\mathbf{u}}{\partial\sigma} = 0 \quad \text{at } \sigma = 0, \quad (5)$$

$$w = i\omega\eta \quad \text{at } \sigma = 0, \quad (6)$$

$$w = 0 \quad \text{at } \sigma = -1, \quad (7)$$

$$\frac{K_M}{H} \frac{\partial\mathbf{u}}{\partial\sigma} = r\mathbf{u} \quad \text{at } \sigma = -1, \quad (8)$$

where $r(x, y)$ is the bottom drag coefficient.

A depth-integrated continuity equation is obtained with use of (3), (6), and (7):

$$i\omega\eta + \nabla \cdot \int_{-1}^0 H\mathbf{u} d\sigma = 0. \quad (9)$$

At the side boundaries Γ the normal current $\mathbf{u} \cdot \mathbf{n}$ is specified pointwise:

$$\mathbf{u} \cdot \mathbf{n}|_{\Gamma} = \begin{cases} 0 & \text{at the coast} \\ u_{\text{OB}} & \text{at the open boundary.} \end{cases} \quad (10)$$

Since horizontal advective terms are dropped from the equations, the problem with boundary condition (10) is formally well posed, in the sense that it has a unique solution which is stable to changes in inputs (Oliger and Sundström 1978; see also Bennett 1992, chapter 9). The open boundary condition (10) is in general reflective. However, we are going to treat the boundary condition as imperfect and correct open boundary baroclinic fluxes. This can be done in such a way to improve wave radiation out of the domain.

Once the dynamics are defined, the GIM requires specification of a cost functional that is a weighted sum of terms, each penalizing errors in the inputs such as the model equations, boundary conditions, or data. This can be minimized in an efficient way if the problem is separated into a series of so called adjoint and forward problems (see Bennett 1992, 2002; Egbert et al. 1994). Technically, there are two approaches to formulating the generalized inverse problem that eventually differ in how the adjoint model is built (Sirkes and Tziperman 1997). In the first approach, the cost functional penalizes the errors in the continuous equations and boundary conditions (2)–(10), and the finite-difference implementation is formulated independently for the forward and adjoint problems derived in continuous form. In the second approach, to be taken here, the cost functional penalizes errors in the discretized dynamical equations

and boundary conditions. The advantage of this approach in our case is that the discrete adjoint solver will be obtained as a matrix transpose of the forward solver, simplifying derivation, as well as computer coding, for representer calculation.

b. Model discretization

The equations and boundary conditions are discretized on a staggered Arakawa C grid, with details as in Blumberg and Mellor (1987), and Mellor (1998). Then, state variables at the points of the discrete grid are combined into vectors of unknowns \mathbf{u} , \mathbf{w} , $\boldsymbol{\eta}$, and $\boldsymbol{\rho}$, corresponding to continuous $H\mathbf{u}$, w , η , and ρ . Discretized model equations can be written in matrix notation as

$$\boldsymbol{\Omega}\mathbf{u} + \mathbf{G}\boldsymbol{\eta} + \mathbf{P}\boldsymbol{\rho} = \mathbf{f}_u, \quad (11)$$

$$i\omega\boldsymbol{\eta} + \mathbf{D}_a\mathbf{u} = \mathbf{f}_a, \quad (12)$$

$$\mathbf{W}\mathbf{w} + \mathbf{D}\mathbf{u} = \mathbf{f}_w, \quad (13)$$

$$\mathbf{B}\boldsymbol{\rho} + \mathbf{Z}_p\mathbf{w} + \mathbf{D}_p\mathbf{u} = \mathbf{f}_p, \quad (14)$$

where (11) represents horizontal momentum equations (2) combined with boundary conditions (5), (8), and (10); (12) represents the depth-integrated continuity equation (9); (13) is for the continuity equation (3) combined with the bottom boundary condition (7); and (14) is the discrete representation of the equation for the potential density (4). In (11)–(14), $\boldsymbol{\Omega}$, \mathbf{G} , and so on, are the coefficient matrices resulting from discretization, and \mathbf{f}_u , \mathbf{f}_a , etc., are forcing vectors. If the state variables and the forcing vectors are combined into

$$\boldsymbol{\nu} = \begin{pmatrix} \mathbf{u} \\ \mathbf{w} \\ \boldsymbol{\eta} \\ \boldsymbol{\rho} \end{pmatrix} \quad \text{and} \quad \mathbf{f} = \begin{pmatrix} \mathbf{f}_u \\ \mathbf{f}_w \\ \mathbf{f}_a \\ \mathbf{f}_p \end{pmatrix}, \quad (15)$$

then Eqs. (11)–(14) can be formally written together as

$$\mathbf{S}\boldsymbol{\nu} = \mathbf{f}, \quad (16)$$

where \mathbf{S} is the model operator (matrix). In general, the forcing vector \mathbf{f} can be written as $\mathbf{f} = \mathbf{f}_0 + \boldsymbol{\epsilon}$, where \mathbf{f}_0 is the vector of boundary values and $\boldsymbol{\epsilon}$ represents errors in the discretized equations and boundary conditions. Although in our study most of the elements of $\boldsymbol{\epsilon}$ are always zero, we retain the general weak constraint formulation for future convenience (see section 3c). An efficient approach that we utilize to solve (16) for an arbitrary \mathbf{f} is based on the direct factorization of the model operator (Egbert and Erofeeva 2002), as described in appendix A.

c. Cost function and inverse solution

Data to be assimilated can be written in the general form

$$d_k = \mathbf{I}'_k\boldsymbol{\nu} + \boldsymbol{\epsilon}_{dk}, \quad (17)$$

where the vector \mathbf{I}_k is the discrete data functional, the prime denotes the complex conjugate matrix transpose, d_k is the k th scalar observation value, $\boldsymbol{\epsilon}_{dk}$ is the data error, and $k = 1, \dots, K$. All the data can be combined in one matrix equation:

$$\mathbf{d} = \mathbf{L}'\boldsymbol{\nu} + \boldsymbol{\epsilon}_d. \quad (18)$$

The inverse solution minimizes the cost function:

$$J(\boldsymbol{\nu}) = (\mathbf{S}\boldsymbol{\nu} - \mathbf{f}_0)' \mathbf{cov}^{-1}(\mathbf{S}\boldsymbol{\nu} - \mathbf{f}_0) + (\mathbf{d} - \mathbf{L}'\boldsymbol{\nu})' \mathbf{cov}_d^{-1}(\mathbf{d} - \mathbf{L}'\boldsymbol{\nu}), \quad (19)$$

where \mathbf{cov} and \mathbf{cov}_d are the Hermitian non-negative definite covariance matrices of errors in the model equations and the data, respectively. These have to be specified prior to inversion. In the statistical interpretation of the GIM, the errors are treated as random vectors, so $\mathbf{cov} = \langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}' \rangle$ and $\mathbf{cov}_d = \langle \boldsymbol{\epsilon}_d\boldsymbol{\epsilon}'_d \rangle$, where angle brackets denote ensemble average. It is also assumed that dynamics and measurements are unbiased: $\langle \boldsymbol{\epsilon} \rangle = 0$, and $\langle \boldsymbol{\epsilon}_d \rangle = 0$.

In our case, the first term in (19) incorporates penalties on both the dynamical equations and boundary conditions. If the errors in the dynamical equations and boundary conditions are assumed to be uncorrelated, this term can be written as the sum of penalties on the errors in these two sources. Note that, in general, dynamical and open boundary condition errors may be correlated (Bogden 2001). The inverse solution can be expressed as

$$\boldsymbol{\nu}_{\text{inv}} = \boldsymbol{\nu}_0 + \sum_{k=1}^K b_k \mathbf{r}_k, \quad (20)$$

where $\boldsymbol{\nu}_0 = \mathbf{S}^{-1}\mathbf{f}_0$ is the prior solution satisfying error-free equations (11)–(14),

$$\mathbf{r}_k = \mathbf{S}^{-1}\mathbf{cov}(\mathbf{S}')^{-1}\mathbf{I}_k \quad (21)$$

are representer vectors, coefficients b_k in (20) are elements of vector \mathbf{b} , of size K :

$$\mathbf{b} = (\mathbf{R} + \mathbf{cov}_d)^{-1}(\mathbf{d} - \mathbf{L}'\boldsymbol{\nu}_0), \quad (22)$$

and \mathbf{R} is the representer matrix with elements $\mathbf{R}_{kj} = \mathbf{I}'_k\mathbf{r}_j$.

The representer shows the zone of influence of each observation in the model domain. It can be partitioned into fields each providing correction to prior u , v , w , η , and ρ . In (21), $(\mathbf{S}')^{-1}\mathbf{I}_k$ is the adjoint representer solution, which gives the sensitivity of the k th observation to changes in forcing and boundary conditions. It is multiplied by the covariance, which usually has a smoothing effect (see Egbert et al. 1994). Last, the “forward” problem (21) is solved. Our solution method for (16) allows rapid computation of a large number of representers (see appendix A). In our formulation, all the equations and boundary conditions are written as weak constraints. Strong constraint cases are recovered simply by setting the appropriate parts of \mathbf{cov} to 0.

In the statistical interpretation of the GIM, representers give the prior error covariance for the model

solution. Specifically, if $\mathbf{e}_0 = \mathbf{v}_0 - \mathbf{v}_{\text{true}}$ is the error in the prior solution, then

$$\mathbf{r}_k = \langle \mathbf{e}_0 \mathbf{e}_0' \rangle \mathbf{I}_k. \quad (23)$$

In this interpretation, $(\mathbf{R}_{jj})^{1/2}$ is the expected prior error of the sampled quantity. Since the representer does not depend on the actual observed value, but only on the form of \mathbf{I}_k and assumed **COV** [see (21)], the expected variability of the model solution can in principle be assessed for any variable at any point. This property will be used to compute the open boundary covariance by nesting (section 4).

In the area shown in Fig. 1b harmonic constants for the radial components of surface velocities from the HF radars are available at over 900 locations. Computation of the full set of representers would be costly. Consequently, a reduced basis approach is taken (Egbert and Erofeeva 2002), as detailed in appendix B. A set of 68 representers, chosen as a reduced basis for model correction, is computed for data functionals $\hat{\mathbf{I}}_k$ corresponding to u and v at a fixed set of 34 surface locations (see Fig. 1b). The sum in (20) is replaced by a linear combination of the representers of the reduced basis [see (B1)]. A vector $\hat{\mathbf{b}}$ of 68 representer coefficients is sought to minimize the original cost function (19), that is, with all the data in the second penalty term. Note that no actual u and v data are measured at the reduced basis locations and the radial components from the two HF radars do not have to be reprocessed to obtain u and v components. The inverse solution $\hat{\mathbf{v}}_{\text{inv}}$ obtained with the reduced representer basis should be a close approximation of \mathbf{v}_{inv} , obtained with the full basis, provided representers for all the radial data can be approximated well as a linear combination of representers from the reduced basis set. With this approach, the same relatively small representer basis is used to find the solution in every time window, while the number and location of harmonically analyzed HF radar data meeting our accuracy threshold (see section 2) vary with time.

d. Computational setup

Most computations have been performed for the 38 km \times 56 km area shown in Fig. 1b that has a maximum depth of 250 m and includes Stonewall Bank, a moderate bathymetric rise. In this area, the data from both HF radars, at Yaquina Head and Waldport, provide information about two components of surface currents (Kosro et al. 1997). The computational grid has 1-km resolution in the horizontal and 21 σ levels in the vertical. Some computations described in sections 4 and 5 have also been performed for larger domains, at 2- and 3-km resolution.

Some dissipation is needed for numerical stability, to smear sharp gradients across wave beams resulting from the singular eddy forcing of the adjoint solution [see (21)]. Vertical eddy viscosity is taken to be horizontally uniform and varying with depth as $K_M = 10^{-2} \exp[-(z/10)^4] +$

10^{-4} ($\text{m}^2 \text{s}^{-1}$). Thus K_M approaches $10^{-2} \text{m}^2 \text{s}^{-1}$ in the upper 10 m (Wijesekera et al. 2003) and $10^{-4} \text{m}^2 \text{s}^{-1}$ in deeper water [a minimum value based on the parameterization of Pacanowski and Philander (1981)]. Increasing K_M to a constant value of $10^{-2} \text{m}^2 \text{s}^{-1}$ everywhere in the computational domain makes representer functions smoother and has only minor effects on the inverse solution. The linear bottom friction coefficient is a constant $r = 2.5 \times 10^{-4} \text{m s}^{-1}$. Unless otherwise specified, the mean of summer observations, 1961–71, at a station 46 km offshore of Yaquina Head (Smith et al. 2000) defines $\bar{\rho}(z)$ (the solid line in Fig. 2).

The prior model is forced at the open boundary by barotropic (depth averaged) currents, taken from a coastal-scale tidal inverse model in which TOPEX/Poseidon altimetry data are assimilated into the shallow-water equations (Egbert and Erofeeva 2002). The same prior solution is used in each time window. The prior solution provides an accurate estimate of the barotropic semidiurnal flow in the area (Erofeeva et al. 2003). However, deviations from the depth average are much smaller in the prior solution than in the ADP data (cf. Figs. 3b, and 3a), suggesting that most baroclinic signal propagates into the study area from outside. It is thus essential to estimate baroclinic fluxes at the open boundary (OB). This is done by means of DA.

To obtain an inverse solution, **COV** and **COV_d** should be specified. A great deal of uncertainty still exists about error statistics of HF radar data. Contributing factors include instrumental and processing errors, environmental conditions that affect signal scattering, and representation of the measurement in the dynamical model (i.e., the choice of \mathbf{I}_k). In the absence of detailed information about all these factors, in this study we assume the simplest diagonal covariance matrix for data errors, **COV_d** = $\sigma_d^2 \mathbf{I}$, where $\sigma_d = 0.02 \text{m s}^{-1}$ is based on typical MLS errors of the harmonic constants, and \mathbf{I} is the identity matrix.

