JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 99, NO. C12, PAGES 24,821–24,852, DECEMBER 15, 1994

TOPEX/POSEIDON tides estimated using a global inverse model

Gary D. Egbert and Andrew F. Bennett
College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis

Michael G.G. Foreman
Institute of Ocean Sciences, Sidney, British Columbia

Abstract. Altimetric data from the TOPEX/POSEIDON mission will be used for studies of global ocean circulation and marine geophysics. However, it is first necessary to remove the ocean tides, which are aliased in the raw data. The tides are constrained by two distinct types of information: the hydrodynamic equations which the tidal fields of elevations and velocities must satisfy, and direct observational data from tide gauges and satellite altimetry. Here we develop and apply a generalized inverse method, which allows us to combine rationally all of this information into global tidal fields best fitting both the data and the dynamics, in a least squares sense. The resulting inverse solution is a sum of the direct solution to the astronomically forced Laplace tidal equations and a linear combination of the representers for the data functionals. The representer functions (one for each datum) are determined by the dynamical equations, and by our prior estimates of the statistics of errors in these equations. Our major task is a direct numerical calculation of these representers. This task is computationally intensive, but well suited to massively parallel processing. By calculating the representers we reduce the full (infinite dimensional) problem to a relatively low-dimensional problem at the outset, allowing full control over the conditioning and hence the stability of the inverse solution. With the representers calculated we can easily update our model as additional TOPEX/POSEIDON data become available. As an initial illustration we invert harmonic constants from a set of 80 open-ocean tide gauges. We then present a practical scheme for direct inversion of TOPEX/POSEIDON crossover data. We apply this method to 38 cycles of geophysical data records (GDR) data, computing preliminary global estimates of the four principal tidal constituents, M_2, S_2, K_1, and O_1. The inverse solution yields tidal fields which are simultaneously smoother, and in better agreement with altimetric and ground truth data, than previously proposed tidal models. Relative to the "default" tidal corrections provided with the TOPEX/POSEIDON GDR, the inverse solution reduces crossover difference variances significantly (≈20–30%), even though only a small number of free parameters (≈1000) are actually fit to the crossover data.

1. Introduction

Almost 30 years ago, in their discussion of the application of new time series methods to tidal spectroscopy, *Munk and Cartwright* [1966, p. 536] admitted "it can be said that we are here attempting to improve the one geophysical prediction that works tolerably well already; to this charge we plead guilty." A similar plea must be entered here. Of the many factors which must be considered and accounted before satellite altimetric data can be usefully applied to studies of global ocean circulation or marine geophysics, the ocean tides are perhaps the best understood. Certainly among all altimetric corrections, the tides have the longest and most extensive history as a subject of scientific research (see *Cartwright* [1977] and *Hendershott* [1981] for reviews). However, the tides also represent far and away the largest single correction to the altimetry data. In an analysis of Geosat data, *Ray et al.* [1991] (see also *Ray* [1993]) found that ocean tidal corrections represented 82% (94% if solid Earth tides are included) of the total reduction in variance in collinear differences, resulting from all standard corrections. The sheer size of the tidal correction thus places severe demands on the relative accuracy of tidal models. In a review of the situation on the eve of the TOPEX/POSEIDON (T/P) launch, *Ray* [1993] concluded that no available global ocean tide model was sufficiently accurate to satisfy mission goals. Thus while the physics of ocean tides is comparatively well understood, and while numerical models which reproduce most of the principal tidal features have been available for some time, there is still a great need for significant improvement in the modeling of global ocean tides.

The ocean tides are constrained by two distinct types of information. First, we have the laws of physics (i.e., the continuum equations for momentum and mass conservation, together with the astronomical tide-generating

Paper number 94JC01894
0148-0227/94/94JC-01894$05.00

force), which completely determine the tidal elevations and currents, at least in principal. In practice, use of this sort of information requires us to approximate the dynamical equations and seek solutions on a numerical grid. The linearized shallow water equations are typically assumed for tidal modeling [*Hendershott*, 1981], with parameterized corrections for dissipation, tidal loading, and ocean self-attraction. In this approximate form the dynamics obviously do not exactly determine the ocean tides. Early numerical modeling efforts led to solutions which exhibited only qualitative agreement amongst themselves and with the available observations [*Hendershott*, 1977]. More recent modeling efforts for individual ocean basins, using variable-mesh finite element schemes and more complex parameterizations of dissipation, demonstrate that numerical calculations can yield substantially more accurate tidal solutions [*Le Provost and Vincent*, 1991]. However, even with these more sophisticated modeling approaches, the discrepancy between solution and observations far exceeds the 2-cm accuracy goal of *Koblinsky et al.* [1992]. Indeed, the root-mean-square (RMS) misfit for the principal lunar semidiurnal constituent (M_2) in the Indian Ocean is about 6 cm [*Le Provost and Lyard*, 1993], comparable to the expected magnitude of the oceanographic mesoscale signal of interest (5-10 cm).

