Estimating Open-Ocean Barotropic Tidal Dissipation: The Hawaiian Ridge

EDWARD D. ZARON AND GARY D. EGBERT

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

ABSTRACT

The generalized inverse of a regional model is used to estimate barotropic tidal dissipation along the Hawaiian Ridge. The model, based on the linear shallow-water equations, incorporates parameterizations for the dissipation of energy via friction in the bottom boundary layer and form drag due to internal waves generated at topographic slopes. Sea surface height data from 364 orbit cycles of the Ocean Topography Experiment (TOPEX)/Poseidon satellite mission are used to perform inversions at eight diurnal and semidiurnal tidal frequencies. It is estimated that the barotropic M_2 tide loses energy at a rate of 19 GW, of which 88% is lost within 250 km of the ridge, presumably via conversion to the internal or baroclinic tide. Uncertainty in the assumed model error and wave drag in the forward model suggest that M_2 dissipation values from 18 to 25 GW are consistent with the altimetric observations. Other barotropic tidal constituents are estimated to lose a total of 5.7 GW. The spatial distribution of barotropic dissipation along the ridge is similar to that inferred from three-dimensional primitive equation models, and it is largely insensitive to details of assumed model and data errors. Dissipation at semidiurnal frequencies is most intense at the French Frigate Shoals with lesser, but significant, contributions at other sites. Diurnal tidal dissipation is concentrated to the east of the French Frigate Shoals, at the Gardner Pinnacles. Further work with three-dimensional models will be necessary to determine the fate of the energy that is removed from the barotropic tide.

1. Introduction

Prior to the era of satellite altimetry, it was believed that the vast majority of tidal dissipation occurred in shallow seas and coastal areas (Jeffreys 1920). A wide variety of analyses of altimetry data now suggest that 70%–75% of the dissipation does indeed occur in shallow areas; however, that leaves 25%–30% to take place in the deep ocean, primarily in association with major bathymetric features (Egbert and Ray 2001). This result has had broad significance because the deep-ocean tidal dissipation may provide as much as one-half of the mechanical energy necessary to maintain the global thermohaline circulation (Munk and Wunsch 1998).

The Hawaii Ocean Mixing Experiment (HOME) has sought to elucidate the pathways that link the barotropic surface tide to small-scale turbulence and mixing (Rudnick et al. 2003). The observational components of the experiment include measurements along the ridge at hot spots of active mixing and baroclinic conversion

(the so-called Nearfield), as well as measurements at some distance from the ridge (the Farfield). By developing a better understanding of the linkage between the Nearfield sources of baroclinic energy and the Farfield propagation and dissipation of this energy, an energetically consistent understanding of deep-ocean mixing may be obtained.

The overall characterization of the tidal energy budget surrounding the ridge may conceptually be divided into three components: the energy flux convergence of the barotropic surface tide (the barotropic tidal dissipation, hereinafter), the energy flux divergence of the baroclinic tide, and turbulent dissipation. The barotropic tidal dissipation is primarily a consequence of two processes: the nonlinear transfer of energy to turbulence in the bottom boundary layer and the linear transfer of energy to internal waves (baroclinic conversion). At open-ocean sites with significant cross-isobath flows, such as are found along the Hawaiian Ridge, the process of baroclinic conversion is dominant, and turbulence plays a lesser role. The energy flux divergence of the baroclinic tide is also the consequence of the same processes, baroclinic conversion and turbulence; however, the nonlinear cascade of energy to smaller scales

Corresponding author address: Edward D. Zaron, 104 Ocean Admin. Bldg, Oregon State University, Corvallis, OR 97331.
E-mail: ezaron@coas.oregonstate.edu

may be local in space (where shear-driven turbulence occurs) or nonlocal (where parametric subharmonic instability and induced diffusion take place).

While the observational program of HOME has made direct observations of the localized turbulent dissipation and internal waves (Rudnick et al. 2003), direct measurements of the barotropic tidal dissipation are exceedingly difficult because of the vast length scale of the barotropic tide. In the present study we use satellite data, together with a model based on the Laplace tidal equations, to refine estimates of barotropic tidal dissipation around the Hawaiian Ridge and relate the pattern of dissipation to other modeling and observational studies in HOME.

As mentioned above, the dissipation of barotropic surface tides may take place via bottom drag or baroclinic conversion, which is a type of form drag. The former results in the direct production of turbulence, while the latter results in internal waves that either radiate away from the site of genesis or dissipate locally (St. Laurent and Garrett 2002). Linear theories of baroclinic conversion predict that barotropic tidal flow over topography will lose energy at a rate given by

$$C = \rho_0 N_B U^2 \sqrt{1 - \frac{f^2}{\omega^2}} T, \quad \text{(W)} \qquad (1)$$

where ρ_0 is the density of seawater, N_B is the buoyancy frequency at the ocean bottom, U is the magnitude of the barotropic transport ($m^2 s^{-1}$), f is the Coriolis parameter, ω is the tidal frequency, and T (having units of area) is related to the power spectrum of the bottom topography and the orientation of the barotropic flow (Llewellyn Smith and Young 2002). Perturbation techniques have been applied to study the tidal flow in the limit of small-amplitude, subcritical topography (Bell 1975; Balmforth et al. 2002). There have also been extensions of these analyses to finite-amplitude, supercritical topography (St. Laurent et al. 2003; Petrelis et al. 2003), which generally amount to first-order corrections to the linear theory and less than order-of-magnitude changes in computed conversion rates. Three-dimensional models, which explicitly model the internal waves, reveal baroclinic conversion at topographic sites that is at least in qualitative agreement with the analytical predictions (Kang et al. 2000; Merrifield and Holloway 2002; Khatiwala 2003; Simmons et al. 2004; Di Lorenzo et al. 2006).

The intensification of barotropic tidal energy dissipation near topographic features in the open ocean, which has been inferred from data-assimilative tidal models, is broadly consistent with the theory of baroclinic conversion (Egbert and Ray 2001). Because these models have been run at relatively coarse resolution, quantitative comparisons with theory or measurement at specific topographic features is problematic. Here, we look at the Hawaiian Ridge in some detail to see if the dissipation can be localized and constraints on its total value refined. Although we emphasize the dominant constituent, M_2, dissipation at seven other diurnal and semidiurnal frequencies is also estimated. Our central tool shall be the generalized inverse of a shallow-water model, the Oregon State University tidal inversion software (OTIS; Egbert and Erofeeva 2002).

This paper is organized as follows. Section 2 summarizes the unique aspects of regional open-ocean tidal modeling and briefly reviews the principles of generalized inversion. In section 3 we present dissipation estimates in the vicinity of the Hawaiian Ridge inferred from a large set of altimetric data. Sensitivity to various assumptions is examined in section 4. The last section briefly summarizes our quantitative results and suggests further steps for research.

2. Methods

The problem of regional, open-ocean, barotropic tidal inversion is to find that set of tidal fields, namely, surface elevation ζ and volume transport $\mathbf{U}$, that are consistent with both a dynamical model and a set of data. Here, the dynamical model is based on the Laplace tidal equations, together with boundary conditions, while the data are altimeter measurements of surface elevation at the tidal frequencies. The information from the model and data are combined to produce an optimal estimate of the tidal fields using the generalized inverse methods developed previously by Egbert et al. (1994) and Egbert and Erofeeva (2002). Here, we summarize the method and provide details specific to our application.

a. Dynamical model

We assume that the barotropic tide is governed approximately by the Laplace tidal equations, extended to include the effects of self-attraction and loading (SAL), and frictional or dissipative stress:

$$\frac{\partial \mathbf{U}}{\partial t} + f\mathbf{k} \times \mathbf{U} = -gH\boldsymbol{\nabla}(\zeta - \zeta_{\text{EQ}} - \zeta_{\text{SAL}}) - \mathcal{F} \qquad (2)$$

and

$$\frac{\partial \zeta}{\partial t} = -\boldsymbol{\nabla} \cdot \mathbf{U}. \qquad (3)$$

The symbols are used here in conformance with Egbert and Ray (2001); f is the Coriolis parameter, g is gravi-

tational acceleration, H is the bottom depth, ζ_{EQ} is the equilibrium tidal potential corrected for the body tide, and ζ_{SAL} is the correction to the tidal potential that accounts for the effects of solid earth loading and ocean self-attraction. The normal component of the transport, $\mathbf{U} \cdot \mathbf{n}$, is specified on open boundaries, with boundary data supplied by a larger-scale data-assimilative model, and no-normal-flow is specified on coastlines. The divergence and gradient operators are two-dimensional on the surface of the sphere. The body force associated with dissipative effects, $\mathcal{F}$, is represented as a sum of two terms,

$$\mathcal{F} = \mathcal{F}_B + \mathcal{F}_{BC}, \qquad (4)$$

where $\mathcal{F}_B$ is the vertically integrated turbulent stress caused by the bottom boundary layer, and $\mathcal{F}_{BC}$ is the form drag caused by the baroclinic flow. These terms are discussed in more detail below.

To facilitate the solution of Eqs. (2) and (3) in the frequency domain, a linear drag law is used to parameterize the bottom boundary layer stress, namely,

$$\mathcal{F}_B = \frac{c_d v_f}{H} \mathbf{U}, \qquad (5)$$

with nondimensional drag coefficient, $c_d = 0.003$, and friction velocity $v_f = 0.1 \text{ m s}^{-1}$. The standard quadratic drag law can be linearized around a tidal solution to improve the accuracy of the model in shallow seas and coastal areas (Egbert and Erofeeva 2002); however, solutions to the tidal equations in our regional domain are insensitive to the form of the bottom drag. Thus, we use Eq. (5) for its simplicity. In all of our solutions, the bottom drag dissipates about 0.5 GW (summed over all constituents) along the Hawaiian Ridge, a negligible fraction of the total.

The form drag is also parameterized by a linear drag law in which the drag coefficient is explicitly proportional to the square of the topographic slope,

$$\mathcal{F}_{BC} = c_{BC} \frac{N_B (\nabla H)^2}{\kappa H} \mathbf{U}, \qquad (6)$$

where c_{BC} is a nondimensional constant, κ is the wavenumber of the first internal mode at the M_2 tidal frequency, N is the buoyancy frequency, and N_B is its value at the ocean bottom; that is, $N_B = N(-H)$. This expression for $\mathcal{F}_{BC}$ is based on a Wentzel–Kramers–Brillouin approximation for the energy conversion in the linear weak-topography approximation (Llewellyn Smith and Young 2002). The $\kappa^{-1}(\nabla H)^2$ term in Eq. (6) approximates an expression involving the power spectrum of the topography. To apply Eq. (6) we assume that $N(z) = N_0 e^{z/L}$, with $N_0 = 5.24 \times 10^{-3} \text{ s}^{-1}$ and $L =$ 1300 m. This model of drag due to internal waves contains the free parameter, c_{BC}; more is said about the value of this constant below.