4. Error covariance for the OB condition

Although errors in the linearized equations are surely not zero, we hypothesize that the major source of solution error in our small domain is the open boundary condition (10) and take this to be the only weak constraint. Thus, the rhs of Eqs. (11)–(14) will all be 0 except in the rows corresponding to open boundary nodes, where \mathbf{f}_u is equal to the specified barotropic current plus the error that is estimated through DA. Accordingly, in the matrix **COV** [see (21)] the only nonzero entries are the elements corresponding to the spatial covariance of OB fluxes: $C(\mathbf{x}_j, \sigma_j; \mathbf{x}_k, \sigma_k) = \langle q(\mathbf{x}_j, \sigma_j) q'(\mathbf{x}_k, \sigma_k) \rangle$, where $(\mathbf{x}, \sigma) = (x, y, \sigma)$ are positions on the OB, and q represents u at the western boundary and v at the northern or southern boundary of our rectangular domain.

In an oceanographic context, the input error covariances are known only approximately, at best. In practice,

the choice of covariance is guided by both the need to satisfy (at least approximately) a number of physical constraints and by ease of implementation. In our case, physical constraints with which the open boundary covariance must be consistent include the following: (i) at the OB, correction is provided only to deviations from the depth-average current (recall that deviations from the depth-average at the OB are 0 in the prior solution, while depth-averaged currents are believed to be reasonably accurate); (ii) u at the west, and v at the north and south boundary segments are correlated in a way to avoid nonphysical solution behavior in the corners of the computational domain; (iii) the expected prior error of the velocity at the OB is of the order of a few centimeters per second (this requirement can be readily satisfied by scaling the covariance); (iv) low vertical modes are dominant; and (v) the covariance has a dynamically consistent spatial structure since, for instance, propagating modes are correlated at different parts of the boundary in accordance with their phase speed and direction of propagation. This list of constraints for the covariance can be refined and completed as more experience is gained in regional DA modeling, for example, via the analysis of the representers and the inverse solution.

We tested two covariances that differ in the extent to which criterion (v) is addressed. The Type I covariance is constructed in an ad hoc manner guided by the principles outlined above. The hope with such an approach, which has been frequently used in the past, is that the covariance enforces smoothing and thus regularizes the inverse problem, but details of structure do not matter so much. The Type II covariance is obtained following a nesting approach, using the representer solution in a larger domain.

The Type I covariance is assumed to be separable as

$$C(\mathbf{x}_j, \sigma_j; \mathbf{x}_k, \sigma_k) = C_1(\mathbf{x}_j, \mathbf{x}_k)C_2(\sigma_j, \sigma_k), \quad (24)$$

where C_1 and C_2 are both symmetric and positive definite. To ensure regularity of u and v around corners (criterion ii), velocity \mathbf{u} in a σ layer is partitioned into a nondivergent flow, described by a streamfunction ψ , and an irrotational flow, described by a potential ϕ . Spatial correlations for both $\langle\psi\psi\rangle$ and $\langle\phi\phi\rangle$ are assumed to have a Gaussian form, with $\langle\psi\phi\rangle = 0$. Then C_1 is obtained in the standard way as a linear combination of derivatives of $\langle\psi\psi\rangle$ and $\langle\phi\phi\rangle$. To derive C_2 , the errors in the OB boundary condition are projected onto local flat bottom rigid-lid baroclinic vertical modes associated with stratification $N(z)$, and the modal amplitudes are assumed to be uncorrelated (Kurapov et al. 2002).

Our construction of the Type I covariance is already quite involved, and yet some rather obvious issues (e.g., appropriate phase velocities along the boundaries) have not even been considered. Nesting, based on the interpretation of the representers as covariances [see (23)], provides a simpler and more general approach to construction of a physically consistent covariance for the

OB condition. The Type II covariance for u and v at the boundary of the local grid can be obtained as a result of the representer computation on a larger, coarser-resolution grid. In our application, the local grid is nested in a large-scale grid covering an area of $177 \text{ km} \times 435 \text{ km}$ with 3-km resolution in the horizontal and 11 σ surfaces in the vertical (this domain is larger than the area shown in Fig. 1a); maximum water depth is set to 500 m. For the large-scale grid, the open boundary condition is a strong constraint, and the horizontal momentum equation (2) is a weak constraint. To form matrix **cov** for the large-scale model, the covariance for the errors in each u and v component of Eq. (2) is written as in (24) with \mathbf{x}_j and \mathbf{x}_k now inside the domain. Here C_1 is Gaussian with a decorrelation length scale of 10 km and C_2 is as above, based on the local flat bottom modal decomposition. The Type II covariance is computed on the large-scale grid as a representer matrix for velocity observations at the locations corresponding to the OB nodes of the small, nested domain (observations are for u at the west, and v at the north and south).

Figure 5 illustrates the difference in the spatial pattern of the Type I and Type II covariances at the northern boundary. Distinctive features of the Type II covariance are determined by the superinertial wave dynamics in the stratified medium. The first baroclinic mode is preserved, indicated by the 180° phase difference between the surface and bottom. In contrast to the strictly real-valued Type I covariance, the Type II covariance is complex-valued allowing for progressive phase changes along the boundary. The fact that energy (and hence information) propagates along wave characteristics manifests itself in the pattern of phase lines oriented at an angle to the horizontal. For the assumed $\bar{p}(z)$, the bottom slope near the coast is close to critical (see Fig. 1a), and here energy density may be intensified near the bottom. Accordingly, the covariance amplitude, as well as the expected prior error, is increased at this spot.

In the “hand-made” Type I covariance (see Fig. 5b), the first baroclinic mode is also dominant, by construction. However, the other dynamically realistic features evident in the Type II covariance are missing.

5. Computations with synthetic data: Effect of the OB condition error covariance

When inverting HF radar surface currents (section 7), ADP velocities are available to validate the data assimilation results only at one location. To see better how the choice of the covariance for the open boundary condition (section 4) affects the solution everywhere in the computational domain, tests with synthetic data have been performed. To generate synthetics we first compute the model solution for the larger $94 \text{ km} \times 150 \text{ km}$ area shown in Fig. 1a, forced with depth-averaged currents from the shallow-water inverse model. In this area, an internal tide with an amplitude comparable to observations at the ADP site is generated over the shelf break.

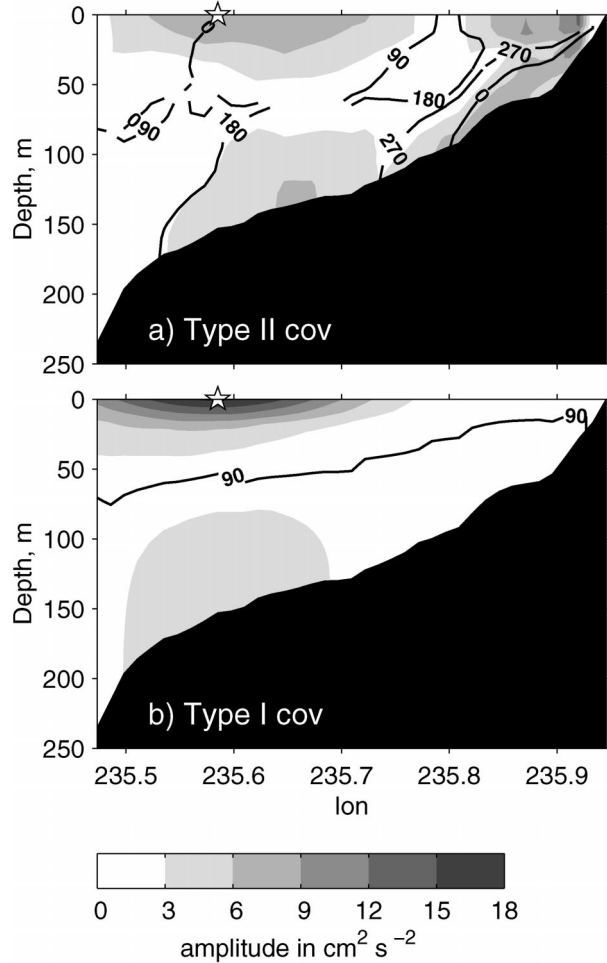


FIG. 5. Error covariance of v at the location shown as the star and v' everywhere else at the northern boundary: (a) Type II covariance (obtained by nesting), complex amplitude is indicated by shading, phase is contoured; (b) Type I covariance [real valued, 90° phase line divides positive (at the top) and negative (near the bottom) values].

This solution is obtained with resolution $2 \text{ km} \times 2 \text{ km}$ in the horizontal, and 11 σ surfaces in the vertical, with the maximum water depth set at 500 m. The total, depth-dependent currents obtained from the 2-km resolution solution are then used to force the model in the small, higher-resolution domain to generate a validation solution. Synthetic u and v data are sampled from this solution at 34 surface locations, shown as circles in Fig. 1b, and random noise of amplitude 0.02 m s^{-1} is added. The inverse solution obtained with these synthetic data is compared with the validation solution by computing the rms difference of the complex-valued three-dimensional velocity fields.

Since the choice of covariances in the cost function (19) is always a hypothesis, it is important to learn about sensitivity of the inverse solution to their specification. In reality, uncertainty exists first of all about the magnitude of the errors, both for the model and the data. The relative amplitude of the assumed errors in the in-

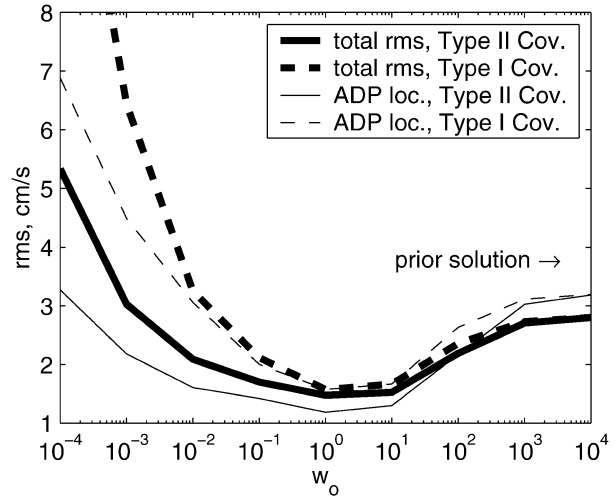


FIG. 6. Effect of the OB error covariance on the rms difference between the inverse and validation u for a range of weights w_o in the computations with synthetic data. Thick lines: total rms; thin lines: depth-average rms at the ADP location. Solid lines correspond to the Type II covariance (obtained by nesting) and dashed lines to the Type I covariance.

puts can be controlled by scaling the covariances in (19). Let us replace the covariance \mathbf{cov} with $w_o^{-1} \mathbf{cov}$, where w_o is a scaling factor for the open boundary penalty term in the cost functional (we again consider a strong constraint case, so only OB elements of \mathbf{cov} are non-zero). The rms error of the inverse solutions, computed with the Type I and Type II covariances, is plotted as a function of w_o in Fig. 6, where the total rms error is averaged over the whole domain. For large values of w_o the solution rms error is close to that for the prior solution. The minimum rms error is attained for $w_o = 1$ both for Type I and Type II covariances. For lower w_o , the surface data is fit better, but the DA model performance deteriorates. Very low values of w_o correspond to the case of effectively no open boundary condition penalty term in the cost function. The problem becomes ill-conditioned; solution irregularities originate near the open boundary and propagate inside the domain (Foreman et al. 1980; Kivman 1997). In our case, spurious small scale baroclinic vertical modes appear in the solution for low w_o .

For the full range of w_o , the inverse solution obtained with the Type II covariance has a smaller total rms error than that obtained with the Type I covariance. Furthermore, for $w_o < 1$ the Type II solution is much less sensitive to weight misspecification than Type I. Overestimating the expected open boundary error amplitude by an order of magnitude corresponds to $w_o = 0.01$ and for this weight the Type I covariance yields a solution with rms error already worse than the prior, while the Type II solution still provides improvement. For small w_o , the total rms error is larger than the depth-average rms error at the ADP location, indicating that the solution quality is worse at the boundaries. This is illus-