Estimates of the tides should also be guided by observational data. Until recently these data were limited to sea surface elevations measured by coastal gauges and a very small number of widely separated pelagic tide gauges. By themselves, these data are of limited use for estimating open-ocean tides. To incorporate these data into a tidal model, *Schwiderski* [1978] developed a "hydrodynamic interpolation" scheme, in which the estimated harmonic constants were used as boundary conditions (at the tide gauge locations) in a numerical solution of the dynamical equations. Imposing these extra boundary conditions required local adjustment of the bottom-friction coefficient and relaxation of the no-flow boundary conditions at the coast. It is a testament to the value of this crude approach (which inserts observational data into the dynamics) that this tidal model has been the standard to which others have been compared for more than a decade [*Ray*, 1993].

With the advent of satellite altimetry, globally distributed measurements of sea level have become available [*Cartwright*, 1991], and with the approach developed by *Cartwright and Ray* [1990, 1991] (hereinafter referred to as CR), direct empirical estimates of tidal constituents across the ocean are now possible. Their approach makes possible the construction of global fields of tidal constituents, without using the dynamical information in any way. Such fields are especially valuable for validating other tidal solutions, for they are not biased by any dynamical assumptions. However, as the success of *Schwiderski*'s [1980a, b] model suggests, there is much useful information in both the dynamics and the data. If the goal is to produce the most accurate estimate of the global ocean tides, all of the available information, both data and dynamics, should be used.

Inverse methods provide a direct approach to this problem, allowing us to seek solutions which fit both data and dynamics "well enough" [e.g., *Bennett*, 1992]. These ideas were first discussed explicitly in the context of tidal problems by *Bennett and McIntosh* [1982] and by *McIntosh and Bennett* [1984], who used inverse methods to construct a regional scale solution for the tides in Bass Strait. *Zahel* [1991] has subsequently applied a similar approach to tidal inversions on a global scale. The hydrodynamical interpolation scheme of *Schwiderski* [1978] could also be interpreted, in a very loose sense, as an inverse method. To accommodate the data, *Schwiderski* [1978] adjusted his a priori assumptions about the dynamics. However, a more formal treatment of the inverse problem offers many advantages. The trade-off between fitting the data and satisfying the dynamics is made explicit, and realistic a priori weighting schemes for data and dynamics can be incorporated. Using formal inversion methods, it is straightforward to incorporate data of different types, and of varying quality. For instance, as we show below, we can directly invert time domain altimetry data, yielding spatially smooth frequency domain tidal fields for all constituents simultaneously. We could also incorporate harmonically analyzed tide gauge data and current meter data in the same inverse calculation. Furthermore, with the inverse approach developed in this paper we have complete control over conditioning, so that we can minimize the degree to which noise (or nontidal oceanography) in the data is propagated into our tidal solution. Finally, the inverse formalism provides a natural framework in which to analyze the stability of the solution with respect to the data.

The paper is organized as follows. In section 2 we give a general overview of the data, the assumed dynamics, and our general approach to the inverse problem. In section 3 we consider the comparatively simple case of inverting dynamics and harmonically analyzed data for a single tidal constituent. We illustrate these initial results by inverting harmonic constants from a set of 80 open-ocean tide gauges [*Ray*, 1993]. Finally, in section 4 we build on the results of section 3 to develop a practical scheme for direct inversion of T/P data for multiple tidal constituents. We apply this method to 38 cycles of TOPEX data, computing preliminary global estimates of the four principal tidal constituents (M_2, S_2, K_1, and O_1) on a 512×256 grid (≈80×65 km at the equator).

We are currently extending and refining our calculations to include additional tidal constituents, coastal tide gauges, and pelagic current meter data, in an inverse solution on a 1024×512 grid. Results of these calculations, together with a more detailed analysis of posterior errors (in the data and the dynamics) will be given elsewhere.