Equation (6) should be regarded as a scale relation for the dissipation and not as a definitive representation of a Reynolds stress. This parameterization is similar in form to those used in previous studies. Jayne and St. Laurent (2001) used $(1/2)\kappa \langle h^2 \rangle N_B \mathbf{U}/H$, in which wavenumber κ is a tunable parameter and $\langle h^2 \rangle$ is estimated from subgrid-scale topographic roughness. Two other parameterizations of the internal wave drag were treated by Egbert et al. (2004). In the first scheme the ocean bottom is approximated as a set of rectangular prisms, each of which independently scatters the barotropic tide into internal waves (Gustafson 2001; Sjöberg and Stigebrandt 1992). This approximation is known to produce resolution-dependent drag (St. Laurent et al. 2003), but it results in an expression dimensionally similar to Eq. (6). The second was based on the computation of tidal conversion due to barotropic cross-isobath flow under a small-amplitude approximation. The result, which provides a good fit to the observations without any free parameters, expresses the internal wave drag as a convolution over both space and time and, hence, requires some additional simplification for computational practicality. Egbert et al. (2004) found that, after tuning, all schemes yielded similar improvements in solution accuracy; hence, we chose to retain Eq. (6) for its simplicity. Indeed, as mentioned in the introduction, theoretically derived estimates of baroclinic conversion differ by $O(1)$ factors depending on the assumed water depth, bottom slope, stratification, and boundary conditions (e.g., Llewellyn Smith and Young 2002; St. Laurent et al. 2003).

b. Performance of the forward model

The data-assimilation procedure requires that we estimate errors in the model. Therefore, it is worthwhile to compare the solution of the forward model with the altimeter data. That is, how well does the model actually predict the altimetric observations? In particular, this comparison permits one to assess the impact of the baroclinic conversion parameterization on the model.

Note that the forward model already indirectly incorporates tidal elevation measurements via the boundary conditions, which are taken from a data-assimilative tidal model of the North Pacific (PAC6). The PAC6 model is a ⅙°-resolution, basin-scale refinement of the ½°-resolution, global, data-assimilative model TPXO.5 (Egbert and Ray 2003). Both the TPXO.5 and PAC6 solutions predict about 25 GW (1 GW = 10^9 W) of barotropic dissipation, integrated over the rectangular domain of the regional model. Note that this value does
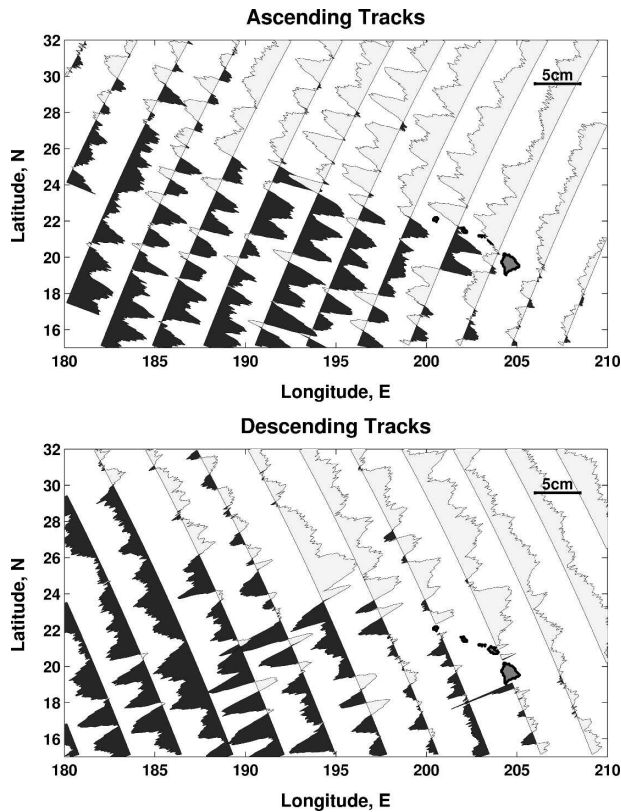
Ascending Tracks

Descending Tracks

FIG. 1. The panels show the difference between the T/P observations and the prior solution at the M_2 frequency. The prior solution solves Eqs. (2) and (3) exactly, with boundary conditions on the volume transport taken from a larger-scale data-assimilative model (PAC6, described in the text). The real part of the complex tidal amplitude is shown. Comparison with Fig. 7 shows that the data-assimilative solution makes a very modest correction to the tidal elevation field.

not contradict the estimated 18 GW of barotropic dissipation that was published previously (Egbert and Ray 2001) for a smaller patch of ocean around the Hawaiian Ridge (see Fig. 4).

Figure 1 illustrates the discrepancy between the model solution and harmonically analyzed complex altimetric data along both ascending and descending tracks. These residuals make evident several properties of the model and the data. First, the data contain small-scale noise that is uncorrelated along the altimeter tracks. This noise has three sources: 1) there is instrumental noise in the raw altimetric time series; 2) there is noise caused by aliasing of the tidal signal and spectral leakage; and 3) there is noise caused by unknown variations in the phase of the tidal signal. Taken together, these sources generate noise with an amplitude of roughly 0.4 cm. Second, the figure illustrates the surface expression of the internal tide, evident as the 150-

km wavelength signal along the tracks with amplitude generally near 3 cm (Ray and Mitchum 1996).

Last, the figure shows the larger-scale discrepancy between the model and data, which consists of a systematic offset of perhaps 2 cm between the northeast and southwest sides of the ridge. Because the shallow-water equations contain no physics to generate the signal associated with the internal tide, the impact of the data assimilation will be to adjust this large-scale offset across the ridge.

In the model–data comparison shown in Fig. 1, the root-mean-square (rms) misfit of the model and the observations is 1.36 cm. For comparison, the rms M_2 tidal signal in the altimeter data is 13.4 cm. Thus, this model explains about 99.0% of the M_2 tidal variance. When the explained variance is used as a measure of success, it is apparent that the open-ocean tide is well described by Eqs. (2) and (3).

Because the performance of the prior model essentially reflects the accuracy of the boundary conditions (obtained from the PAC6 solution), the actual value of c_{BC} has only a modest impact on the predictions of the regional model. The value of c_{BC} used above was chosen to yield the same dissipation as the PAC6 and TPXO.5 solutions in this domain (the numerical value, $c_{BC} = 1.51$, is of no particular significance). When c_{BC} is set to zero, that is, when no baroclinic conversion occurs, the model still explains 98.5% of the M_2 variance. A visual comparison of the predicted and measured surface elevation (not shown) is virtually indistinguishable from that shown in Fig. 1.

In contrast to what has been found with the global models (e.g., Jayne and St. Laurent 2001), we find that it is not possible to constrain the value of c_{BC} by minimizing the misfit of the prior model in a highly accurate regional model such as this. Based on this metric (i.e., the squared misfit of the modeled and observed tidal elevations), the optimal value for c_{BC} is about 2.4; however, this yields approximately 32 GW of barotropic dissipation via baroclinic conversion, a value that is unsupported by any other evidence and is, indeed, contradicted by our results discussed below.

Because the numerical value of c_{BC} is a free parameter, and its role in improving the model is quite negligible, we chose to set c_{BC} to zero in the prior model. An estimate of the internal wave drag based on Eq. (6) is used only to scale the model error, as discussed below. The benefit of this approach is that the barotropic dissipation estimates presented in section 3 are not biased to high values by the prior model. Essentially all of the barotropic dissipation is a consequence of the data–assimilation; it does not arise from a particular choice of c_{BC} in the prior model.

c. Data assimilation

The data-assimilation method is based on minimizing a quadratic penalty functional,

$$J[\mathbf{u}, \sigma] = (\mathbf{Lu} - \mathbf{d})^{\dagger} \Sigma_e^{-1} (\mathbf{Lu} - \mathbf{d})$$
$$+ \sigma(\mathbf{Su} - \mathbf{f})^{\dagger} \Sigma_f^{-1} (\mathbf{Su} - \mathbf{f}), \qquad (7)$$

which is the squared weighted sum of data and model misfits. As in Egbert and Erofeeva (2002), the tidal state, model, and measurement operators are represented with a terse notation. Specifically, $\mathbf{u}$ represents the tidal state ($\mathbf{U}$ and ζ fields), $\mathbf{S}$ denotes the dynamical operators [Eqs. (2) and (3), together with boundary conditions], $\mathbf{f}$ represents the model inhomogeneities (astronomical forcing, SAL corrections, and the normal transport at the boundaries), $\mathbf{L}$ represents the measurement functionals, and $\mathbf{d}$ is a vector of data; Σ_e and Σ_f are the covariances for the data and dynamical errors, respectively.

The scalar σ is introduced here as a regularization parameter that controls the relative fit to the data and the dynamics. If the model and data errors are normally distributed and their covariances are known, then the minimizer $\mathbf{u} = \hat{\mathbf{u}}$ of $J[\mathbf{u}, 1]$ is the maximum likelihood estimate for the tidal state. However, for tidal estimation, we cannot pretend to satisfy the prerequisites of true statistical estimation, and σ provides a single parameter for adjusting the relative scale of the model and data errors. The value of σ is estimated using generalized cross validation (Wahba 1990), described in more detail in the appendix.

1) DATA ERRORS

The residual data shown in Fig. 1 have been computed from nearly 10 years (364 orbit cycles) of Ocean Topography Experiment (TOPEX)/Poseidon (T/P) altimetry data collected from 1992 to 2002. These data consist of over 2 million individual measurements at 7281 sites within the model domain.

The precision of the altimetric data may be estimated from the nontidal variance of the time series. For the M_2 frequency, this precision is approximately 0.16 $(\text{cm})^2$. This is the variance of the smallest-scale noise in the along-track data presented in Fig. 1. Unfortunately, this estimate of the precision is not the same as the data error variance. As is apparent from the figure, the altimeter data contains a signal due to the surface expression of the internal tide, which has an amplitude of 2–4 cm, correlated over a scale of 100 km (Ray and Mitchum 1996; Dushaw 2002). As the tidal model considered here does not contain the dynamics sufficient to generate these features, they must be regarded as error in the data. Evidently, this "error of representation"

has variance roughly a factor of 10 larger than the basic precision estimate would suggest, and it is correlated along track.