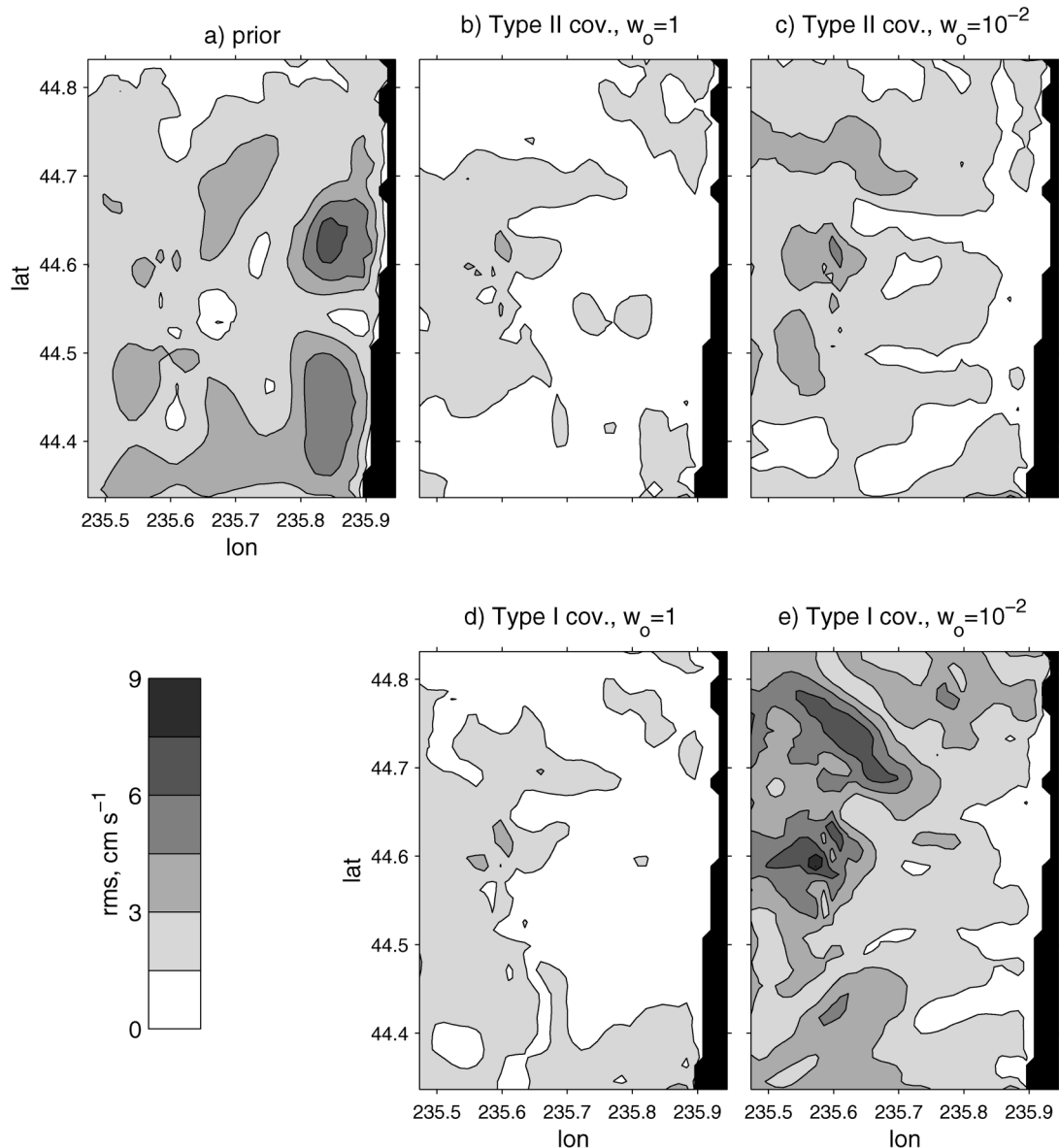


FIG. 7. Effect of the OB covariance on the spatial pattern of the depth-averaged rms error, in the computations with synthetic data: (a) rms of the prior \mathbf{u} ; (b) rms of the inverse \mathbf{u} , obtained with the Type II covariance, $w_o = 1$; (c) same as (b), but with underspecified weight $w_o = 10^{-2}$; (d) rms of the inverse \mathbf{u} , obtained with the Type I covariance; (e) same as (d), but with underspecified weight $w_o = 10^{-2}$.

trated in Fig. 7 with 2D plots of depth-averaged rms. Although for $w_o = 1$ the two covariances yield solutions of comparable quality (cf. Figs. 7b and 7d), for the Type I covariance the solution corresponding to an underestimated weight ($w_o = 0.01$) is much more irregular (cf. Figs. 7c and 7e). Thus, results from experiments with synthetic data indicate improvements with the Type II covariance, which we use to obtain results in the next sections unless otherwise specified.

Note that the choice of the open boundary error covariance affects transparency of the boundary to outgoing waves. Even if perfect radiation boundary con-

ditions are used, but treated as a weak constraint, a poor choice of covariance can result in partial reflection. Conversely, in data assimilation there is an opportunity to make reflective boundary conditions more nearly radiative by an appropriate choice of the covariance. We hypothesize that the improved performance of the Type II covariance should have better radiation properties since it is computed in a larger domain that “does not know” about the existence of the small domain. Improvement of wave radiation by the choice of the OB covariance in a DA model should be explored

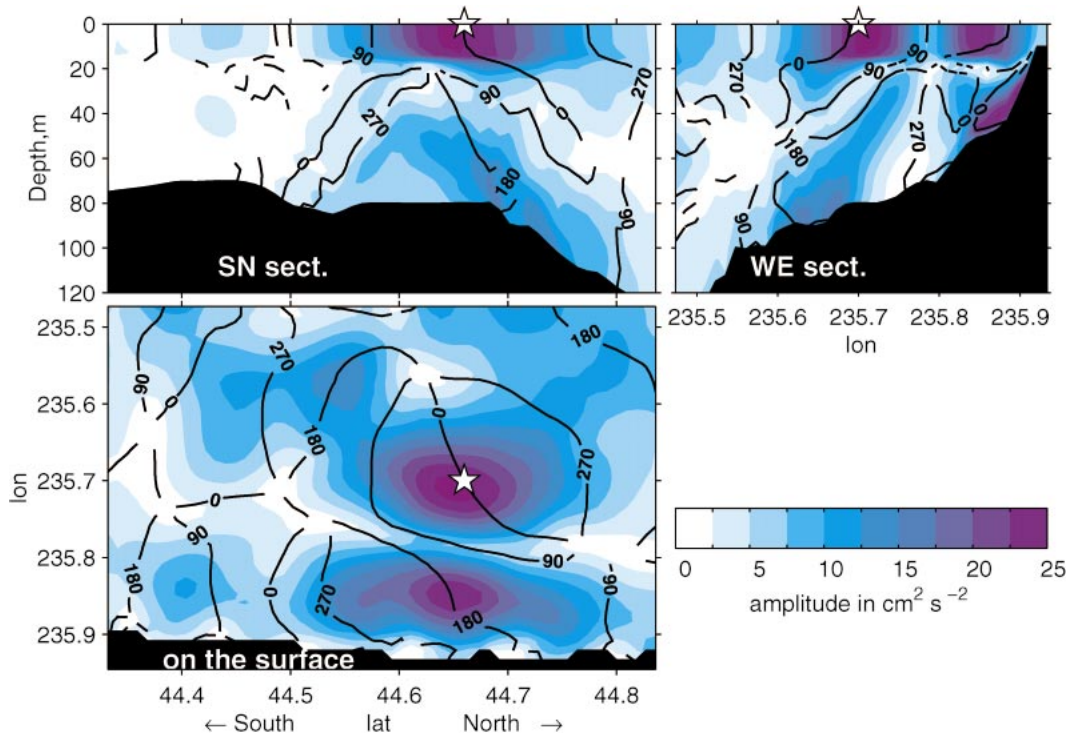


FIG. 8a. Component v of the representer for the surface v measurement computed with Type II covariance. (upper) South–north (SN) and west–east (WE) cross sections through the observation location (at 44.66°N , 235.7°E marked as the star). (lower) Plan view of representer on the surface. Here and in Figs. 14 and 15 amplitude is shaded, and phase angle, contoured, increases in the direction of phase propagation.

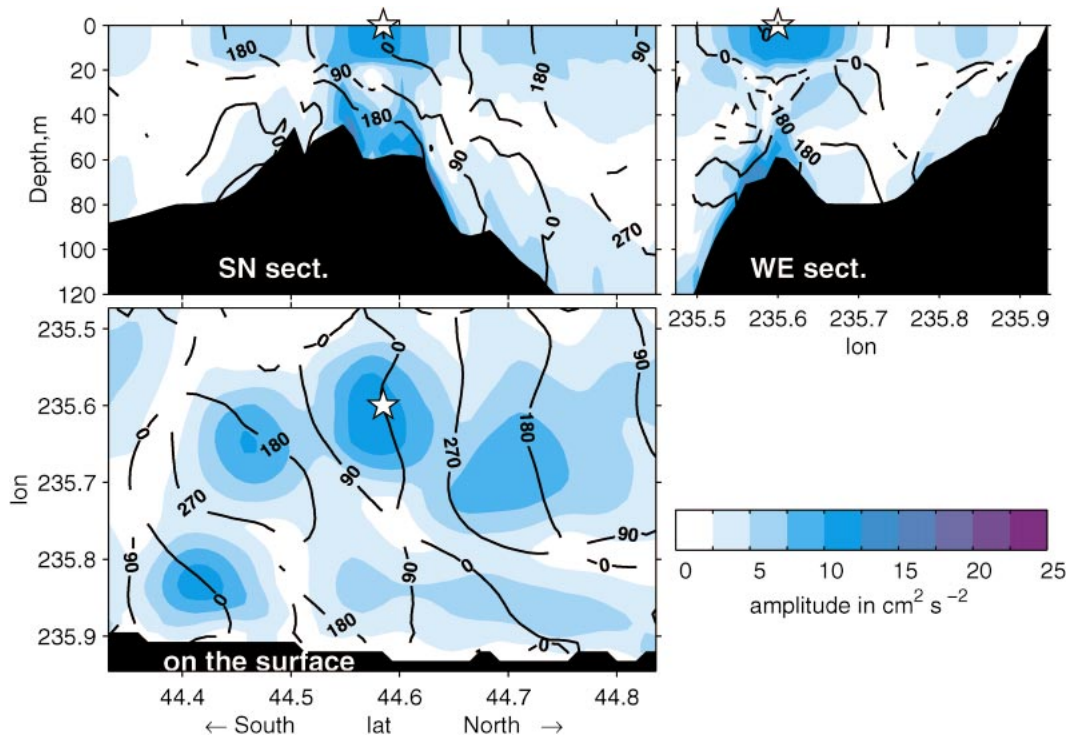


FIG. 8b. As in Fig. 8a but for observation at 44.58°N , 235.6°E over Stonewall Bank.

in detail in future studies, since it may be applicable in coastal data assimilation in a broader context.

6. Representer structure

Analysis of the representer is useful since, for instance, it shows how information from the data projects (in space and time) into the inverse solution. Figure 8 shows the v component of the representer for a surface v measurement. Zero phase contours in the alongshore, south–north (SN) section extend from the observation point downward along wave characteristics, giving a clear illustration of the direction of information propagation. For a measurement made northeast of Stonewall Bank (Fig. 8a), two such zero phase lines are apparent to the north and south of the observation point. In the west–east (WE) cross-shore section, the representer structure is similar to that of the Type II input covariance at the northern boundary (see Fig. 5a) with a complicated phase pattern resulting from superposition of incoming and reflected waves. For a v measurement made over Stonewall Bank (Fig. 8b), the representer structure is more strongly influenced by bathymetry. Slopes of phase contours are close to the bathymetric slope at the northern flank of the bank; that is, the slopes are close to critical.

In both Fig. 8a and 8b, the representer phase on the surface implies propagation from northwest to southeast. This pattern may to some degree result from the OB covariance (Type II), and hence reflect dynamics in the larger domain. However, a similar pattern can be seen in the representer obtained with the Type I covariance, so the preferred phase direction may well be the effect of the local dynamics.

A representer for a sea surface elevation is qualitatively similar to that for a surface velocity, suggesting that η measurements, for instance, from a satellite altimeter, may in principle be a valuable source of information about baroclinic tides. However, the computed expected prior error for η in our domain is less than 1 cm. A signal of such small amplitude contaminated by observational noise would be difficult to observe on the Oregon shelf.

7. Inversion of HF radar surface currents off Oregon, May–July 1998

Analysis of the representer suggests that information from the surface velocity measurements propagates to depth along wave characteristics. This fact, as well as results from the experiments with synthetic data, shows that in principle assimilation of harmonically analyzed surface currents can be useful for predicting internal tidal flows at depth. Assimilation of real data allows a rigorous test of the actual value of this data and our model. It also shows where efforts should be first directed to improve the inverse model.

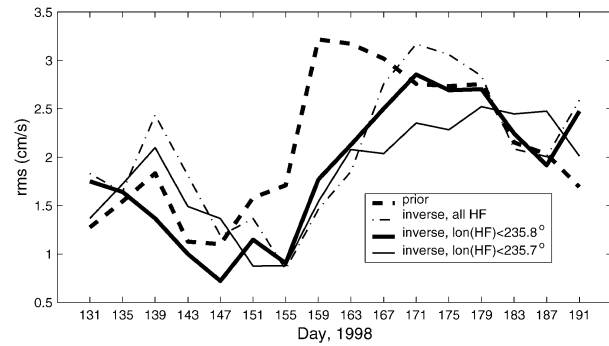


FIG. 9. Rms difference of the ADP validation data and model solutions, for each time window: prior solution (dashed thick line), inverse solution with all the radial data assimilated (thin dot-dashed line), inverse solution with radial data assimilated west of 235.8° (solid thick line), and inverse solution with data assimilated west of 235.7° (solid thin line).

a. Comparison with ADP validation data

Figure 9 shows the depth-average rms difference between the computed solutions and ADP validation data for each of the 16 overlapping 2-week time segments. Assimilation of the whole set of HF radar harmonic constants yields a reduction in rms misfit compared with the prior solution only for days 151–167 (compare the thin dot-dashed and thick dashed lines in Fig. 9). When only data west of longitude 235.8° are assimilated, improvement is obtained for a number of earlier and later time segments (thick solid line in Fig. 9). For days 171–179 the inverse solution rms misfit is less than the prior solution rms misfit if only radial currents west of longitude 235.7° are assimilated (thin solid line in Fig. 9). However, data in the stripe $235.7^{\circ} < \text{lon} < 235.8^{\circ}$ still contain valuable information about flow in the beginning of the assimilation period (compare the thin and thick solid lines for days 139–147). The reductions in misfit that result from not fitting the near-coast data suggest that the model assumption of horizontally uniform background stratification may be poor for at least a part of the study period. Since the winds during the study period are generally upwelling favorable (see Fig. 4), isopycnals are presumably lifted near the coast (although we have no measurements to verify this). During short relaxation events (e.g., days 155–157 and 160–162; see Fig. 4) the pycnocline will not return to horizontal if the cross-shore pressure gradient is geostrophically balanced by alongshore currents, and so deviations from our assumed, horizontally uniform stratification will tend to increase throughout the study period. Our model is likely to get worse as the upwelling continues. In these conditions, model wave characteristics near the coast may especially be expected to have the wrong slope, so representers will project information from the near-coastal data to the wrong locations offshore.