2. The Tidal Inverse Problem

To set notation we begin with a general formulation of the tidal inverse problem, defining the state space and its relation to observable data. We then set out the hydrodynamic model which forms the basis of our prior information concerning the tides. We close this section with a summary overview of our general approach to the tidal inverse problem, based on explicit calculation of the representers of the data functionals [e.g., *Bennett*, 1992]. For this initial overview we keep the discussion fairly

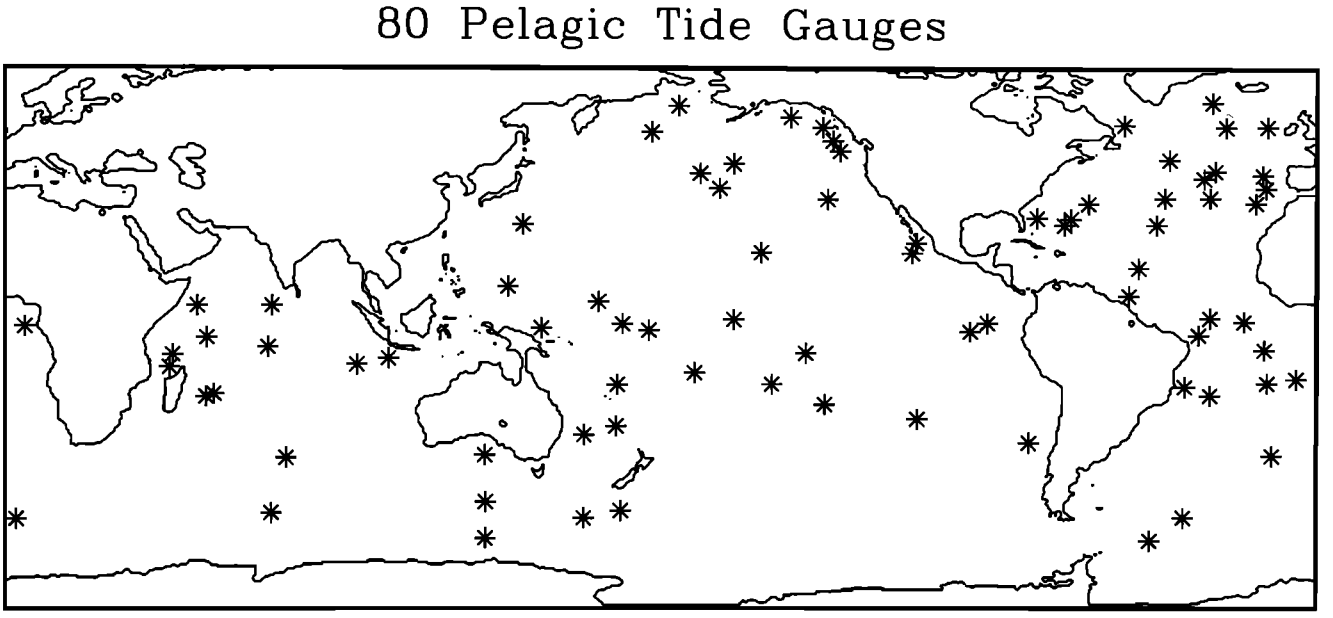

Figure 1. Locations of the 80 pelagic and island tide gauges of *Ray* [1993] used for the initial inversion.

general. Further details are provided in the next two sections where we treat two specific examples: inversion of data from a set of 80 pelagic tide gauges [*Ray*, 1993] (Figure 1), and inversion of the initial T/P altimetry data. We will be rather casual about technical mathematical details and only state the necessary results from generalized inverse theory. Where possible we use the notation of *Bennett* [1992], who gives a systematic development of the theory used here.

The State Space and the Observable Data

In general, our model of the ocean tide will be limited to a finite number L of principal harmonic constituents, with frequencies ω_l, $l = 1, L$. These are also designated traditionally by M_2, S_2, K_1, O_1 etc. For constituent l, the ocean tide is described by the complex amplitudes of the zonal and meridional components of volume transport u^l and v^l, respectively, plus sea level elevation h^l. These quantities are all complex-valued functions of position $\mathbf{x}$. The complete description of the tidal state is thus given by the complex $3L$-dimensional vector field

$$\mathbf{u}(\mathbf{x}) = \begin{bmatrix} \mathbf{u}^1(\mathbf{x}) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}^L(\mathbf{x}) \end{bmatrix} \quad \text{where} \quad \mathbf{u}^l(\mathbf{x}) = \begin{bmatrix} u^l(\mathbf{x}) \\ v^l(\mathbf{x}) \\ h^l(\mathbf{x}) \end{bmatrix}, \quad (1)$$

from which observable tidal quantities can be computed. For example, the sea level which would be observed by a tide gauge is given as a function of universal time and position by

$$h(\mathbf{x}, t) = \mathrm{Re}[\sum_{l=1}^{L} h^l(\mathbf{x}) e^{i\omega_l(t-t_0)+V_l(t_0)}], \quad (2)$$

where $V_l(t_0)$ is the astronomical argument at time t_0 for constituent l. We denote the full space of possible tidal states by τ; see Appendix A for a more precise definition. For any $\mathbf{u} \in \tau$ we will always refer to the three-component vector associated with a single tidal constituent using the superscript l, that is $\mathbf{u}^l$. Note also that for some purposes it will be useful to treat the real and imaginary parts of u^l, v^l and h^l separately, and represent tidal states explicitly as $6L$-dimensional real vector fields. We will denote the corresponding real vector state space as τ_R.