It is not obvious how to account for the structure of the data errors in this context. The so-called error is, in fact, the signature of a deterministic process. If one were to use a covariance to describe the error, it would be inhomogeneous in both magnitude and correlation scale across the model domain. Equivalently, one might attempt to low-pass filter the data because of the scale separation between the baroclinic and barotropic waves. In fact, near bathymetric features like the Hawaiian Ridge, the scale separation between baroclinic and barotropic signals does not obtain: bathymetric features on the same length scale as the baroclinic waves can cause small-scale features in the barotropic tide. It is possible that an a posteriori analysis of the residuals could be used to find a data error covariance structure function (or, equivalently, an along-track smoothing kernel), but we have not attempted this.

For simplicity, we have decided to assume that the data errors are uncorrelated and allow changes in the variance scale via the regularization parameter σ. Based on this approach, one can make an a priori estimate of the regularization parameter. The along-track spacing of the measurements is about 7 km, and the decorrelation scale is on the order of 70 km. Thus, these serial correlations suggest that the data should be de-weighted by a factor of 10. Likewise, the variance due to the internal tide is about 10 times the precision estimate, which suggests a further de-weighting of the data by another factor of 10. When Σ_e in Eq. (7) is based on the data precision, this heuristic argument suggests that the optimal value of σ should be near 100. Indeed, this crude estimate is consistent with results obtained from the generalized cross validation discussed below.

2) MODEL ERRORS

The tidal fields should not be expected to satisfy the model equations exactly for various reasons. The effects of the elastic earth, gravitational self-attraction, and the stresses modeled by $\mathcal{F}$ are approximated, and there is uncertainty in the bathymetry itself. In this study, we estimate the spatially varying magnitude of errors in the model by taking simple percentages of the terms in the momentum equations, following Egbert et al. (1994). For example, an uncertainty of 5% in the bathymetry data suggests a 5% error in the pressure gradient, and the error in the bottom drag is assumed to be 50%. An exception to this simple approach concerns errors due to baroclinic conversion. The magnitude of the error in $\mathcal{F}_{\text{BC}}$ is assumed to be
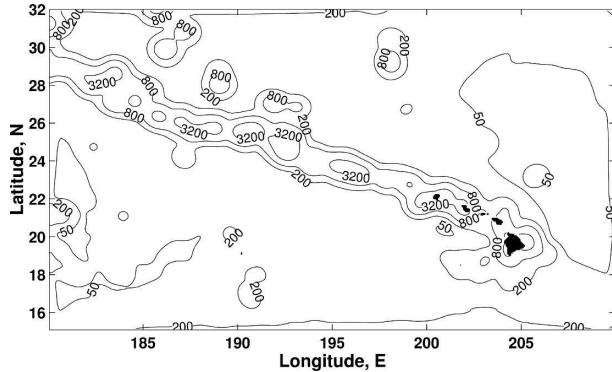
FIG. 2. The assumed model error variance for the zonal momentum equation is shown $[(m \, s^{-1})^4 \times 10^9]$. Note that the contours are evenly spaced on a logarithmic scale.

$$\overline{\mathcal{F}}_{BC} = \overline{c}_{BC} \frac{N_B (\nabla H)^2}{\kappa H} \mathbf{U}_0, \qquad (8)$$

where $\mathbf{U}_0$ is the amplitude of the transport in the forward model solution [i.e., the solution to Eqs. (2) and (3) computed with $c_{BC} = 0$]. The parameter $\overline{c}_{BC}$ is set to 1.51, which yielded the same nominal dissipation as found in the PAC6 and TPXO.5 solutions. Each source of error in the momentum equations is assumed to be independent, so the magnitude of each can be simply squared and summed to obtain the total assumed model error variance. The small-scale features in this error variance field are smoothed by convolution with a bell-shaped (Gaussian) kernel with a 50-km correlation scale. As an example, the variance of errors in the zonal momentum equation is plotted in Fig. 2. It may be seen that the error is amplified over topographic features where baroclinic effects are likely prominent. The model errors are assumed to be spatially correlated with a constant e-folding scale of 50 km, isotropic on the sphere.

Errors in the boundary conditions, which are assumed to be spatially correlated and inhomogeneous, have been computed from a suite of 60 Monte Carlo simulations with the PAC6 model. Each simulation consisted of a complete inversion of synthetic altimeter data in which representative noise was added to the model and synthetic data, following the method described in Egbert and Erofeeva (2002). Based on these simulations, the open boundary data (normal component of volume transport) are estimated to be accurate to within 1%–5% of their magnitude. Near submarine topographic features (i.e., in the northwest corner of the domain), the boundary condition uncertainty rises to as much as 10%.
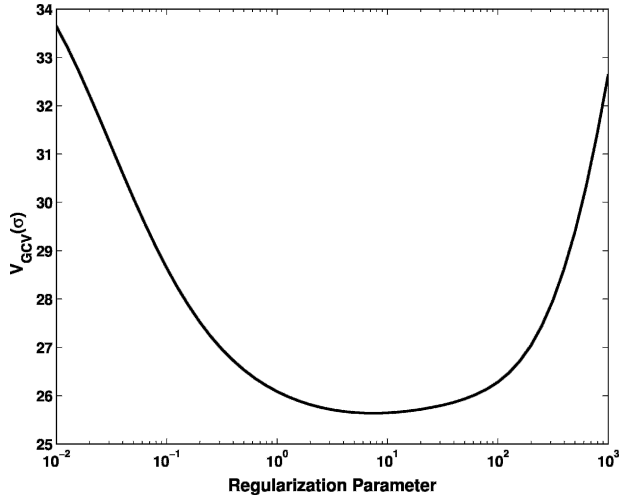

FIG. 3. The generalized cross-validation function is plotted as σ, the regularization parameter, is varied. Evidently, values of σ between about 1 and 100 are equally plausible.

d. Regularization parameter

The value of σ has been determined by generalized cross validation (Wahba 1990). This procedure, which is explained in the appendix, is carried out using the singular value decomposition of matrices that are created during the generalized inversion. The inversion for each constituent is performed separately, so an independent cross validation is performed to determine the value of σ at each tidal frequency.

The cross validation yields unambiguous results for the major semidiurnal constituents, M_2 and S_2. Figure 3 plots the generalized cross-validation function, $V_{GCV}(\sigma)$, for the M_2 constituent. This function is the scaled predictive error of the inverse solution. The minimum value of $V_{GCV}(\sigma)$ is obtained when σ is about 10, and this was also obtained for the S_2 constituent. The generalized cross-validation function for the other semidiurnal constituents is nearly flat, which suggests that there is little signal to fit, but the value $\sigma = 10$ was used for these cases as well.

Among the diurnal species, only the K_1 constituent had a convex generalized cross-validation function. The optimal value of σ was 100. Presumably, the larger value of σ is a consequence of the larger correlation scale of the internal tidal signal in the data. This same value of σ was used for all of the diurnal constituents.

3. Dissipation estimates

The generalized inversion of Eqs. (2) and (3), together with boundary conditions and data, was accomplished using OTIS (Egbert and Erofeeva 2002). The
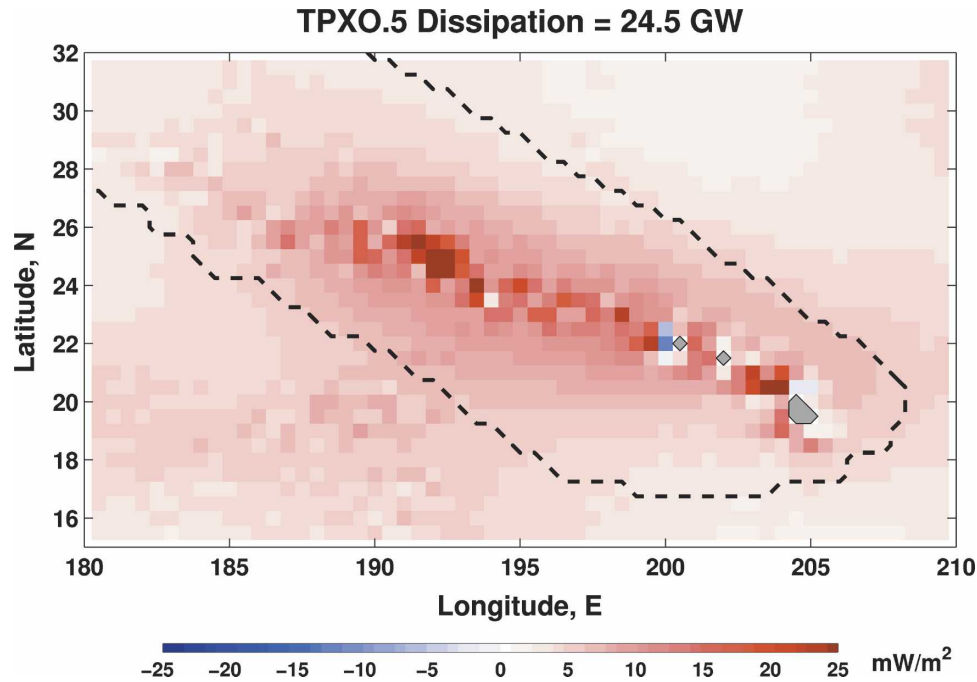
TPXO.5 Dissipation = 24.5 GW

FIG. 4. The TPXO.5 calculations were performed on a global grid at ½° resolution using a model that included the effects of the turbulent bottom boundary layer but not baroclinic conversion. As such, the barotropic dissipation shown here is entirely a consequence of the data assimilation (the effects of bottom friction are confined to a small number of very shallow passages along the Hawaiian Ridge). Because the vast majority of the barotropic dissipation is believed to occur via baroclinic conversion in association with bathymetric features, the spatial distribution of the barotropic dissipation in the figure is unrealistic—only 35% of the dissipation occurs within 250 km of the ridge. Note that the smaller value of dissipation along the ridge, 18 GW, published previously by Egbert and Ray (2001), was obtained by integrating over the area within the dashed curve.

equations were solved on a ¹⁄₁₅° finite-difference grid with bottom depth H obtained from the ETOPO2 database (Smith and Sandwell 1997). The tidal potential, self-attraction, and loading were computed as described in Egbert and Ray (2001), with ζ_{SAL} computed from the TPXO.5 solution.