The inverse solution does not offer improvement in terms of rms for days 131 and 135 when the ADP data

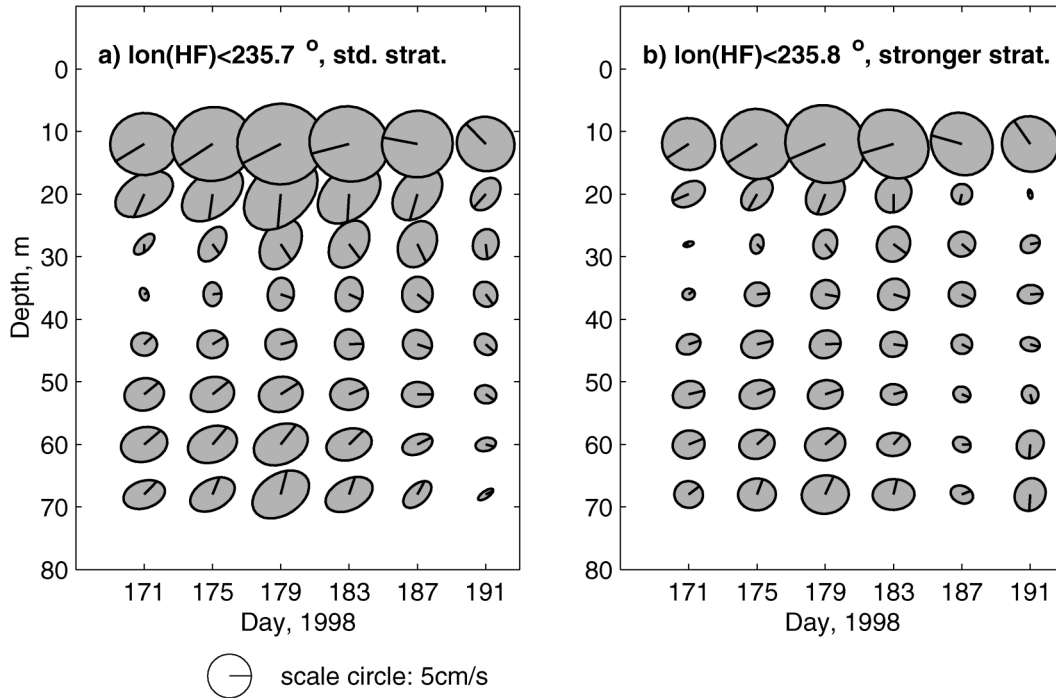


FIG. 10. Ellipses of horizontal baroclinic velocity at the ADP data locations for days 171–191, obtained by (a) assimilating HF radar radials only west of longitude 235.7°E (instead of 235.8°E as plotted in Fig. 3c), or (b) assuming stronger stratification (dashed line in Fig. 2) in the first 50 m. Both of these variants on the standard case of Fig. 3c have surface-intensified baroclinic currents, similar to the ADP data.

show low baroclinic signal (see Fig. 3a), or beyond day 179. Incorrect background stratification in the model may explain the poorer performance at the end of the studied period. Note that the relative increase in the rms misfit after day 159 coincides with the rise in baroclinic current amplitudes (see Fig. 3a).

Figure 3c shows horizontal tidal ellipses for the computed baroclinic currents at the ADP location for the case $\text{lon}(\text{HF}) < 235.8^\circ$. Qualitatively, the inverse model captures much of the internal tide variability in the validation data (cf. Fig. 3a). Starting with day 171, the ADP data and inverse solutions have significantly different phases near the surface. Also, in the ADP data the internal tide is surface intensified, a feature not reproduced by the inverse solution. Surface intensification is reproduced in the inverse solution if more data are dropped near the coast [case $\text{lon}(\text{HF}) < 235.7^\circ$, Fig. 10a]. Alternatively, surface-intensified currents are also obtained if the assumed stratification is increased in the upper 50 m (on a flat bottom, such change in stratification would have a similar effect on the shape of the first baroclinic vertical mode). Stratification corresponding to the latter experiment is shown as a dashed line in Fig. 2, and the resulting baroclinic tidal ellipses at the ADP location in Fig. 10b.

Covariability of the validation data and computed solutions can be quantified by the complex correlation

$$CC(\mathbf{q}_1, \mathbf{q}_2) \equiv |CC|e^{i\phi_{cc}} = \frac{\mathbf{q}'_1 \mathbf{q}_2}{(\mathbf{q}'_1 \mathbf{q}_1)^{1/2} (\mathbf{q}'_2 \mathbf{q}_2)^{1/2}} \quad (25)$$

and the gain G , which complements $|CC|$ in quantifying the relative magnitude of \mathbf{q}_1 and \mathbf{q}_2 :

$$G(\mathbf{q}_1, \mathbf{q}_2) = \frac{|\mathbf{q}'_1 \mathbf{q}_2|}{\mathbf{q}'_2 \mathbf{q}_2}. \quad (26)$$

Here the vector of complex-valued entries \mathbf{q}_1 represents the computed solution, \mathbf{q}_2 the corresponding validation data, and prime again denotes complex conjugate transpose. The vectors \mathbf{q}_1 and \mathbf{q}_2 may represent vertical profiles of currents for a given day, or alternatively time series of harmonic constants at a given depth.

Covariability in the vertical is assessed for each day for the case $\text{lon}(\text{HF}) < 235.8^\circ$ (Fig. 11), separately for u (dashed lines) and v (solid lines). The prior solution, which does not vary in time, is used to provide reference levels for all parameters (shown in Fig. 11 as thin lines). The inverse solution (bold lines) captures variability in the vertical better than the prior solution in terms of each $|CC|$, G , and ϕ_{cc} for most of the study period. Covariability with time is assessed at each vertical level where ADP data is available. When the whole time series (days 131–191) is used, the inverse solution is of better quality than the prior solution in terms of $|CC|$ and G in the larger part of the water column. Improvement in terms of ϕ_{cc} is marginal. If the statistics are

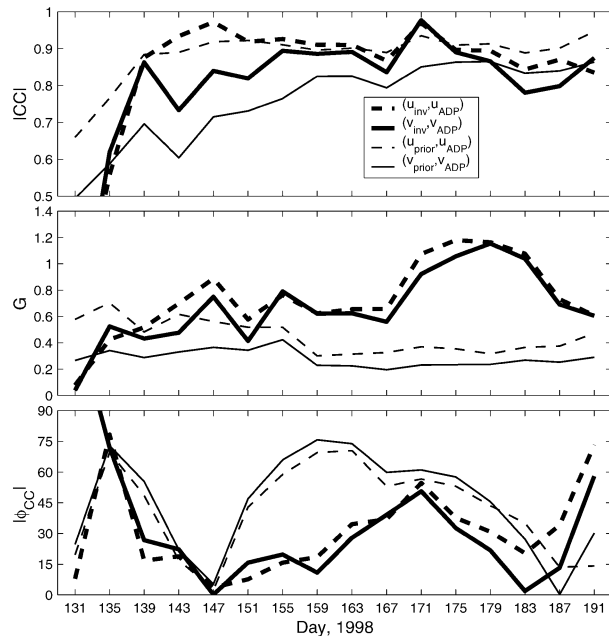


FIG. 11. Functions showing ADP-prior and ADP-inverse solution covariability in the vertical, for each day: amplitude $|CC|$ and phase ϕ_{CC} of complex correlation, and gain G (26), where \mathbf{q}_2 represents the ADP measurements; lon(HF) < 235.8°E.

computed only for days 139–167 for which the solution rms error is improved, the inverse solution is of better quality than the prior solution in terms of all the three criteria (Fig. 12). Note that at middepths, where the magnitude of the baroclinic currents is low, estimates of CC and G lose meaning. The fact that all the criteria are improved below 40 m provides a clear demonstration of the value of surface currents from HF radars in constraining subsurface tidal flows.

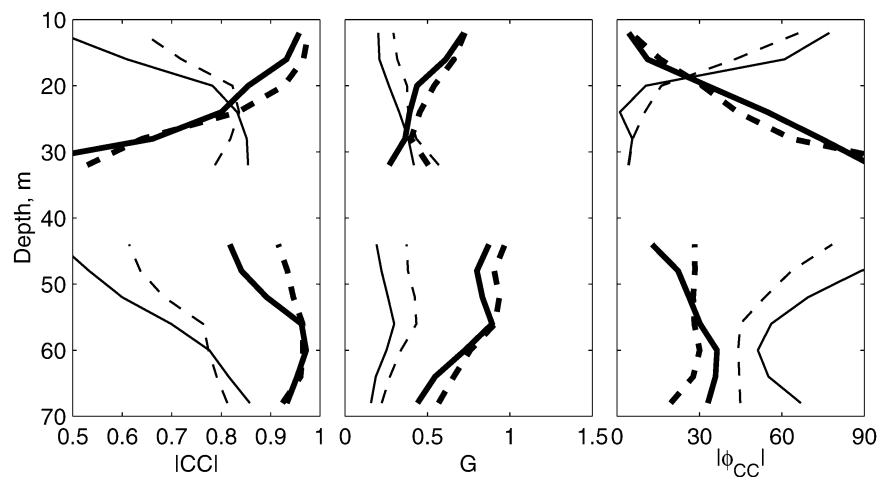


FIG. 12. Functions showing ADP-prior and ADP-inverse solution covariability with time, at each depth [lon(HF) < 235.8°E]: statistics for days 139–167, for which the solution rms error is improved. Legend as in Fig. 11. Values corresponding to rms amplitudes of the ADP fluctuation currents < 1 cm s⁻¹ are not shown.

Our results suggest that the assumption of horizontally uniform background stratification is a significant deficiency in our model. To the extent that this is true, our initial hypothesis that the only significant source of error is in the open boundary forcing is questionable. The hypothesis about the errors can be tested formally (Bennett 2002, his section 2.3.3). Under the assumption that the errors are Gaussian with zero mean and covariance \mathbf{cov} and \mathbf{cov}_d , twice the cost function $2J(\mathbf{v}_{inv})$ is itself a random variable having a χ^2 distribution with $2K$ degrees of freedom (a factor of 2 appears since the data are complex valued); recall K is the number of assimilated data ($K \approx 540$ in the case lon(HF) < 235.8°). In particular, the mean value of $2J(\mathbf{v}_{inv})$ should be $2K$ and the variance $4K$. Note that the reduced basis estimate $J(\hat{\mathbf{v}}_{inv})$ (B2) is larger than the full basis estimate $J(\mathbf{v}_{inv})$, although these values should be close since $\hat{\mathbf{v}}_{inv} \approx \mathbf{v}_{inv}$. For our series of inverse solutions $J(\hat{\mathbf{v}}_{inv})/K$ is in the range of 2–12 implying that hypothesis about errors should be rejected. The data term in $J(\hat{\mathbf{v}}_{inv})$ accounts for about 95% of the total cost function value, so changing \mathbf{cov}_d should have a stronger effect on $J(\hat{\mathbf{v}}_{inv})$ than \mathbf{cov} . In this respect, better knowledge of the error model for the data would be desirable.

By itself the χ^2 criterion does not say where the deficiency in our hypothesis is; e.g., either one or both of \mathbf{cov} and \mathbf{cov}_d could be misspecified. Increasing our estimate of the data error standard deviations from 2 to 3 cm s⁻¹ decreases $J(\hat{\mathbf{v}}_{inv})/K$ to a value near 1 over much of the first half of the study. Such an increase in data error is quite plausible. However, in the second half of the study, when $J(\hat{\mathbf{v}}_{inv})/K > 6$, unreasonably large increases in the data error would be required to bring the test statistics down to acceptable levels, implying that some modifications to \mathbf{cov} are required at least for this time period. Increasing the expected magnitude of errors

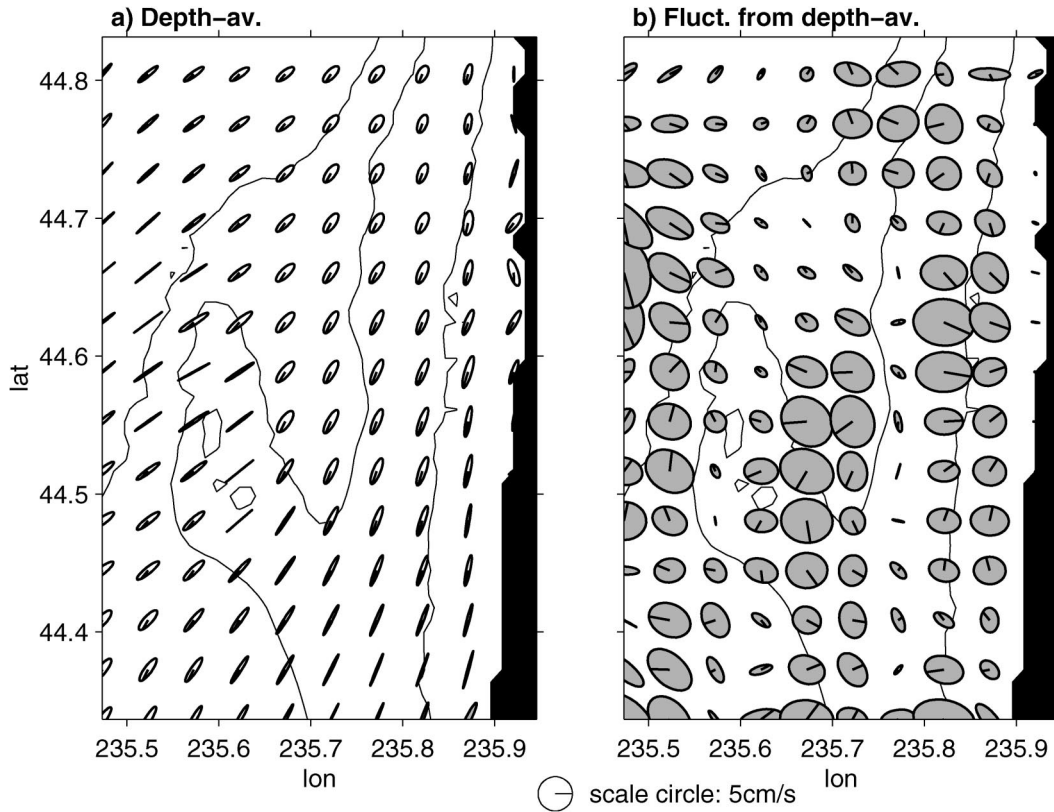


FIG. 13. Tidal ellipses of the inverse solution, day 139: (a) depth-averaged current and (b) deviations from the depth-average near the surface. Contours are isobaths of 50, 75, and 100 m.

in the baroclinic boundary conditions has little effect on the value of $J(\hat{\mathbf{v}}_{\text{inv}})$. Furthermore, similar values are obtained for the test statistics with both the Type I and Type II covariances. This suggests that as long as we retain a strong constraint formulation, we cannot make reasonable adjustment to \mathbf{cov} and \mathbf{cov}_d that are statistically consistent with observations. Only by admitting errors in our equations (and hence significantly changing our assumption about \mathbf{cov}) could we achieve consistency between the HF radar data and our simplified dynamical model.