Previous estimates of barotropic M_2 tidal dissipation obtained from global-scale models range from 18 to 22 GW along the Hawaiian Ridge (Egbert and Ray 2001). Figure 4 shows the results of a more recent calculation, TPXO.5, which is a global, coarse-resolution (½°), data-assimilative model (Egbert and Ray 2003). Hawaii is hardly resolved, and the dissipation occurs in general association with the ridge. Note that the color scale of this figure, which is different from that used below, emphasizes the small dissipation, 5–10 mW m⁻², which occurs over a vast area. Only about 35% of the barotropic dissipation in this estimate occurs within 250 km of the ridge axis. The energy dissipation via bottom drag is quite negligible at the Hawaiian Ridge, so it is presumed that the vast majority of the barotropic dissipation should occur in close association with bathy-

metric features via baroclinic conversion. Hence, the small fraction of the dissipation that does occur near the ridge is taken as evidence that the resolution of this model is insufficient for detailed open-ocean regional studies.

The dissipation in the area around Hawaii computed from the ¹⁄₆°-resolution North Pacific basin model (PAC6), which provides the boundary conditions for the regional model, is shown in Fig. 5. Dissipation in this higher-resolution inverse solution is largely confined to the ridge and is most intense near French Frigate Shoals. Note that the PAC6 and TPXO.5 models included no parameterization for baroclinic conversion. Hence, the barotropic dissipation inferred from the data-assimilative runs is entirely a consequence of data assimilation.

The estimate of the barotropic M_2 tidal dissipation around Hawaii, obtained from the regional data-assimilative model with parameters as described above, is shown in Fig. 6. In this solution, the total barotropic dissipation in the model domain is 18.9 GW, of which 18.4 GW is converted to the internal tide. Note that the
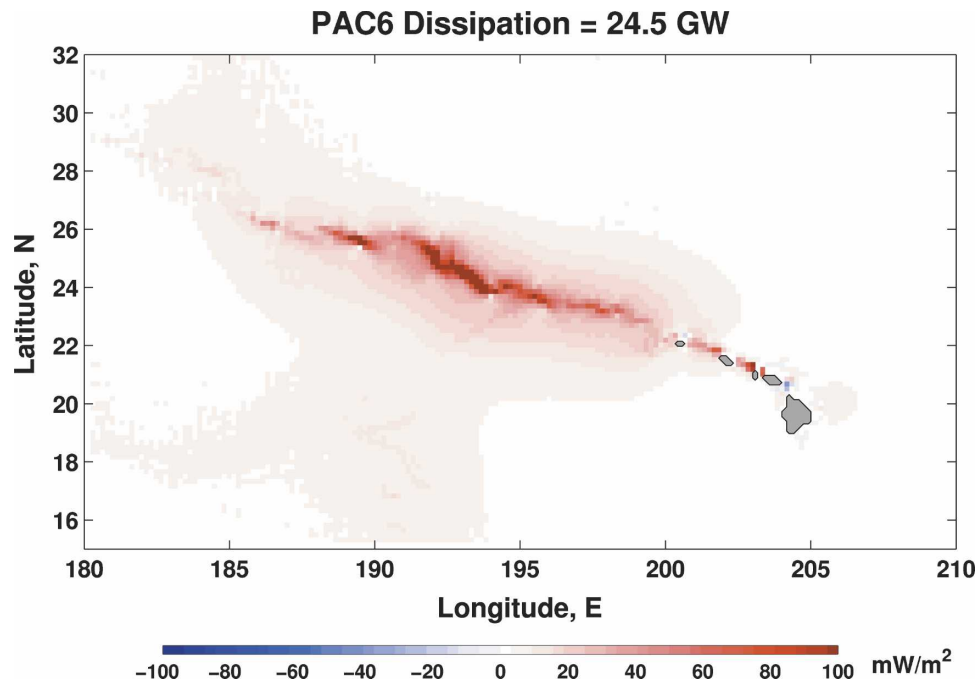
FIG. 5. The inverse solution from the ⅙° resolution North Pacific basin model (PAC6), which provides boundary conditions to the regional model of the Hawaiian Ridge, shows a large elevation in dissipation levels at French Frigate Shoals. Note that, as above, the original model contained no representation of the baroclinic conversion process. The barotropic dissipation presented in the figure is entirely a consequence of the data assimilation. When one adds a forcing to the momentum equations, which corresponds to the pictured dissipation, the surface elevation predicted by the model is a least squares best fit to the altimeter observations.

prior model used $c_{BC} = 0$, so this 18.4 GW of baroclinic conversion results entirely from the data assimilation. The data assimilation procedure has added the forcing necessary to the right-hand side of Eq. (2) to bring the surface elevation into agreement with the observed data. This extra forcing, the model forcing residual, obtains its maximum amplitude along the ridge axis. The barotropic dissipation shown in Fig. 6 reveals that the model residual is arranged so that it extracts work from the barotropic flow almost everywhere; that is, it has added an energetic sink to the model, in conformance with our expectations for the direction of energy transfer via baroclinic conversion. In other words, the data assimilation has allowed us to estimate the form drag in the momentum equations necessary to reconcile the model with the large-scale features in the data.

Figure 7 compares the inverse solution with the T/P data (cf. Fig. 1). The model forcing residual, which is an estimate of the form drag, has produced a large-scale adjustment on the northeast and southwest sides of the ridge, which has reduced the spatially correlated errors in these regions. The residuals clearly show the signal of the baroclinic internal tide, with maximum amplitude near the ridge between 190° and 195°E. Because the model equations do not contain the physics capable of reproducing the baroclinic waves in the data, the residual data in Fig. 7 are an estimate of just the baroclinic waves. A comparison of Figs. 6 and 7 shows that the region of maximum barotropic dissipation is roughly coincident with the area of the largest internal-tidal signal. This bolsters our confidence that the barotropic dissipation we have estimated is an accurate estimate of the actual baroclinic conversion.

It is important to reiterate that the barotropic dissipation presented in the figures is the sum of actual turbulent dissipation in the bottom boundary layer and dissipation via (nonturbulent) baroclinic conversion at bathymetric features. Furthermore, the barotropic dissipation in Fig. 6 results from the sum of the explicitly modeled drag forces ($\mathcal{F}_B$ and $\mathcal{F}_{BC}$) and the model forcing residual that the data-assimilation procedure provides. In order not to bias the dissipation estimates toward any particular numerical value, the explicitly modeled wave drag ($\mathcal{F}_{BC}$) was set to zero in the prior model (i.e., $c_{BC} = 0$ in the prior model). In essence, the purpose of the $\mathcal{F}_{BC}$ term (with nonzero c_{BC}) was solely to provide a scale estimate for the variance of the model forcing residual. The sensitivity of our results to
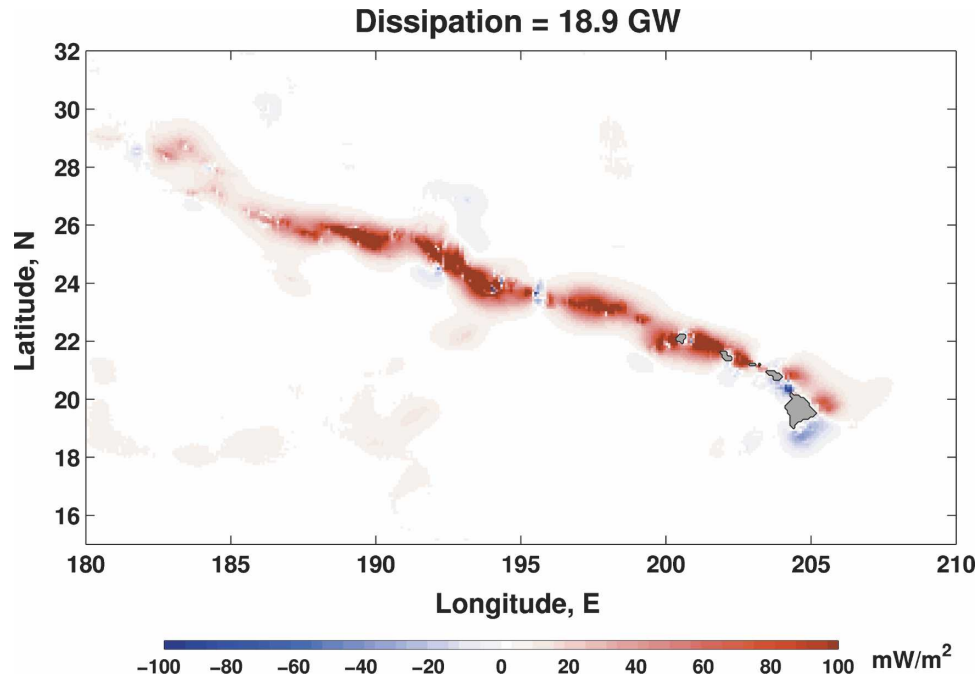
FIG. 6. The barotropic dissipation inferred from the generalized inverse of the tidal model and the T/P altimeter data. The dissipation results from a combination of the explicitly modeled bottom friction (0.5 GW) and the model forcing residual, which is interpreted as conversion to baroclinic tidal motion (18.4 GW). Virtually all of the inferred dissipation is a result of assimilating the altimeter data.

the particular value of c_{BC} in the prior model is analyzed in the next section, below.

The spatial distribution of the barotropic M_2 dissipation within 250 km of the ridge is shown in Fig. 8a. Approximately 88% of the energy loss occurs within 250 km of the ridge axis, with the majority occurring in the region between the Gardner Pinnacles and French Frigate Shoals. Three other sites are also significant: Mara Reef, Nihoa Island, and Kauai Channel. If one assumes that 100% of the energy lost from the barotropic tide is converted to baroclinic energy and radiated as a coherent beam of internal waves oriented normal to the ridge, one would see a peak wave energy flux of 15 kW m^{-1} to both the northeast and southwest near French Frigate Shoals. Other sites of dissipation could radiate as much as 5–7 kW m^{-1}.

Barotropic dissipation estimates were also made from inversions at seven other tidal frequencies: S_2, N_2, K_2, K_1, O_1, P_1, and Q_1. Table 1 shows the total dissipation at these frequencies, as well as the percentage that occurs within 250 km of the ridge. The non-M_2 constituents lose approximately an additional 5.7 GW total. The ratio of barotropic dissipation to squared equilibrium tidal amplitude for the diurnal constituents is uniformly less than 25% of a similar ratio for the semidiurnal constituents. This is consistent with the observation from global models that the diurnal tides lose a larger fraction of their energy in marginal seas than the semidiurnal tides (Egbert and Ray 2003).