Although our analysis of the χ^2 criterion suggests that the inverse solutions satisfying our simplified dynamical equations should be considered suboptimal, they are still a significant improvement over the prior. Moreover, our simplified model has allowed us to achieve the goals of the study, stated in the introduction. From here we can improve our model or attempt to use the dynamics as a weak constraint.

b. Variability of the tidal fields

The internal tide is a significant contributor to M_2 tidal flow at the surface. Deviations at the surface frequently exceed the depth-averaged tidal current (Fig. 13). The computed internal tide varies on a scale of about 10 km, in accordance with the theoretical esti-

mate. The horizontal tidal velocity vectors of the barotropic flow rotate counterclockwise and ellipses are strongly polarized, consistent with shelf-modified Kelvin wave dynamics (e.g., Munk et al. 1970). At the same time, vectors of the horizontal baroclinic current (defined as the deviation from the depth-averaged) are less polarized and rotate clockwise, consistent with kinematics of free planar waves. Indeed, the direction of rotation depends on the phase angle α between complex amplitudes u and v , such that, by our convention, $\sin \alpha < 0$ corresponds to clockwise rotation. If the rotary components of the current are defined as $v^\pm = (u \pm iv)/2$ then $|v^\pm|^2 = (|u|^2 + |v|^2 \pm 2|u||v| \sin \alpha)/4$. So, for the vector rotating clockwise, $|v^+| < |v^-|$. It is easily shown that for a planar wave (i.e., a wave involving a pressure gradient only in one direction) $|v^+|/|v^-| \approx (\omega - f)/(\omega + f)$, and so in our case $|v^+| \ll |v^-|$ and rotation should be clockwise. To see clearly the phase propagation of the baroclinic wave, we plot the v^- component of the baroclinic current at the surface for days 139, 155, and 167 (Fig. 14). A complicated phase pattern is seen and is due to the superposition of incident waves and those reflected from topography (or the coast). Despite day-to-day variability in details, dominance of phase propagation from the northwest to the southeast of the area is a persistent feature.

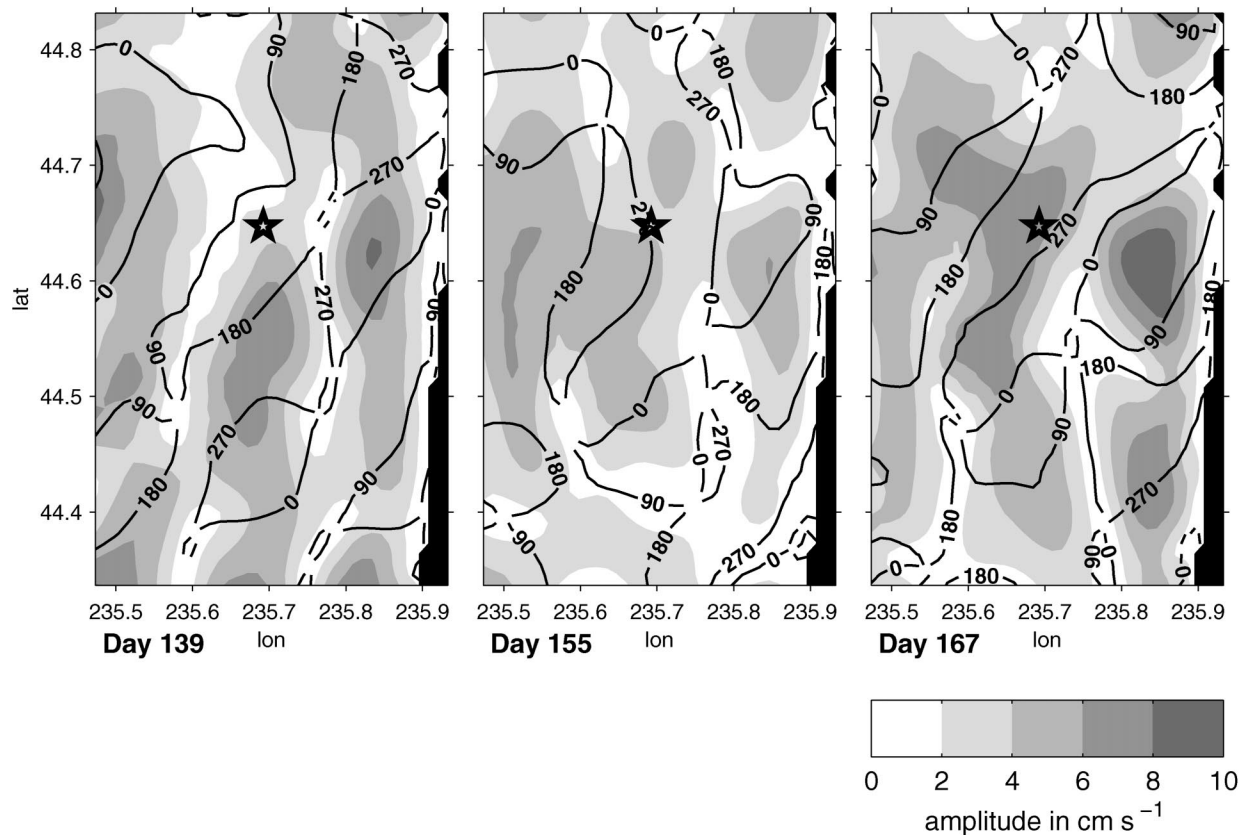


FIG. 14. Rotary current $(u - iv)/2$ at the surface, deviations from the depth average, for days 139, 155, and 167. The star shows an ADP location.

The same phase propagation pattern can be found in isopycnal displacements. The dominance of shoreward phase propagation is seen most clearly in the cross-shore section through the northern part of the study area, where topography is relatively simple (Fig. 15, left panels). In the cross section through Stonewall Bank (right panels) topography complicates the phase pattern. For instance, for day 155, the phase of the isopycnal displacements suggests scattering off Stonewall Bank to the west and east. The magnitude of the computed isopycnal displacements at middepth is 2–6 m, sometimes reaching 10 m over Stonewall Bank.

The magnitude of the internal tide can also be quantified by computing baroclinic kinetic energy (per unit mass) averaged over the tidal period, $KE_{BC} = |\mathbf{u} - \mathbf{u}_1|^2/4$, where \mathbf{u}_1 is the complex amplitude of the depth-averaged current. Near the surface, superposition of waves incident and reflected from the coast results in zones of higher and lower KE_{BC} . If KE_{BC} is averaged over a series of days, zones of higher internal tide energy near the surface appear to be aligned along the coast (Fig. 16a). This pattern may superficially suggest surface bounce points of beams of internal tide oriented perpendicular to the coast. However, this pattern and cross-shore spatial scales are in fact consistent with energetics of a Poincaré wave in the presence of a coast

(e.g., Gill 1982, chapter 10). This can be shown using an analytical model of the stratified flow over the flat bottom (Wunsch 1975), where the flow is separated into a series of vertical modes and the dynamics of individual modes are governed by shallow-water-type equations:

$$i\omega\mathbf{U} + \mathbf{f} \times \mathbf{U} = -\nabla P, \quad (27)$$

$$i\omega P + (1/\lambda^2)\nabla \cdot \mathbf{U} = 0. \quad (28)$$

Here $\mathbf{U}(x, y)$ and $P(x, y)$ are modal amplitudes of \mathbf{u} and p/ρ_0 , correspondingly. In the case $N = \text{const}$, $\lambda \approx \pi/(NH)$ for the first baroclinic mode. If the coast is straight and is aligned with the y axis at $x = 0$, then the solution is sought as a superposition of an incident wave $P = P_a \exp(-ikx - ily)$ and a wave reflected from the coast $P = P_b \exp(ikx - ily)$. Let us put $P_a = 1$. To satisfy a no-flow boundary condition at $x = 0$, $P_b = -c_1/c_1^*$, where $c_1 = -(f + i\omega k)/(\omega^2 - f^2)$, and the asterisk denotes complex conjugate. For a Poincaré wave, $k^2 + l^2 = \lambda^2(\omega^2 - f^2)$, and so the cross-shore component of the wave vector k depends on the stratification, depth, and the direction α of the incident wave phase propagation. Proceeding with the solution, one obtains

$$|U|^2 = 2|c_1|^2[1 - \cos(2kx)], \quad (29)$$

$$|V|^2 = 2|c_2|^2[1 - \cos(2kx + \phi)], \quad (30)$$

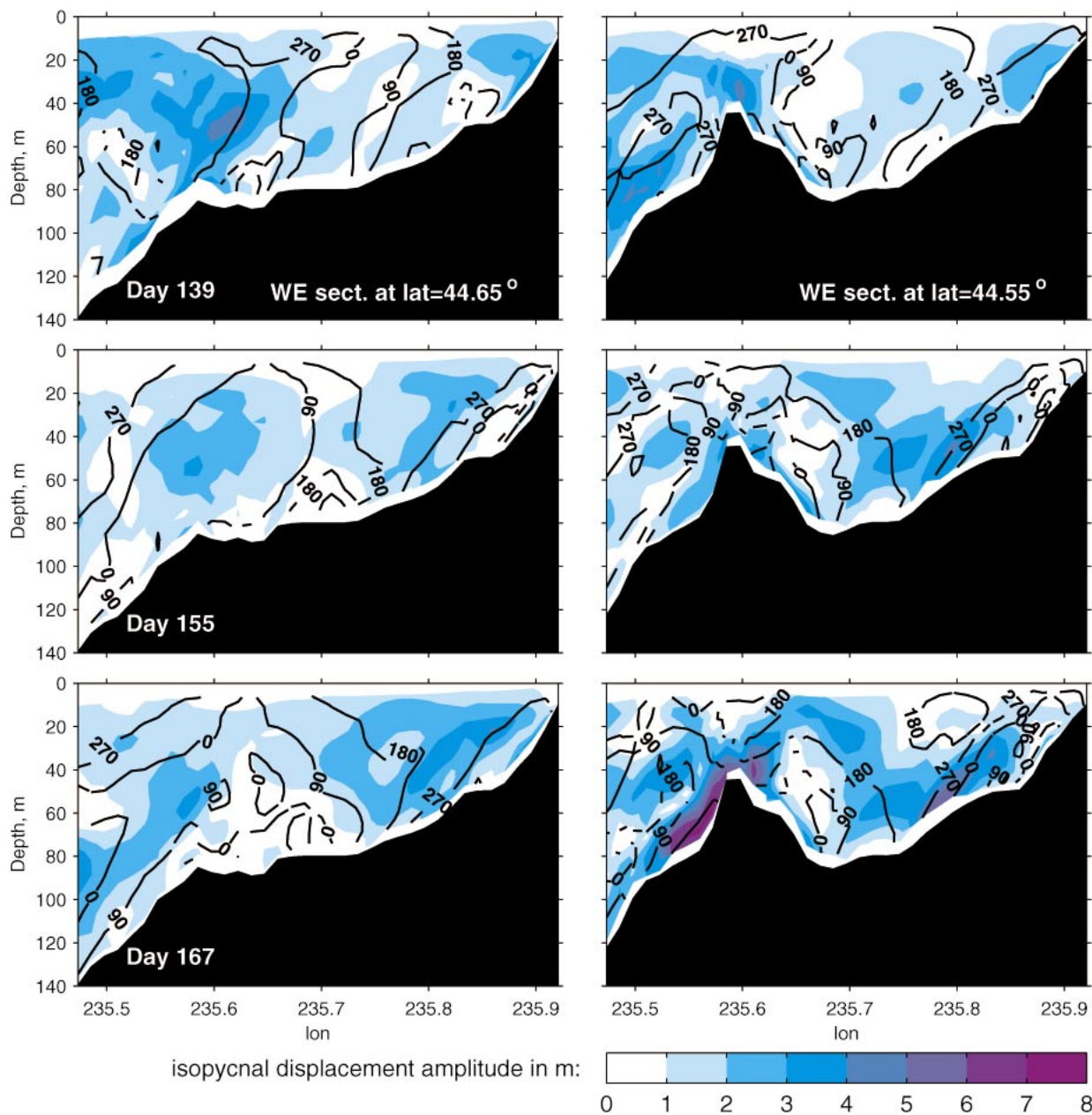


FIG. 15. Amplitude and phase of isopycnal displacements, shown in two west–east (WE) cross sections, (left) one north of Stonewall Bank and (right) another through the bank, for days 139, 155, and 167.

where U and V are cross-shore and alongshore components of \mathbf{U} , $c_2 = (l\omega + ifk)/(\omega^2 - f^2)$, and $\phi = \arg[c_1 c_2^*/(c_1^* c_2)]$. Thus, near the coast, the kinetic energy of a baroclinic mode [the sum of (29) and (30)] does not depend on the alongshore coordinate and has zones of higher and lower magnitude in the cross-shore direction. Relevant to our study, we take an estimate of $N = 0.01 \text{ s}^{-1}$, $H = 100 \text{ m}$, and $\alpha = -45^\circ$ (a wave from the north-west) so that for baroclinic mode 1 $|c_1| = |c_2|$, $\phi = -34^\circ$, and $\pi k^{-1} = 15 \text{ km}$. Then, the first three zones of max KE_{BC} should be 7, 22, and 37 km from the shore, in close agreement with results of Fig. 16a.