The spatial distribution of barotropic dissipation along the ridge is similar for all of the semidiurnal species. In every case, the dissipation is maximum along the topographic features from the Gardner Pinnacles to French Frigate Shoals. In contrast, the diurnal species all have a sharp peak at the Gardner Pinnacles, with little dissipation at French Frigate Shoals or elsewhere. These results are robust to changes in the regularization parameter and the model forcing error covariance. It must be acknowledged that the N_2, K_2, O_1, P_1, and Q_1 constituents have a low signal-to-noise ratio in comparison with the M_2 signal. Also, the S_2 tide certainly contains a component that is driven by daily insolation-induced atmospheric pressure variations (Cartwright and Ray 1994), and the K_1 data contain some aliased semiannual signal (Andersen and Knudsen 1997). For these reasons, a complete sensitivity analysis of the dissipation was performed only for the M_2 constituent.

4. Sensitivity to assumptions

The estimated barotropic dissipation and tidal fields depend on assumptions regarding the forward tidal model and its errors. We have verified that the results are completely insensitive to reasonable changes in the
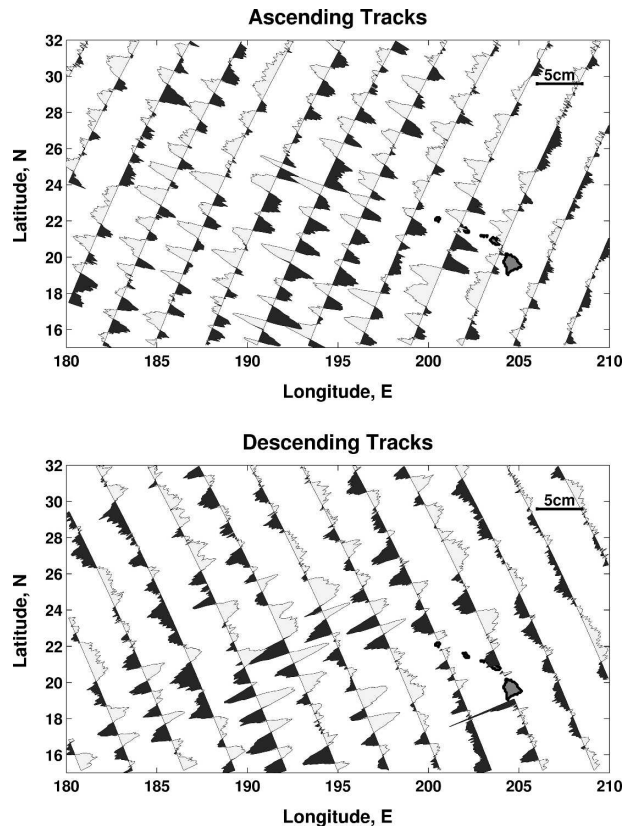
FIG. 7. The difference between the T/P observations and the inverse solution for the M_2 constituent. The primary effect of the inversion has been to introduce a large-scale tilt amounting to a 2-cm change in the surface elevation across the domain.

friction velocity, v_f, and to changes in the assumed error of the no-normal-flow condition at coastlines. Changes to the open boundary condition error covariance results in approximately ±1 GW changes in dissipation for the M_2 constituent.

Three parameters lead to more significant uncertainty in the dissipation, and we now discuss these in more detail. These parameters are

1) the level of baroclinic conversion in the prior model, that is, the value of c_{BC};
2) the correlation length scale of the assumed model error; and
3) the ratio of the assumed model and data errors, as controlled by the regularization parameter σ.

Experiments were conducted varying c_{BC} in the prior model from 0 (no conversion) to 3.62 (40-GW conversion). Figure 9 shows that the dissipation computed from the inverse solution is remarkably insensitive to the value of c_{BC} in the forward model. Over this entire range, the inverse dissipation only varied from approximately 19 to 28 GW. It is encouraging to see that the

inverse estimate of dissipation is relatively insensitive to the value of c_{BC}; nonetheless, this analysis provides no basis for rejecting an M_2 dissipation level as high as 25 GW. Dissipation values above this would require a prior model that dissipates some 30 GW or more, which we reject a priori. These higher levels of dissipation, if they occur, are simply not observable by the altimetry using the present formalism.

The e-folding scale of the correlation structure function for the assumed model errors was 50 km. While there is no rigorous justification for this scale, it is comparable to the length scale observed in the inferred dissipation. To test the sensitivity of the dissipation estimates to this length scale, several experiments were conducted using a 200-km correlation scale; Fig. 10 shows the pattern of dissipation. There is 1.6 GW more dissipation than the nominal best case shown in Fig. 6. There is less nonphysical negative dissipation in comparison with the standard case, and the dissipation is still localized along the ridge.

The dissipation estimates depend weakly on the regularization parameter σ. This parameter controls the trade-off between a perfect fit to the prior forward model ($\sigma \to \infty$) and a perfect fit to the data ($\sigma \to 0$). Figure 3 suggests that a range of σ values between 1 and 100 are equally justified; however, the generalized cross-validation procedure tends to overfit the data when the data errors are correlated (Wahba 1990), as is the case here. Inspection of the actual tidal fields suggests that $\sigma \in [10, 100]$ is the plausible range. Varying σ over this range changes the dissipation by 2 GW.

A more limited series of experiments were conducted to assess the sensitivity of the non-M_2 dissipation estimates to the aforementioned factors. These constituents were found to have a sensitivity proportional to the M_2 case.

Figure 8 summarizes the above considerations and presents a detailed view of the barotropic M_2 dissipation along the Hawaiian Ridge. The dissipation has been integrated across a 500-km swath centered on the ridge axis. The heavy solid line in Fig. 8b corresponds to the nominal best estimate in Fig. 6. The sensitivity to the three factors considered above is indicated by the shaded gray band, which shows the range of dissipation values obtained in all of the sensitivity experiments. The distribution of the dissipation is remarkably insensitive to reasonable changes in these parameters. The highest dissipation level is found at French Frigate Shoals with other local maxima occurring at Kauai Channel, Nihoa Island, Gardner Pinnacles, and Maru Reef. Overall, 88% ± 1% of the total dissipation occurs within 250 km of the ridge axis.

A) Bathymetry

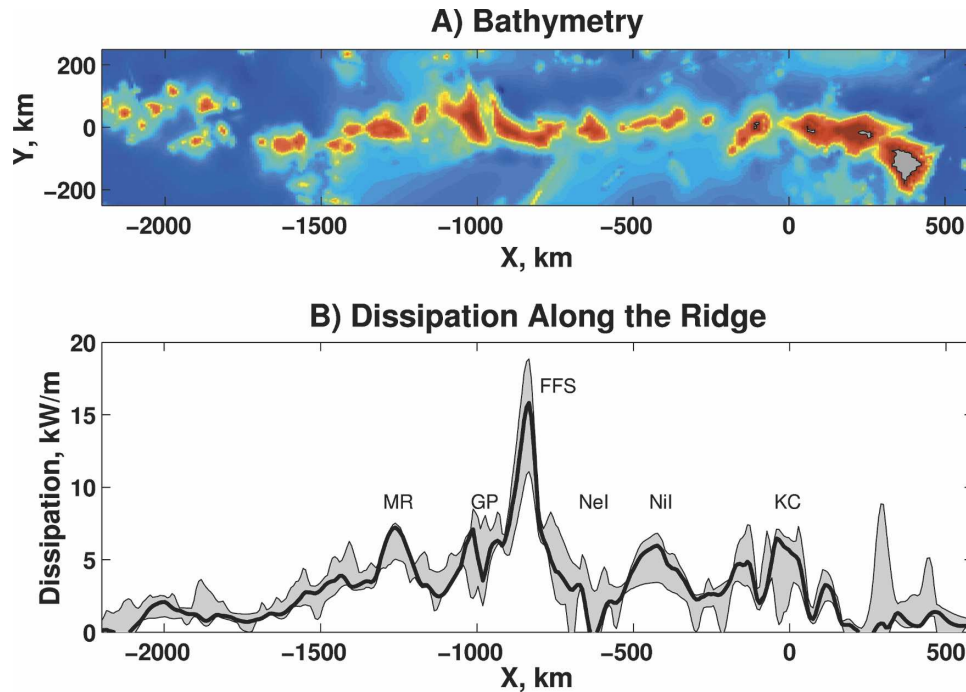## B) Dissipation Along the Ridge

FIG. 8. (a) The model bathymetry. (b) The estimated barotropic dissipation integrated across a 500-km strip, centered on the Hawaiian Ridge. The following geographic sites are noted: Maru Reef (MR), Gardner Pinnacles (GP), French Frigate Shoals (FFS), Necker Island (NeI), Nihoa Island (NiI), and Kauai Channel (KC). The heavy curve shows the best estimate of the dissipation, while the gray band indicates the range of values computed during the sensitivity tests. Variations of the regularization parameter σ, the value of c_{BC} in the forward model, and the correlation length scale of the assumed model errors all result in dissipation estimates contained in the gray band.

Estimating the uncertainty in our estimate of the total barotropic dissipation is a subjective exercise. Based on the sensitivity analysis for M_2, we conclude that the range of values from 17.5 to 25.0 GW cannot be rejected. Uncertainty in just the size and correlation of the model forcing errors would lead to a narrower estimate of 19 ± 1.5 GW, and our preferred estimate is 19 GW; however, an unbiased view, which a priori accepts dissipation from 0 to 30 GW, leads to the possibility of 25.0 GW as the upper extreme.

5. Discussion

Our estimates for the barotropic dissipation of tidal energy along the Hawaiian Ridge are a refinement of previous efforts. When all of the constituents are included, nearly 25 GW of power is available to drive internal waves and mixing over the model domain. Of this, 19 GW is attributed to the M_2 tide. The majority of the remaining 5.7 GW is due to the S_2, K_1, and N_2 constituents. Approximately 88% of the total, or 22 GW, is lost from the barotropic tide within 250 km of the ridge. Presumably, this means that 22 GW of energy

is pumped into the baroclinic wave field at the ridge; some unknown fraction of this energy is used to drive local turbulence, while the remainder radiates into the deep ocean where it may provide energy for turbulence in the main thermocline. It must be emphasized that our use of the term "barotropic dissipation" refers to the rate of energy loss from the barotropic flow: The vast majority of this energy directly forces the internal tidal field, and only a small fraction of the of barotropic kinetic energy is lost directly to turbulent dissipation.