Near the bottom, the intensity of the internal tide is determined more directly by the interaction with local topography, with larger KE_{BC} found on the flanks of Stonewall Bank (Fig. 16b). In the vertical, west–east cross section north of the bank (Fig. 16c), an energy minimum at middepth indicates dominance of the first baroclinic mode. In the west–east section through Stonewall Bank, high KE_{BC} in the water column over the bank indicates that internal tide beams generated over this topographic feature extend to the sea surface.

Day-to-day evolution of KE_{BC} averaged over the tidal period and over the whole computational domain (Fig.

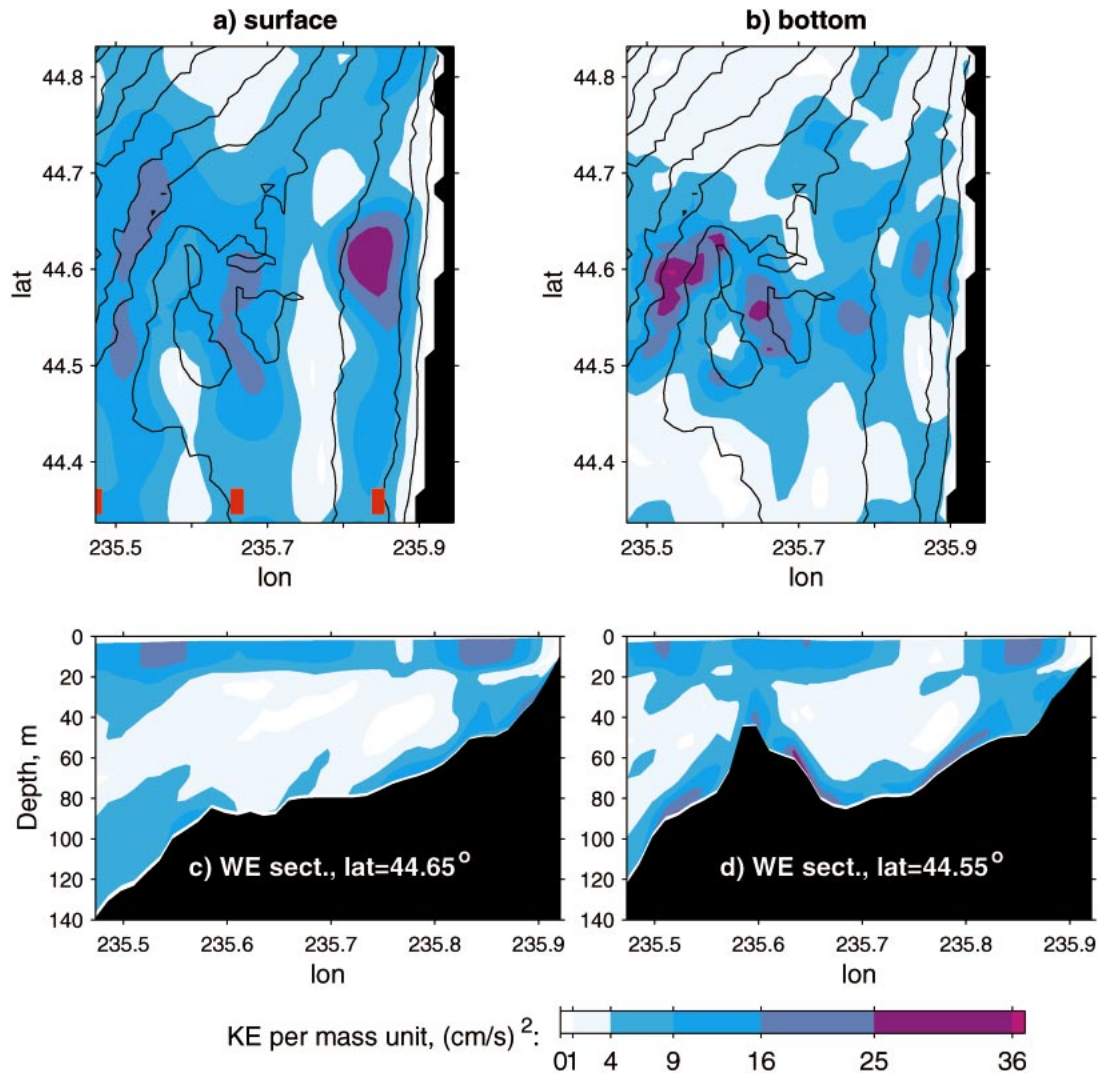


FIG. 16. Plot of KE_{BC} averaged over a series of days 139–167: (a) at the surface [contour lines are bathymetry every 20 m; red marks show locations of zones of maximum KE_{BC} predicted by the theoretical Poincaré wave solution (29)–(30)]; (b) at the bottom; (c) west–east cross section north of Stonewall Bank; and (d) west–east cross section through Stonewall Bank.

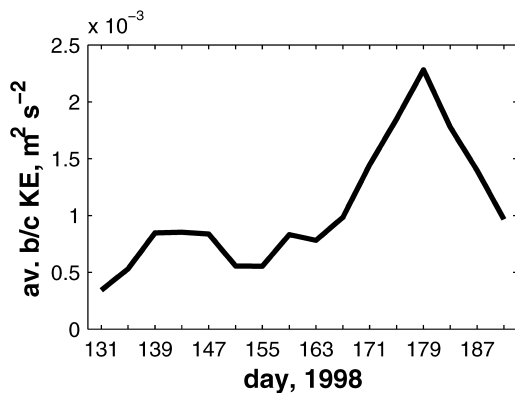


FIG. 17. Plot of KE_{BC} averaged over the whole computational domain and over the tidal period, shown for a series of days.

17) shows larger values between days 171 and 187, consistent with the impression from Fig. 3. Note that the level of KE_{BC} for days 139–147 is the same as for days 159–167 while both the validation data (see Fig. 3a) and the inverse solution at the ADP location (see Fig. 3c) show a lower baroclinic signal for days 139–147 than for days 159–167. This suggests that for days 139–147 the baroclinic tide was stronger away from the ADP location than near the ADP site, as seen in Fig. 14a. So it may be misleading to attempt to assess the intensity of the internal tide in the whole area from information collected at one location.

c. Energy balance

Unless the bottom is flat, separation of a flow into barotropic and baroclinic parts is not obvious, and the

analysis and interpretation of energetics will depend to some extent on how this partitioning is formally defined. If the barotropic current \mathbf{u}_1 is defined as the depth average (Cummins and Oey 1997; Holloway 2001), and the baroclinic current as the deviation from the depth average, $\mathbf{u}_2 = \mathbf{u} - \mathbf{u}_1$, then the barotropic (depth integrated) energy equation is

$$\begin{aligned} & \frac{\partial}{\partial t} \frac{1}{2} \rho_o (g\eta^2 + H|\mathbf{u}_1|^2) \\ & = -\nabla \cdot (H\mathbf{u}_1 p_1) - \rho_o r \mathbf{u}_1 \cdot \mathbf{u}(-H) - \text{EC}. \end{aligned} \quad (31)$$

Here the field variables are time-dependent, rather than complex amplitudes, $p_1(x, y, t)$ is the depth-averaged

pressure, $p_2(x, y, z, t) = p - p_1$, and the first term on the rhs of (31) is minus the barotropic energy flux divergence; EC is the barotropic-to-baroclinic energy conversion rate:

$$\text{EC} = g \int_{-H}^0 \hat{w}^{\text{cart}} \rho \, dz, \quad (32)$$

where \hat{w}^{cart} is the vertical velocity (aligned with the vertical z axis in Cartesian coordinates) associated with the depth-averaged flow:

$$\hat{w}^{\text{cart}} = \frac{z}{H} \mathbf{u}_1 \cdot \nabla H + \frac{(z+H)}{H} \frac{\partial \eta}{\partial t}. \quad (33)$$

The baroclinic, depth-integrated energy equation is

$$\frac{\partial}{\partial t} \frac{1}{2} \int_{-H}^0 \left(\rho_o |\mathbf{u}_2|^2 + \frac{g\rho^2}{-\bar{\rho}_z} \right) dz = -\nabla \cdot \int_{-H}^0 \mathbf{u}_2 p_2 \, dz - \rho_o r \mathbf{u}_2 \cdot \mathbf{u}(-H) - \rho_o \int_{-H}^0 K_M \left| \frac{\partial \mathbf{u}}{\partial z} \right|^2 dz + \text{EC}, \quad (34)$$

where the first term on the rhs is minus the baroclinic energy flux divergence. Defining the barotropic and baroclinic flows as above has the advantage that the energy conversion term in (31) finds its exact counterpart with opposite sign in (34); adding (31) and (34) yields the energy equation for the total flow so that the total energy in the model can be accounted for.

In our analysis, we consider terms in the energy equations averaged over a tidal period. The depth-integrated barotropic energy flux is from the south to the north (consistent with Kelvin wave dynamics), on the order of 10 kW m⁻¹, with relatively smaller values over the shallows of Stonewall Bank (Fig. 18a). The barotropic flux has low day-to-day variability. The depth-integrated baroclinic energy flux is much smaller, on the order of tens of watts per meter, and experiences substantial temporal variability (Figs. 18b–d). The persistent feature for all days studied is the general direction of baroclinic energy flux from northwest to southeast, close to the direction of phase propagation. This suggests that the continental shelf break slope northwest of the studied area may be a generation site of energetic M_2 internal tide which can propagate onto the shelf. Internal tides are probably also generated due west and south of our computational domain, but the steeper slopes in these areas (Fig. 1a) reflect most of the internal tide energy toward deeper water.

The time-averaged terms in the energy equations for day 139 are shown in Figs. 19a–d and 19g. The divergence of the barotropic and baroclinic energy fluxes, and the energy conversion rate, are only a few milliwatts per meter squared, at most. This is several orders of magnitude lower than the values seen in areas of strong internal tide generation such as the Hawaiian Ridge (Merrifield and Holloway 2002). This is partly due to

the shallowness of the area. Also, while the continental shelf slope outside our domain is steep, we do not expect much conversion here compared to Hawaii since barotropic flow near the coast mostly follows bathymetric contours (Baines 1982).

If (33) is substituted into (32), the energy conversion term can be separated into two parts:

$$\text{EC} = -p_2|_{z=-H} \mathbf{u}_1 \cdot \nabla H + \frac{\partial \eta}{\partial t} \frac{g}{H} \int_{-H}^0 (z+H) \rho \, dz, \quad (35)$$

where the first term is the topographic energy conversion (e.g., see Llewellyn-Smith and Young 2001). The second term is associated with sea surface variations, and its physical interpretation is not entirely intuitive. Its appearance is the consequence of our approximate division into barotropic (depth averaged) and baroclinic flows, with the elevation and density variations attributed solely to the barotropic and baroclinic components respectively. Consideration of the classical problem of internal waves over a flat bottom with a free surface (Phillips 1966; Wunsch 1975) shows that this division is not exactly correct. For that simple analytical case the flow can be decomposed into a series of vertical modes such that the depth averages of the baroclinic pressure modes are not identically zero. Potential energy for both the barotropic (mode 0) and baroclinic (all other modes) components include contributions from both elevation and density variations. Since the modes are dynamically uncoupled [cf. (27)–(28)] there will be no barotropic to baroclinic energy conversion. In particular, the second term in (35), which is present even over a flat bottom in our approximate treatment, would not appear with a proper division into barotropic and baroclinic modes.

In regions of strong bathymetric variation (such as

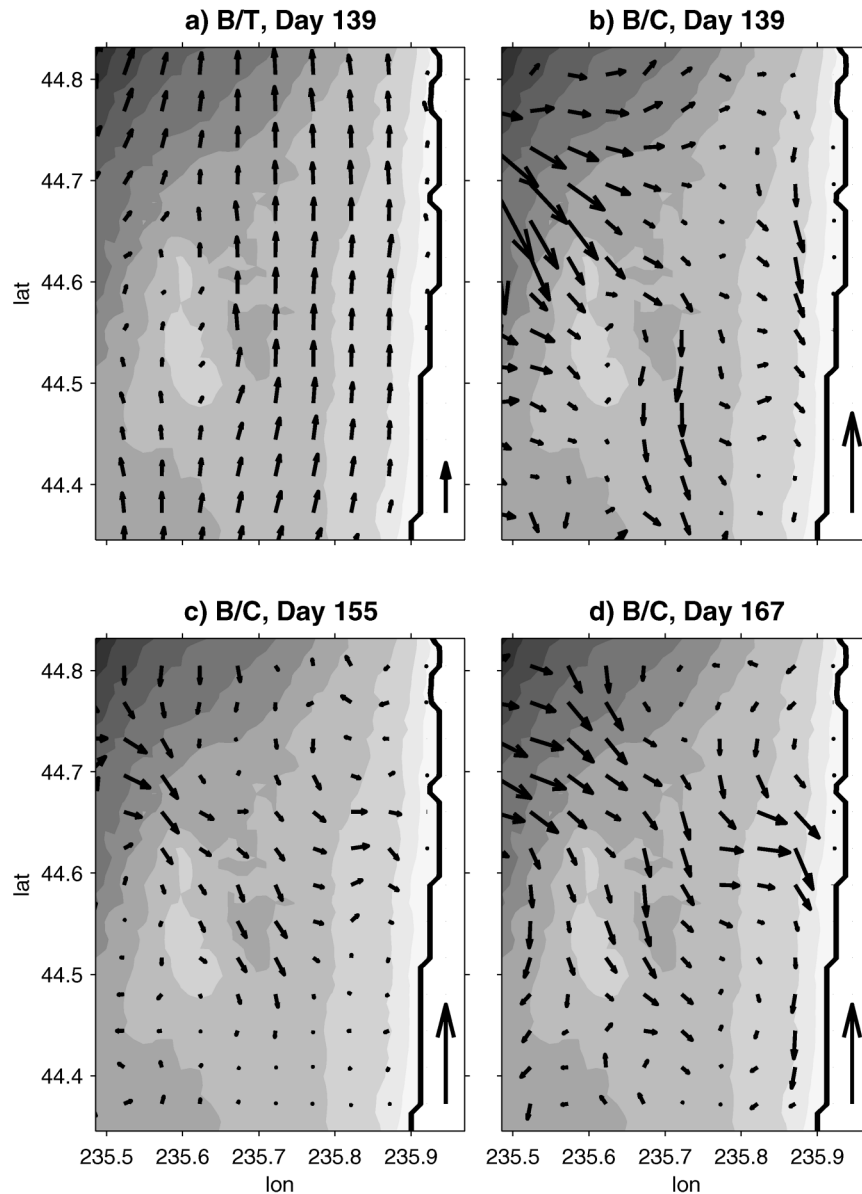


FIG. 18. Energy flux (EF), depth-integrated and averaged over the tidal period: (a) barotropic EF, for day 139, scale vector in the lower right corner is 15 kW m^{-1} ; (b)–(d) baroclinic EF, for days 139, 155, and 167, scale vector is 40 W m^{-1} . Shaded contours are bathymetry, every 20 m.

ocean ridges or the continental shelf break) the first term dominates in (35). However, within our study area topographic variation is relatively mild such that the vertical velocity at the surface is comparable with that at the bottom. Accordingly, both components of EC are of similar magnitude (cf. Figs. 19d–f and 19g–i, where the terms are shown for days 139, 155, and 167). However, they have different, distinctive patterns. For days 155 and 167, topographic conversion is positive and strongest near the western slope of Stonewall Bank. However, for day 139, the topographic EC is negative in the same area, and strong

positive conversion occurs at the north slope. Reverse topographic energy conversion, from the baroclinic to barotropic flow, is possible because the sign of the time-averaged EC depends on the phase shift between \hat{w}^{cart} and p_2 at the bottom; since most of the baroclinic wave propagates into the domain from outside, p_2 is not necessarily phase locked to \hat{w}^{cart} .

The phase difference between ρ and η determines the sign of the second term of (35), averaged over a tidal period. The plots for this term (see Figs. 19g–i) highlight the baroclinic wave propagating from the northwest and apparently reflecting from topography and the coast.

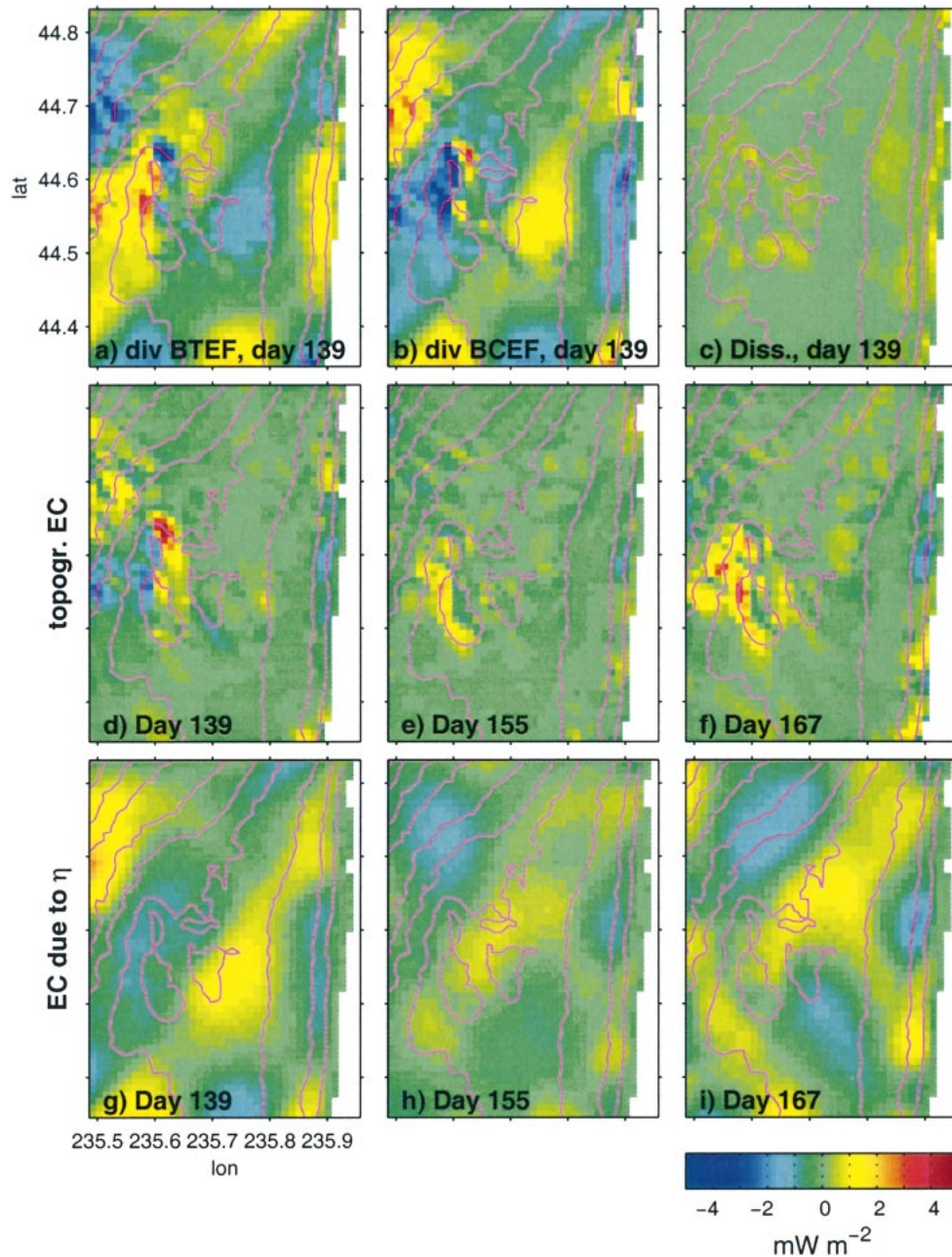


FIG. 19. Terms in the energy equations averaged over the tidal period: (a) divergence of the barotropic energy flux, day 139; (b) divergence of the baroclinic energy flux, day 139; (c) total dissipation, day 139; (d)–(f) topographic *EC* for days 139, 155, and 167, respectively; (g)–(i) second term in (35), for the same days. Contours are bathymetry every 20 m.

Terms on the rhs of (34) averaged over the whole computational area and over the tidal period are shown in Fig. 20a, for every time window. Plotted are the baroclinic energy flux into the area through the open boundary [OB EF, the spatial integral of the first term on the rhs of (34)], the dissipation [combining the second and the third terms on the rhs of (34)], the energy conversion split into the two terms [see (35)], and the

residual (numerical error) in the energy balance computation. The energy conversion terms are much smaller than the OB EF, demonstrating that most baroclinic energy comes through the open boundary with the net influx balanced by dissipation.

Integrated over the 13 km × 22 km area of the Stonewall Bank, the time-averaged energy balance (Fig. 20b) differs from that integrated over the whole computa-

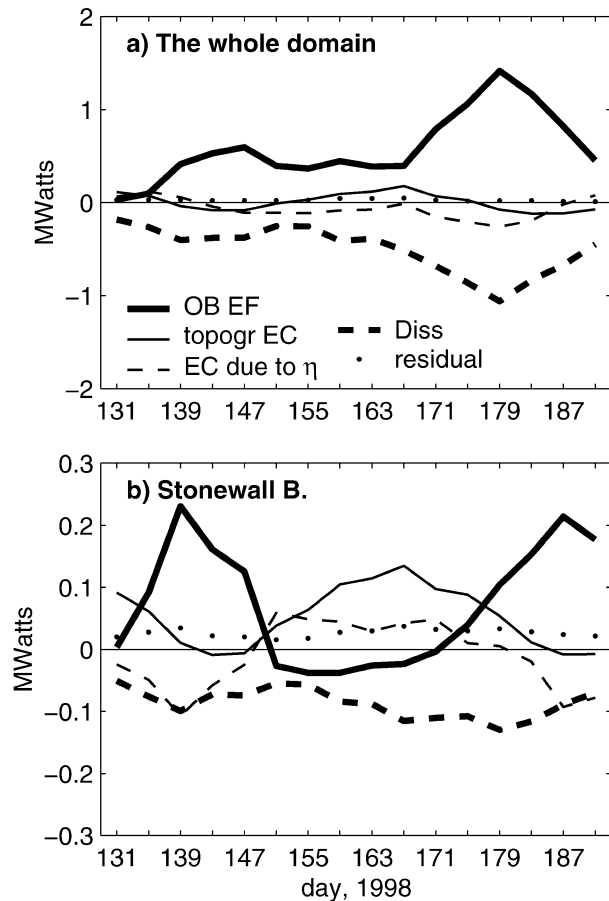


FIG. 20. Terms in the baroclinic energy equation (34) averaged over the tidal period and (a) averaged over the whole computational domain, or (b) averaged over the area of the Stonewall Bank. Case lon(HF) < 235.8°E. Dots show the residual (numerical error) in the energy balance.

tional domain (Fig. 20a). Over the bathymetric rise, topographic EC is comparable to the energy flux into the smaller area and is biased toward positive values.

8. Summary

Our study is the first attempt to model superinertial internal tide variability on the shelf using data assimilation. Since the tidal flow depends critically on hydrography and open boundary currents that are poorly known, data assimilation seems a promising approach. The representer solution and computations with synthetic data show that in principle measurements of surface currents can be used to map tidal flow at depth. For the superinertial M_2 tide, information from the surface is projected in space along wave characteristics. Application of our frequency domain inverse model to HF radar data harmonically analyzed in a series of overlapping 2-week windows provides a uniquely detailed picture of the temporal and spatial variability of internal tide on the central

Oregon shelf. Comparison with the ADP data shows that assimilation of harmonically analyzed surface current measurements from the coast-based HF radars captures temporal variability, as well as variability in the vertical, of the internal tide.

Although M_2 baroclinic flows on the mid-Oregon shelf are variable and intermittent, some features are found to be persistent. For instance, during the study period the phase and energy of the baroclinic tide propagated in the same direction, from north-west to south-east. At the surface, baroclinic surface tidal currents (deviations from the depth-averaged current) can be 10 cm s^{-1} , 2 times as large as the depth-averaged current. Allowing for spring-neap variations and more rapid changes in the internal tide in response to changes in ocean conditions, peak baroclinic velocities in the semidiurnal band could be easily 2 times this. The analysis of the energy balance demonstrates that most baroclinic signal comes into the study area via the open boundaries, and so a proper specification of OB conditions is critical.

The generalized inverse method offers great flexibility in the formulation of the DA problem. For instance, weak constraint DA (where dynamical equations are assumed to contain error) is used here to find an appropriate covariance for the open boundary condition by nesting. Then, strong constraint DA (with exact dynamics) is implemented to estimate open boundary baroclinic currents using the measured surface velocities. One of the values of the representer method is that a dynamically consistent prior error covariance for the model solution can be computed, rather than guessed.

From the perspective of data assimilation, this study raises important questions concerning open boundary conditions. Since DA can provide corrections to the boundary values, there is an opportunity to control radiation by the choice of open boundary condition covariance, consistent with the modeled dynamics. In our study, nesting is adopted as the simplest way to obtain a covariance that should improve wave radiation. Computations with synthetic data show that such a covariance yields an inverse solution superior to that obtained with the best covariance that we could make up without nesting.

Although the dynamical model can in principle be used at any frequency, limited accuracy of HF radar data poses difficulties in estimating major tidal constituents, other than M_2 , from short time series. Off Oregon, S_2 is the second largest tidal constituent with barotropic tidal magnitude about 40% of M_2 and a period (12 h) very close to that of M_2 (12.4 h). The K_1 diurnal tidal frequency is subinertial so that free internal waves are not allowed. Tidal signal in HF radar surface data at the K_1 frequency is probably also contaminated by wind-induced currents due to a diurnal sea breeze (Erofeeva et al. 2003). Tidal studies in the

diurnal band should probably make allowance for this additional forcing.

The internal tide inverse solution is sensitive to the background stratification and velocity fields. These could possibly be accounted for within the context of our simple model by means of weak constraint DA. Then careful interpretation of the dynamical error would be necessary to avoid confusion in the analysis of the momentum and energy budget. A better approach to improvement of the internal tide prediction would be to improve the model by allowing for a horizontally nonuniform background stratification. Such modifications to the model should also include the background current to geostrophically balance background horizontal pressure gradients. Linearization with respect to this new basic state would yield dynamical terms that describe advection of tidal perturbations by the background currents. Such a model could be useful for addressing the question of the effect of subinertial currents on internal tides, which remains open. However, as soon as advective terms are included, the task of finding OB conditions that guarantee a well-posed formulation becomes nontrivial (see Bennett 1992, chapter 9). Our example, with a simpler model, shows that finding an appropriate set of OB conditions does not solve all the problems at the OB if the task is to restore OB inputs by data assimilation. Attention should thus also be given to the OB error covariance.

To provide a more realistic background state, a coastal DA system should eventually couple a tidal inverse model and a reliable model of wind-driven circulation. This would provide a more realistic background state for the tidal model, as well as a framework for studying tidal effects on subtidal flows. Our success with inversion of the HF radar data to describe three-dimensional superinertial tidal flow, using simple dynamics, is encouraging for the optimal use of surface current measurements with more advanced models.

Acknowledgments. The research was supported by the Office of Naval Research (ONR) Ocean Modeling and Prediction Program under Grant N00014-98-1-0043, and the National Science Foundation under Grant OCE-9819518.

APPENDIX A

Model Solver

The solution strategy for (11)–(14) is to operate with parts of the discretized model operator as sparse matrices, eliminate \mathbf{u} and \mathbf{w} , then solve the reduced size matrix problem by direct factorization. Elimination of variables is accomplished by a series of

small size matrix inversions locally in each vertical column of the computational grid.