TABLE 1. Dissipation summary.

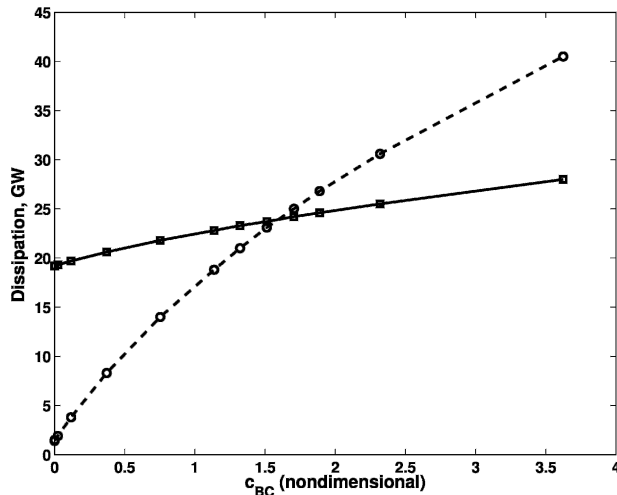| Constituent	Dissipation (GW)	Fraction on ridge (%)
M_2	19	88
S_2	3.1	91
N_2	0.50	88
K_2	0.33	92
K_1	1.4	85
O_1	0.32	85
P_1	3.0×10^{-2}	98
Q_1	6.1×10^{-3}	10

FIG. 9. Sensitivity of the estimated dissipation to c_{BC}: The dashed line indicates the amount of baroclinic conversion in the forward model, and the solid line shows the dissipation in the generalized inverse. As can be seen, the inverse estimate of the dissipation is relatively insensitive to the assumed level of baroclinic conversion in the prior model.

The uncertainty in our estimate of barotropic dissipation is due to a lack of knowledge regarding the structure of the errors in the tidal model. Plausible variations in the assumed model forcing error yield dissipation estimates that differ by as much as 3 GW. In addition, baroclinic conversion may occur at scales too small to be constrained by the altimetric observing array. Consequently, barotropic M_2 dissipation values of up to 25 GW are not inconsistent with the data.

The high-resolution regional model has made it possible to identify specific sites of barotropic dissipation along the ridge. The semidiurnal species lose most of their energy over the topography between the Gardner Pinnacles and French Frigate Shoals. The diurnal dissipation is mostly confined to Gardner Pinnacles. These are areas with strong cross-isobath flows and are good candidates for sites of baroclinic conversion.

In fact, the complete energy budget at tidal frequencies includes a number of processes. First, there is the work done by form drag, which we have been calling baroclinic conversion, and it is assumed that this accounts for much of the barotropic dissipation shown in Fig. 6. Then, there are the truly dissipative processes: small-scale turbulent mixing and boundary layer mixing at the ocean surface and bottom. There are also nonlinear energy exchanges that occur among internal waves of different frequencies.

There are direct measurements of turbulent dissipation at a number of sites along the Hawaiian Ridge. Klymak et al. (2006) extrapolate data from French Frig-

ate Shoals (FFS), Necker Island (NeI), Nihoa Island (NiI), and Kauai Channel (KC) to estimate 4–10 GW of turbulent dissipation for the entire ridge. There is a general qualitative agreement between these observations and our estimates; for example, both indicate that there is less dissipation at NeI than the other sites. However, the significant difference in barotropic dissipation rates between FFS and KC (see Fig. 8b) is not seen in the turbulence data. In the relationship between turbulent dissipation and tidal kinetic energy postulated by Klymak et al. (2006), sites with high tidal kinetic energy will dissipate proportionally less energy as local turbulence, so it may be that FFS actually radiates much more energy than KC. Klymak et al. (2006) estimate maximum turbulent dissipation rates of about 5 kW m^{-1}, peak, at FFS, NiI, and KC. Subtracting one-half of this value from our M_2 ridge-normal energy flux yields wave energy fluxes of roughly 12, 3, and 3 kW m^{-1} emanating from FFS, NiI, and KC, respectively.

A limited number of in situ measurements of the baroclinic wave energy flux have been made (Rudnick et al. 2003). At FFS these fluxes range from 4 to 15 kW m^{-1} normal to the ridge. At NiI and KC the observed fluxes span a large range with a maximum ridge-normal flux of 11 kW m^{-1}. Much smaller flux was observed at NeI. These data, which are each representative of the flux averaged over a single semidiurnal period, are in the same range as wave energy flux inferred from our own estimates of barotropic tidal dissipation.

In the absence of more extensive baroclinc wave flux measurements, comparison with a numerical model is instructive. Figure 11 compares the pattern of barotropic M_2 dissipation found in this study with the divergence of M_2 internal wave energy diagnosed from a three-dimensional primitive equation model (Merrifield and Holloway 2002). What has been plotted as "model dissipation" is, in fact, the difference of baroclinic conversion and turbulent dissipation, so it is strictly less than the barotropic dissipation plotted in the top panel. Nonetheless, there is encouraging agreement as to the sites of significant energy flux divergence. The bottom panel of the figure shows a coincidence of the major peaks of ridge-normal energy flux, except at Maru Reef (which, unfortunately, is located at the edge of one of the numerical model's subdomains). Subtracting 7 GW [the along-ridge dissipation estimate of Klymak et al. (2006)] from 22 GW (the total baroclinic conversion found within 250 km of the ridge at eight tidal frequencies in the present study), leaves a total of 15 GW to be radiated as internal waves. The M_2 calculations of Merrifield and Holloway (2002) find about 60% of this amount. It should be noted that both
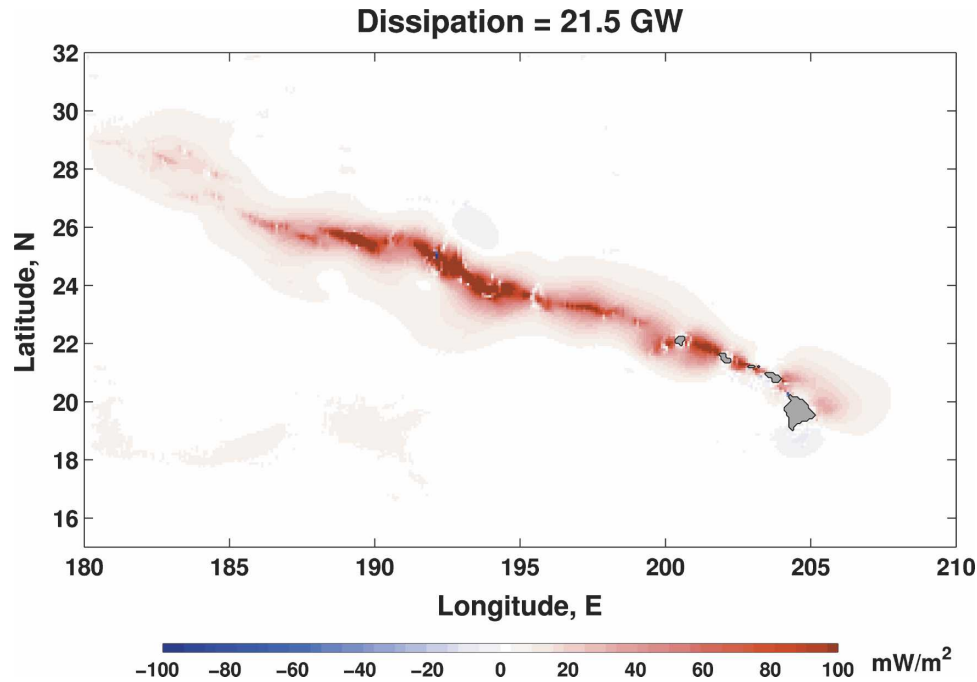
FIG. 10. The plot shows the dissipation when a 200-km correlation scale is assumed for the model errors. The pattern of dissipation is very similar to that found with a 50-km correlation scale. The tidal fields (not shown) are also very similar (surface elevations agree to within 2 mm over most of the domain).

of these estimates greatly exceed the energy radiated by coherent low-mode internal waves as inferred from altimetry (Ray and Cartwright 2001; Dushaw 2002). We eagerly await the results of more in situ measurements of the internal wave energy flux.

A principal limitation of this study is that the free-surface expression of the internal tides, which is present in the altimeter measurements, must be regarded as an error of representativeness in the data. Work is currently under way to develop the generalized inverse of a three-dimensional primitive equation model. With this new tool it will be possible to more directly model the baroclinic conversion process and to map the resulting tidal internal wave field in a dynamically consistent manner. Sources and sinks of internal tidal energy can be diagnosed, and it should be possible to further improve the constraints on the tidal energy budget.

APPENDIX

Generalized Cross Validation on the Reduced Subspace

The numerical value of the regularization parameter σ is found by using a generalized cross-validation method. Because the cross validation is performed using a subset of the data vector after several transformations, it is useful to explain the transformations and inversion procedure in some detail. This section essentially repeats the explanation in Egbert and Erofeeva (2002) with special emphasis on the points of relevance to the determination of σ.