To accomplish this, we approximate (11) by

$$\mathbf{u} = \mathbf{C}\mathbf{f}_u - \Psi_u \boldsymbol{\rho}^+, \quad (\text{A1})$$

where $\Psi_u = \mathbf{C}(\mathbf{G}|\mathbf{P})$,

$$\boldsymbol{\rho}^+ = \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\rho} \end{pmatrix}, \quad (\text{A2})$$

and $\mathbf{C} \approx \boldsymbol{\Omega}^{-1}$. For the original continuous equations the operator corresponding to $\boldsymbol{\Omega}$ in (11) and its inverse are local in the horizontal coordinates x and y . So, a continuous analog of (A1) is exactly equivalent to (2). However, discretized on the staggered C grid the Coriolis term couples the u and v components of \mathbf{u} at neighboring horizontal locations. This destroys the block structure of $\boldsymbol{\Omega}$ and makes $\boldsymbol{\Omega}^{-1}$ a full matrix. To avoid this we directly discretized the horizontally local inverse operator: $\{u, v\}$ is first transformed into $v^\pm = (u \pm iv)/2$ (Lynch et al. 1992; Muccino et al. 1997); the momentum equations are decoupled with respect to v^+ and v^- , so inversion for v^\pm is performed locally at each u and v location of the C grid; finally, backward transformation is made to get \mathbf{u} . All these steps are fixed in the matrix \mathbf{C} .

From (13), we obtain, taking (A1) into account,

$$\mathbf{w} = \mathbf{W}^{-1}\mathbf{f}_w - \mathbf{W}^{-1}\mathbf{D}\mathbf{C}\mathbf{f}_u + \mathbf{W}^{-1}\mathbf{D}\Psi_u \boldsymbol{\rho}^+. \quad (\text{A3})$$

The resulting system of equations for $\boldsymbol{\rho}^+$ is obtained by substitution of (A1) and (A3) into (12) and (14) and then combining them to give

$$\mathbf{A}\boldsymbol{\rho}^+ = \mathbf{F}_u \mathbf{f}_u + \mathbf{F}_w \mathbf{f}_w + \mathbf{f}_{\rho^+}, \quad (\text{A4})$$

where

$$\mathbf{A} = \mathbf{B}_2 + \mathbf{D}_2 \Psi_u, \quad \mathbf{B}_2 = \begin{pmatrix} i\omega \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix},$$

$$\mathbf{D}_2 = -\left(\frac{\mathbf{D}_a}{\mathbf{D}_\rho - \mathbf{Z}_\rho \mathbf{W}^{-1}\mathbf{D}} \right),$$

$$\mathbf{F}_u = \mathbf{D}_2 \mathbf{C}, \quad \mathbf{F}_w = \begin{pmatrix} 0 \\ -\mathbf{Z}_\rho \mathbf{W}^{-1} \end{pmatrix},$$

$$\mathbf{f}_{\rho^+} = \begin{pmatrix} \mathbf{f}_a \\ \mathbf{f}_\rho \end{pmatrix}.$$

The matrix \mathbf{B}_2 is partitioned into blocks such that standard block-matrix multiplication can be performed for \mathbf{B}_2 and $\boldsymbol{\rho}^+$ (A2); \mathbf{F}_w has zeros in the rows corresponding to the depth-integrated continuity equation. The solution to (A4) is found by direct LU factorization of \mathbf{A} using a standard routine for banded matrices (Anderson et al. 1992). Then multiplication by both \mathbf{A}^{-1} and its complex conjugate, matrix transpose $(\mathbf{A}')^{-1}$ can be accomplished rapidly by solving the appropriate triangular systems. Last, the solution to (11)–(14) is

$$\boldsymbol{\nu} \equiv \begin{pmatrix} \mathbf{u} \\ \mathbf{w} \\ \boldsymbol{\rho}^+ \end{pmatrix} = \begin{pmatrix} \mathbf{C} - \Psi_u \mathbf{A}^{-1} \mathbf{F}_u & -\Psi_u \mathbf{A}^{-1} \mathbf{F}_w & -\Psi_u \mathbf{A}^{-1} \\ \Psi_w \mathbf{A}^{-1} \mathbf{F}_u - \mathbf{W}^{-1} \mathbf{D} \mathbf{C} & \Psi_w \mathbf{A}^{-1} \mathbf{F}_w + \mathbf{W}^{-1} & \Psi_w \mathbf{A}^{-1} \\ \mathbf{A}^{-1} \mathbf{F}_u & \mathbf{A}^{-1} \mathbf{F}_w & \mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{f}_u \\ \mathbf{f}_w \\ \mathbf{f}_{\rho^+} \end{pmatrix}, \quad (\text{A5})$$

where $\Psi_w = \mathbf{W}^{-1} \mathbf{D} \Psi_u$, and a 3×3 block-matrix is \mathbf{S}^{-1} . The size of the numerical problem is limited by the storage required for the \mathbf{LU} factors of the banded matrix \mathbf{A} . The advantage of the solution approach described above is that once matrix \mathbf{A} is factored, a large number of representers [see (21)] can be rapidly computed as a series of matrix-vector and vector-vector operations. Also, the adjoint solver $(\mathbf{S}')^{-1}$ is easily obtained as a complex conjugate, matrix transpose of the block matrix in (A5).

APPENDIX B

Reduced Basis Approach

Let $\hat{\mathbf{I}}'_k$ be a data functional corresponding to the assumed observation from the reduced basis, $k = 1, \dots, K_1$. In our study $\hat{\mathbf{I}}'_k \mathbf{u}$ sample u and v at 34 points on the surface, so $K_1 = 68$. Representers of the reduced basis are computed as $\hat{\mathbf{r}}_k = \mathbf{S}^{-1} \mathbf{cov}(\mathbf{S}')^{-1} \hat{\mathbf{I}}'_k$. The inverse solution is sought in the form

$$\hat{\boldsymbol{\nu}}_{\text{inv}} = \boldsymbol{\nu}_0 + \sum_{k=1}^{K_1} \hat{b}_k \hat{\mathbf{r}}_k. \quad (\text{B1})$$

The task is to find the vector $\hat{\mathbf{b}}$ of representer coefficients \hat{b}_k that yields a minimum of the original cost function (19), which is

$$J(\hat{\boldsymbol{\nu}}_{\text{inv}}) = \hat{\mathbf{b}}' \hat{\mathbf{R}} \hat{\mathbf{b}} + (\mathbf{d} - \mathbf{L}' \boldsymbol{\nu}_0 - \mathbf{P} \hat{\mathbf{b}})' \times \mathbf{cov}_d^{-1} (\mathbf{d} - \mathbf{L}' \boldsymbol{\nu}_0 - \mathbf{P} \hat{\mathbf{b}}). \quad (\text{B2})$$

Here $\hat{\mathbf{R}}$ is the representer matrix of the reduced basis, of size $K_1 \times K_1$, with elements $\hat{R}_{kn} = \hat{\mathbf{I}}'_k \hat{\mathbf{r}}_n$; \mathbf{P} is of size $K \times K_1$ with elements $P_{kn} = \mathbf{I}'_k \hat{\mathbf{r}}_n$, where \mathbf{I}'_k are the discrete data functionals for actual observations [see (17)], and $\hat{\mathbf{r}}_n$ is the representer from the reduced basis. Optimal representer coefficients are found as in Egbert and Erofeeva (2002):

$$\hat{\mathbf{b}} = \hat{\mathbf{R}}^{-1/2} \mathbf{Q} \boldsymbol{\Lambda} (\mathbf{I} + \boldsymbol{\Lambda}^2)^{-1} \mathbf{V}' \mathbf{cov}_d^{-1/2} (\mathbf{d} - \mathbf{L}' \boldsymbol{\nu}_0), \quad (\text{B3})$$

where the matrices \mathbf{V} , $\boldsymbol{\Lambda}$ (diagonal), and \mathbf{Q} are the result of singular value decomposition: $\mathbf{V} \boldsymbol{\Lambda} \mathbf{Q}' = \text{s.v.d.} \{ \mathbf{cov}_d^{-1/2} \mathbf{P} \hat{\mathbf{R}}^{-1/2} \}$.

REFERENCES

- Anderson, E., 1992: *LAPACK User's Guide*. Philadelphia, 235 pp.
- Baines, P. G., 1982: On internal tide generation models. *Deep-Sea Res.*, **29**, 307–338.
- Bennett, A. F., 1992: *Inverse Methods in Physical Oceanography*. Cambridge University Press, 346 pp.
- , 2002: *Inverse Modeling of the Ocean and Atmosphere*. Cambridge University Press, 234 pp.
- Blumberg, A. F., and G. L. Mellor, 1987: A description of a three-dimensional coastal ocean circulation model. *Three-Dimensional Coastal Ocean Models*, N. Heaps, Ed., Coastal and Estuarine Science Series, Vol. 4, Amer. Geophys. Union, 1–16.
- Bogden, P. S., 2001: The impact of model–error correlation on regional data assimilative models and their observational arrays. *J. Mar. Res.*, **59**, 831–857.
- Cummins, P. F., and L. Y. Oey, 1997: Simulation of barotropic and baroclinic tides off northern British Columbia. *J. Phys. Oceanogr.*, **27**, 762–781.
- Egbert, G. D., and S. Y. Erofeeva, 2002: Efficient inverse modeling of barotropic ocean tides. *J. Atmos. Oceanic Technol.*, **19**, 183–204.
- , A. F. Bennett, and M. G. G. Foreman, 1994: TOPEX/POSEIDON tides estimated using a global inverse model. *J. Geophys. Res.*, **99** (C12), 24 821–24 852.
- Erofeeva, S. Y., G. D. Egbert, and P. M. Kosro, 2003: Tidal currents on the central Oregon shelf: Models, data, and assimilation. *J. Geophys. Res.*, **108**, 3148, doi:10.1029/2002JC001615.
- Foreman, M. G. G., L. M. Delves, I. Barrodale, and R. F. Henry, 1980: On the use of the Proudman–Heaps tidal theorem. *Geophys. J. Roy. Astron. Soc.*, **63**, 467–478.
- Gill, A. E., 1982: *Atmosphere–Ocean Dynamics*. Academic Press, 662 pp.
- Hayes, S. P., and D. Halpern, 1976: Observations of internal waves and coastal upwelling off the Oregon coast. *J. Mar. Res.*, **34**, 247–267.
- Holloway, P. E., 2001: A regional model of the semi-diurnal internal tide on the Australian North West Shelf. *J. Geophys. Res.*, **106**, 19 625–19 638.
- Kivman, G. A., 1997: Weak constraint data assimilation for tides in the Arctic Ocean. *Progress in Oceanography*, Vol. 40, Pergamon, 179–196.
- Kosro, P. M., J. A. Barth, and P. T. Strub, 1997: The coastal jet: observations of surface currents over the Oregon continental shelf from HF radar. *Oceanography*, **10**, 53–56.
- Kurapov, A. L., G. D. Egbert, R. N. Miller, and J. S. Allen, 2002: Data assimilation in a baroclinic coastal ocean model: ensemble statistics and comparison of methods. *Mon. Wea. Rev.*, **130**, 1009–1025.
- Llewellyn Smith, S. G., and E. R. Young, 2001: Conversion of the barotropic tide. *J. Phys. Oceanogr.*, **32**, 1554–1566.
- Lynch, D. R., F. E. Werner, D. A. Greenberg, and J. W. Loder, 1992: Diagnostic model for baroclinic, wind-driven and tidal circulation in shallow seas. *Cont. Shelf Res.*, **12**, 37–64.
- Mellor, G. L., 1998: Users' guide for a three-dimensional, primitive equation, numerical ocean model. Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ. [Available online at <http://www.aos.princeton.edu/WWWPUBLIC/htdocs.pom/>.]
- Merrifield, M. A., and P. E. Holloway, 2002: Model estimates of M_2 internal tide energetics at the Hawaiian Ridge. *J. Geophys. Res.*, **107**, 3179, doi:10.1029/2001JC000996.
- Muccino, J. C., W. G. Gray, and M. G. G. Foreman, 1997: Calculation of vertical velocity in three-dimensional, shallow water equation, finite element models. *Int. J. Numer. Meth. Fluids*, **25**, 779–802.
- Munk, W., F. Snodgrass, and M. Wimbush, 1970: Tides off shore:

- Transition from California coastal to deep-sea waters. *Geophys. Fluid Dyn.*, **1**, 161–235.
- Oke, P. R., J. S. Allen, R. N. Miller, G. D. Egbert, and P. M. Kosro, 2002: Assimilation of surface velocity data into a primitive equation coastal ocean model. *J. Geophys. Res.*, **107** (C9), 3122, doi: 10.1029/2000JC000511.
- Oliger, J., and A. Sundström, 1978: Theoretical and practical aspects of some initial boundary value problems in fluid dynamics. *SIAM J. Appl. Math.*, **35**, 419–446.
- Pacanowski, P. C., and S. C. H. Philander, 1981: Parameterization of vertical mixing in numerical models of tropical oceans. *J. Phys. Oceanogr.*, **21**, 1186–1201.
- Petruncio, E. T., L. K. Rosenfeld, and J. D. Paduan, 1998: Observations of the internal tide in Monterey Canyon. *J. Phys. Oceanogr.*, **28**, 1873–1903.
- Phillips, O. M., 1996: *Dynamics of the Upper Ocean*. Cambridge University Press, 261 pp.
- Sirkes, Z., and E. Tziperman, 1997: Finite difference of adjoint or adjoint of finite difference? *Mon. Wea. Rev.*, **125**, 3373–3378.
- Smith, R. L., A. Huyer, and J. Fleischbein, 2000: The coastal ocean off Oregon from 1961 to 2000: Is there evidence of climate change or only Los Ninos? *Progress in Oceanography*, Vol. 49, Pergamon, 63–93.
- Torgimson, G. M., and B. M. Hickey, 1979: Barotropic and baroclinic tides over the continental slope and shelf off Oregon. *J. Phys. Oceanogr.*, **9**, 945–961.
- Wijesekera, H. W., J. S. Allen, and P. A. Newberger, 2003: Modeling study of turbulent mixing over the continental shelf: Comparison of turbulent closure schemes. *J. Geophys. Res.*, **108**, 3103, doi: 10.1029/2001JC001234.
- Wunsch, C. H., 1975: Internal tides in the ocean. *Rev. Geophys. Space Phys.*, **13**, 167–182.