The original time series data vector is transformed into a reduced data vector by harmonic analysis of the original time series. This transformation simply projects the data vector into a subspace spanned by the tidal frequency of interest. Next, the reduced data vector and the representer coefficients undergo a rotation that facilitates the solution of the Euler–Lagrange equations on the reduced basis. Last, this reduced and rotated data vector is projected into the subspace spanned by the reduced representer basis. The generalized cross-validation procedure described below is standard, but it is performed on this twice-projected and rotated data vector.
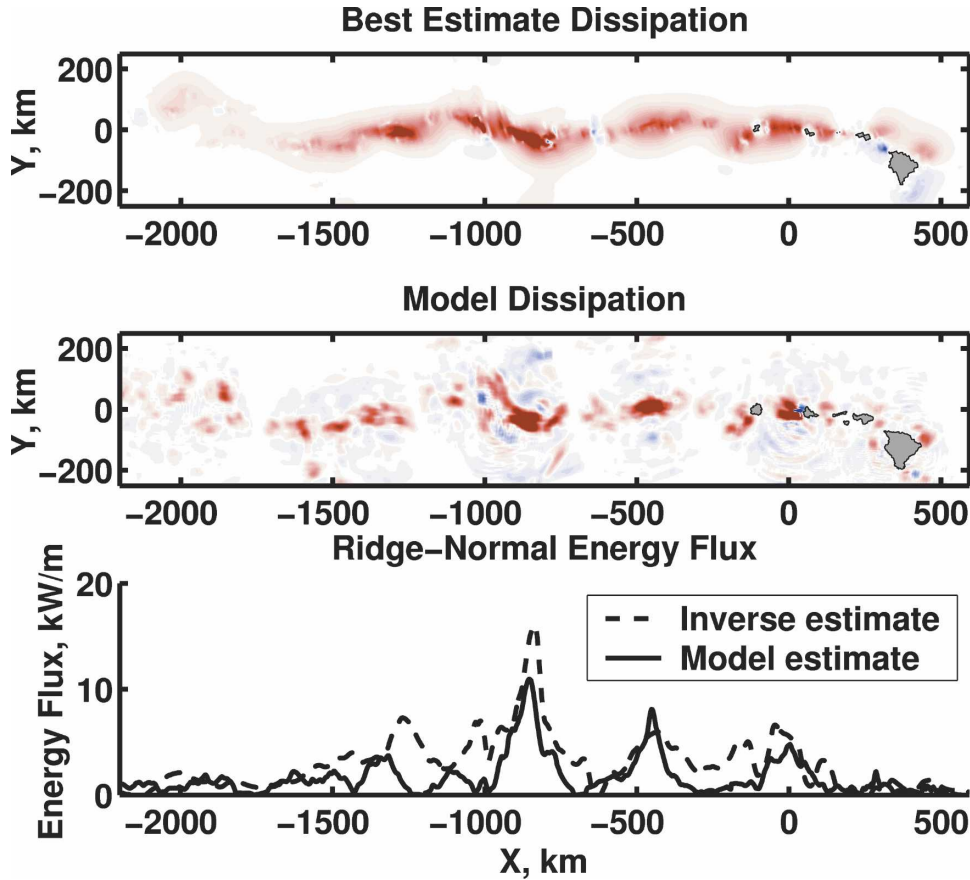
FIG. 11. (top) The M_2 barotropic tidal dissipation in a 500-km-wide swath centered on the Hawaiian Ridge. (middle) The divergence of baroclinic wave energy flux diagnosed from the model of Merrifield and Holloway (2002), which is a balance of baroclinic conversion and turbulent dissipation. The color scales are identical, but reversed in sign; i.e., red denotes energy loss in the top panel, while red denotes energy gain in the middle panel. (bottom) Comparisons of the nominal ridge-normal energy flux inferred from the present study (dashed) and the ridge-normal baroclinic wave flux diagnosed from the model (solid).

The time series data are related to a vector $\mathbf{z}$ of harmonic constants by

$$\mathbf{d} = \mathbf{A}\mathbf{z} + \boldsymbol{\varepsilon}. \qquad (A1)$$

The vector $\mathbf{d}$ contains the time series of altimeter data at every site; it is of dimension $N_d \approx 2 \times 10^6$. The vector $\mathbf{z}$ contains two entries for each measurement site, the real and imaginary parts of the harmonic constants for the frequency of interest. The dimension of $\mathbf{z}$ is $N_z = 2 \times 7281$. The vector $\boldsymbol{\varepsilon}$ contains both measurement error and the nontidal component of the data vector. The matrix $\mathbf{A}$ is defined in Egbert and Erofeeva (2002); it depends in a straightforward way on the times of the satellite measurements.

The $\mathbf{QR}$ decomposition of $\mathbf{A}$ is used to define the reduced data vector. Let $\mathbf{A} = \mathbf{QB}$ be the $\mathbf{QR}$ decomposition of $\mathbf{A}$ (Golub and Van Loan 1989), where $\mathbf{Q}$ is an $N_d \times N_z$ orthogonal matrix (i.e., $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, the $N_z \times N_z$ identity matrix), and $\mathbf{B}$ is an $N_z \times N_z$ upper-triangular matrix. The reduced data vector is given by

$$\tilde{\mathbf{d}} = \mathbf{Q}^T\mathbf{d} = \mathbf{B}\mathbf{z} + \mathbf{Q}^T\boldsymbol{\varepsilon}. \qquad (A2)$$

The least squares estimates of the real and imaginary parts of the harmonic constants at the data sites are given by

$$\mathbf{z} = \mathbf{B}^{-1}\tilde{\mathbf{d}}. \qquad (A3)$$

To project the reduced data vector into the reduced representer subspace, the representer subspace must be defined. First, the minimizer of $J[\mathbf{u}, \sigma]$ is written as

$$\hat{\mathbf{u}} = \mathbf{u}_0 + \sum_{k=1}^{N_z} \beta_k \mathbf{r}_k, \qquad (A4)$$

where $\mathbf{u}_0 = \mathbf{S}^{-1}\mathbf{f}_0$ is the exact solution of the forward model, and the $\mathbf{r}_k$ are the representer functions, $k =$

1, ..., N_z. The representer functions are obtained by first solving for α_k,

$$\mathbf{S}^{\dagger}\alpha_k = \Delta_k, \tag{A5}$$

(where Δ_k is the kernel for the data functional L_k), and then solving for $\mathbf{r}_k$,

$$\mathbf{S}\mathbf{r}_k = \mathbf{\Sigma}_f \alpha_k. \tag{A6}$$

In the reduced basis approach (Parker and Shure 1982; Egbert et al. 1994) a representative subset of the representers is used,

$$\hat{\mathbf{u}} = \mathbf{u}_0 + \sum_{k=1}^{N} \beta_k \mathbf{r}_k, \tag{A7}$$

where $N \leq N_z$. It has been assumed, without loss of generality, that the N_z representers are indexed so that only the first N are taken as the reduced basis. Let $\mathbf{P}$ be the $N_z \times N$ matrix, which is constructed by applying all N_z measurement operators to each of the N representers in the reduced basis, $P_{ij} = L_i \mathbf{r}_j$; and let $\mathbf{R}$ be the $N \times N$ matrix consisting of just the first N rows of $\mathbf{P}$, that is, $R_{ij} = P_{ij}$ for $1 \leq i \leq N$ and $1 \leq j \leq N$. The penalty functional may now be written

$$J[\boldsymbol{\beta}, \sigma] = (\mathbf{d} - \mathbf{Q}\tilde{\mathbf{d}})^{\mathrm{T}} \mathbf{\Sigma}_e^{-1} (\mathbf{d} - \mathbf{Q}\tilde{\mathbf{d}})$$
$$+ (\mathbf{BP}\boldsymbol{\beta} - \tilde{\mathbf{d}}')^{\mathrm{T}} \mathbf{\Sigma}_e^{-1} (\mathbf{BP}\boldsymbol{\beta} - \tilde{\mathbf{d}}') + \sigma \boldsymbol{\beta}^{\mathrm{T}} \mathbf{R} \boldsymbol{\beta}, \tag{A8}$$

where $\boldsymbol{\beta} = [\beta_1, ..., \beta_N]$ is the vector of representer coefficients, and $\tilde{\mathbf{d}}' = \tilde{\mathbf{d}} - \mathbf{BLu}_0$. The penalty functional is the sum of three terms: 1) a nontidal data term, 2) a tidal data error term, and 3) a model error term. Note that the nontidal data term, which is independent of β, is simply a constant and cannot be reduced. Also, as a consequence of the representer expansion, the penalty functional (a functional of the infinite-dimensional field $\mathbf{u}$) is now simply a penalty function (a function of the vector $\boldsymbol{\beta}$).

Before completing the final projection of $\tilde{\mathbf{d}}'$ into the space spanned by the reduced basis, it is convenient to nondimensionalize the reduced data vector and the representer coefficients with the transformations

$$\hat{\mathbf{d}} = \mathbf{\Sigma}_e^{-1/2} \tilde{\mathbf{d}}' \tag{A9}$$

and

$$\mathbf{c} = \mathbf{R}^{1/2} \boldsymbol{\beta}. \tag{A10}$$

In terms of these variables, the penalty function becomes

$$J[\mathbf{c}, \sigma] = (\mathbf{d} - \mathbf{Q}\tilde{\mathbf{d}})^{\mathrm{T}} \mathbf{\Sigma}_e^{-1} (\mathbf{d} - \mathbf{Q}\tilde{\mathbf{d}})$$
$$+ \sum_{n=1}^{N} [(\mathbf{\Sigma}_e^{-1/2}\mathbf{BPR}^{-1/2}\mathbf{c})_n - \hat{d}_n]^2 + \sigma \sum_{n=1}^{N} c_n^2, \tag{A11}$$

where it is understood that the nth element of a vector is indicated by the subscript n.

Let $\mathbf{c}_\sigma$ denote the value of $\mathbf{c}$ that minimizes $J[\mathbf{c}, \sigma]$, keeping σ fixed. The strategy outlined in Egbert and Erofeeva (2002) provides an efficient algorithm for obtaining $\mathbf{c}_\sigma$. Let the $\mathbf{QR}$ decomposition of $\mathbf{\Sigma}_e^{-1/2}\mathbf{BPR}^{-1/2}$ be given by $\mathbf{W}_F\mathbf{Q}_F$, where $\mathbf{W}_F$ is an $N_z \times N$ orthogonal matrix and $\mathbf{Q}_F$ is an $N \times N$ upper-triangular matrix. Let $\overline{\mathbf{d}}$ be defined by projecting the data vector, $\hat{\mathbf{d}}$, into the space spanned by the reduced representer basis, $\overline{\mathbf{d}} = \mathbf{W}_F^{\mathrm{T}}\hat{\mathbf{d}}$. In terms of these variables, the penalty function may be written

$$J[\mathbf{c}, \sigma] = (\mathbf{d} - \mathbf{Q}\tilde{\mathbf{d}})^{\mathrm{T}} \mathbf{\Sigma}_e^{-1} (\mathbf{d} - \mathbf{Q}\tilde{\mathbf{d}}) + (\hat{\mathbf{d}} - \mathbf{W}_F\overline{\mathbf{d}})^{\mathrm{T}}(\hat{\mathbf{d}} - \mathbf{W}_F\overline{\mathbf{d}})$$
$$+ \sum_{n=1}^{N} [(\mathbf{Q}_F\mathbf{c})_n - \overline{d}_n]^2 + \sigma \sum_{n=1}^{N} c_n^2. \tag{A12}$$

If we let the singular value decomposition (SVD) of $\mathbf{Q}_F$ be given by

$$\mathbf{Q}_F = \mathbf{W}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}, \tag{A13}$$

where $\mathbf{W}$ and $\mathbf{U}$ are orthogonal matrices and $\mathbf{\Lambda}$ is the diagonal matrix of singular values of $\mathbf{Q}_F$, then J is minimized by $\mathbf{c} = \mathbf{c}_\sigma$, where

$$\mathbf{c}_\sigma = \mathbf{U}(\mathbf{\Lambda}^2 + \sigma\mathbf{I})^{-1}\mathbf{\Lambda}\mathbf{W}^{\mathrm{T}}\overline{\mathbf{d}}. \tag{A14}$$

An ordinary cross-validation function may be defined in terms of the twice-reduced and rotated data vector, $\overline{\mathbf{d}}$. The ordinary (leaving out one) cross-validation function corresponding to this observable tidal data error in Eq. (A12) is

$$V_0(\sigma) = \frac{1}{N} \sum_{k=1}^{N} (M_k \mathbf{c}_\sigma^{[k]} - \overline{d}_k)^2, \tag{A15}$$

where M_k is the kth row of $\mathbf{Q}_F$, and $\mathbf{c} = \mathbf{c}_\sigma^{[k]}$ is the minimizer of the penalty function

$$J[\mathbf{c}, \sigma, k] = \sum_{n=1, n \neq k}^{N} [(\mathbf{Q}_F\mathbf{c})_n - \overline{d}_n]^2 + \sigma \sum_{n=1}^{N} c_n^2. \tag{A16}$$

The "leaving-out-one" lemma (Craven and Wahba 1979) shows that the ordinary cross-validation function may be written in terms of $\mathbf{c}_\sigma$ as

$$V_0(\sigma) = \frac{1}{N} \sum_{k=1}^{N} \frac{(M_k \mathbf{c}_\sigma - \bar{d}_k)^2}{[1 - A_{kk}(\sigma)]^2}, \qquad \text{(A17)}$$

where

$$A_{kk} = \frac{\partial M_k \mathbf{c}_\lambda}{\partial d_k}$$

are the diagonal elements of the so-called influence matrix, $\mathbf{A}$:

$$\mathbf{A} = \mathbf{W}(\Lambda^2 + \sigma\mathbf{I})^{-1}\Lambda^2\mathbf{W}^{\mathrm{T}}. \qquad \text{(A18)}$$

The importance of Eq. (A17) is that it allows one to evaluate the ordinary cross-validation function *without* having to calculate each value of $\mathbf{c}_\sigma^{[k]}$, for $k = 1, \ldots, N$.

The generalized cross-validation function is independent of rotational transformations of the data vector and, consequently, it is more stable than the ordinary cross-validation function for estimating the predictive error of $\mathbf{c}_\sigma$. The generalized cross-validation function is given by

$$V_{\mathrm{GCV}}(\sigma) = \frac{1}{N} \sum_{k=1}^{N} \frac{(M_k \mathbf{c}_\sigma - \bar{d}_k)^2}{[1 - \mu_1(\sigma)]^2}, \qquad \text{(A19)}$$

where $\mu_1(\sigma) = N^{-1}\mathrm{Tr}(A)$. By using the SVD of $\mathbf{Q}_F$, defined above, the generalized cross-validation function may be written

$$V_{\mathrm{GCV}}(\sigma) = \frac{1}{N} \sum_{k=1}^{N} \frac{[\sigma(\lambda_k^2 + \sigma)^{-1}(\mathbf{W}^{\mathrm{T}} \bar{\mathbf{d}})_k]^2}{[1 - \mu_1(\sigma)]^2}, \qquad \text{(A20)}$$

where λ_k is the kth element on the diagonal of Λ, and

$$\mu_1(\sigma) = \frac{1}{N} \sum_{k=1}^{N} \frac{\lambda_k^2}{\lambda_k^2 + \sigma}. \qquad \text{(A21)}$$

Figure 3 shows how the predictive error of the M_2 inverse solution depends on the regularization parameter, σ. For values of σ above 100, the penalty functional assigns the model too much weight, and the inverse solution oversmoothes the data. For σ less than 1, the penalty functional assigns the data too much weight, and the inverse solution overfits the data. That is to say, the predictive error increases as $\sigma \to 0$ because the inverse solution is interpolating measurement error in this limit. The predictive error is evidently minimum near $\sigma = 10$, which is used to produce the nominal best estimate in the text.

REFERENCES

Andersen, O. B., and P. Knudsen, 1997: Multi-satellite ocean tide modelling—The K_1 constituent. *Progress in Oceanograpy,* Vol. 40, Pergamon, 197–216.

Balmforth, N. J., G. R. Ierley, and W. R. Young, 2002: Tidal conversion by nearly critical topography. *J. Phys. Oceanogr.,* **32,** 2900–2914.

Bell, T., 1975: Topographically generated waves in the open ocean. *J. Geophys. Res.,* **80,** 320–327.

Cartwright, D. E., and R. D. Ray, 1994: On the radiational anomaly in the global ocean tide, with reference to satellite altimetry. *Oceanol. Acta,* **17,** 453–459.

Craven, P., and G. Wahba, 1979: Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.,* **31,** 377–403.

Di Lorenzo, E., W. R. Young, and S. Llewellyn Smith, 2006: Numerical and analytical estimates of M_2 tidal conversion at steep oceanic ridges. *J. Phys. Oceanogr.,* **36,** 1072–1084.

Dushaw, B. D., 2002: Mapping low-mode internal tides near Hawaii using TOPEX/Poseidon altimeter data. *Geophys. Res. Lett.,* **29,** 1250, doi:10.1029/2001GL013944.

Egbert, G. D., and R. D. Ray, 2001: Estimates of M2 tidal energy dissipation from TOPEX/Poseidon altimeter data. *J. Geophys. Res.,* **106** (C10), 22 475–22 502.

——, and S. Y. Erofeeva, 2002: Efficient inverse modeling of barotropic ocean tides. *J. Atmos. Oceanic Technol.,* **19,** 183–204.

——, and R. D. Ray, 2003: Semi-diurnal and diurnal tidal dissipation from TOPEX/Poseidon altimetry. *Geophys. Res. Lett.,* **30,** 1907, doi:10.1029/2003GL017676.

——, A. F. Bennett, and M. Foreman, 1994: TOPEX/Poseidon tides estimated using a global inverse model. *J. Geophys. Res.,* **99,** 24 821–24 852.

——, Ray, R. D., and Bills, B. G., 2004: Numerical modeling of the global semidiurnal tide in the present day and in the last glacial maximum. *J. Geophys. Res.,* **109,** C03003, doi:10.1029/2003JC001973.

Golub, G., and C. Van Loan, 1989: *Matrix Computations.* 2d ed. Johns Hopkins University Press, 642 pp.

Gustafson, K. E., 2001: Comparison of the energy flux to mixing processes via baroclinic wave drag on barotropic tides. *Deep-Sea Res.,* **48,** 2283–2296.

Jayne, S. R., and L. St. Laurent, 2001: Parameterizing tidal dissipation over rough topography. *Geophys. Res. Lett.,* **28,** 811–814.

Jeffreys, H., 1920: Tidal friction in shallow seas. *Philos. Trans. Roy. Soc. London,* **A221,** 239–264.

Kang, S. K., M. G. Foreman, W. R. Crawford, and J. Y. Cherniawsky, 2000: Numerical modeling of internal tide generation along the Hawaiian Ridge. *J. Phys. Oceanogr.,* **30,** 1083–1098.

Khatiwala, S., 2003: Generation of internal tides in an ocean of finite depth: Analytical and numerical calculations. *Deep-Sea Res. I,* **50,** 3–21.

Klymak, J. M., and Coauthors, 2006: An estimate of tidal energy lost to turbulence at the Hawaiian Ridge. *J. Phys. Oceanogr.,* **36,** 1148–1164.

Llewellyn Smith, S. G., and W. R. Young, 2002: Conversion of the barotropic tide. *J. Phys. Oceanogr.,* **32,** 1554–1566.

Merrifield, M. A., and P. E. Holloway, 2002: Model estimates of M2 internal tide energetics at the Hawaiian Ridge. *J. Geophys. Res.,* **107,** 3179, doi:10.1029/2001JC000996.

JUNE 2006 ZARON AND EGBERT 1035

Munk, W. H., and C. Wunsch, 1998: Abyssal recipes II: Energetics of tidal and wind mixing. *Deep-Sea Res.,* **45,** 1977–2010.

Parker, R. L., and L. Shure, 1982: Efficient modeling of the earth's magnetic field with harmonic splines. *Geophys. Res. Lett.,* **9,** 812–815.

Petrelis, F., S. Llewellyn Smith, and W. R. Young, 2003: Tidal conversion at a submarine ridge. *J. Phys. Oceanogr.,* **36,** 1053–1071.

Ray, R. D., and G. T. Mitchum, 1996: Surface manifestation of internal tides generated near Hawaii. *Geophys. Res. Lett.,* **23,** 2101–2104.

——, and D. E. Cartwright, 2001: Estimates of internal tide energy fluxes from Topex/Poseidon altimetry: Central North Pacific. *Geophys. Res. Lett.,* **28,** 1259–1262.

Rudnick, D. L., and Coauthors, 2003: From tides to mixing along the Hawaiian Ridge. *Science,* **301,** 355–357.

Simmons, H. L., R. W. Hallberg, and B. K. Arbic, 2004: Internal wave generation in a global baroclinic tide model. *Deep-Sea Res. II,* **51,** 3043–3068.

Sjöberg, B., and A. Stigebrandt, 1992: Computations of the geographical distribution of the energy flux to mixing processes via internal tides and the associated vertical circulation in the ocean. *Deep-Sea Res.,* **39,** 269–291.

Smith, W., and D. T. Sandwell, 1997: Global sea floor topography from satellite altimetry and ship depth soundings. *Science,* **277,** 1956–1962.

St. Laurent, L., and C. Garrett, 2002: The role of internal tides in mixing the deep ocean. *J. Phys. Oceanogr.,* **32,** 2882–2899.

——, S. Stringer, C. Garrett, and D. Perrault-Joncas, 2003: The generation of internal tides at abrupt topography. *Deep-Sea Res. I,* **50,** 987–1003.

Wahba, G., 1990: *Spline Models for Observational Data.* SIAM Publications, 169 pp.