For tide gauge data at a fixed $\mathbf{x}$ the individual constituents $h^l(\mathbf{x})$ in (2) can be separated by harmonic analysis, provided a long enough time series is available. For shorter time series, and especially for satellite altimetry data, a full harmonic analysis of all significant constituents may not be feasible. In this case a slight modification of (2) is useful. Assuming that the oceanic response to the tide-generating potential (i.e., the tidal admittance; see *Munk and Cartwright* [1966]) varies smoothly with frequency, additional minor constituents can be (at least approximately) taken into account in the truncated harmonic expansion of (2) without increasing L. The simplest application of this idea is the inclusion of nodal corrections, which allow for the slow modulation of the astronomical forcing. With this correction (2) becomes

$$h(\mathbf{x}, t) = \mathrm{Re}[\sum_{l=1}^{L} \alpha_l(t) h^l(\mathbf{x})], \quad (3)$$

where $\alpha_l(t)$ is the exponential function of (2), with long-period (dominantly 18.6 years) modulations in amplitude and phase [*Cartwright and Tayler*, 1971] (Appendix B). From a frequency-domain perspective, (3) essentially amounts to including additional satellite constituents, under the assumption that the admittance is essentially constant in the neighborhood of ω_l. By smoothly interpolating the admittance between major constituents, this idea can be extended to include additional minor constituents at more distant frequencies [e.g., *Le Provost et al.*, 1991]. In this case $h(\mathbf{x}, t)$ is

still expressed as in (3) as a linear combination of the major complex constituent amplitudes $h^l(\mathbf{x})$, but the time dependent coefficients $\alpha_l(t)$ are more complicated, depending on the number and form of basis functions used for interpolation (e.g., see *Cartwright et al.* [1988] and *Le Provost et al.* [1991] for examples).

For the initial illustrative applications discussed in this paper we focus on the four dominant tidal constituents (M_2, S_2, K_1, and O_1), so that $L = 4$. For the direct inversion of T/P crossover differences discussed in section 4 we use a simple linear interpolation of the admittance in frequency to approximately include thirteen additional tidal constituents. The coefficients $\alpha_l(t)$ appropriate for this model (which also incorporates nodal corrections) are given in Appendix B. In future work we intend to expand the state space to include the additional constituents N_2, K_2, Q_1, and P_1 directly (so that $L = 8$), with minor constituents included with a more complicated interpolation scheme.

With altimetric observations of sea level, there are two further complications which must be addressed. First, the altimeter observes the geocentric tide $h'(\mathbf{x}, t)$. Due to the elastic yielding of the Earth, this differs from the Earth-relative (water-depth) tide $h(\mathbf{x}, t)$ measured by a conventional tide gauge. The levels h and h' are related via [e.g., *Ray and Sanchez*, 1989]

$$h' = G_r * h + h_b \quad , \tag{4}$$

where the convolution operator $G_r *$ corrects sea level for tidal loading effects on the elastic Earth [*Farrell*, 1972; *Francis and Mazzega*, 1990], and h_b is the solid Earth body tide. The level h_b is readily calculated in terms of the equilibrium tides [e.g., *Ray and Sanchez*, 1989] and is provided as a data item on the T/P geophysical data record (GDR). The solid Earth tide can thus be easily removed from the altimetric data. The convolution in (4) is reasonably well approximated by the simple scalar multiplication $G_r * h^l \simeq \beta_r^l h^l$, with β_r^l equal to 0.954 for semidiurnal constituents, or 0.940 for diurnal constituents [*Ray and Sanchez*, 1989]. We have made this simplifying approximation for the initial tidal inversions discussed here. A more sophisticated convolution treatment of load tides could be readily incorporated into the definition of the measurement functionals; see below. However, the value of this complication is questionable, particularly if our primary goal is to develop a tidal model for correcting altimetric data. In this case the simple scalar correction factor can be used consistently both for estimating the tidal state $\mathbf{u}$, and for calculation of the corrections $h''(\mathbf{x}, t)$ which will be applied to the altimetry data. With the solid Earth tide corrections, the geocentric sea level observed by the altimeter is related to the tidal state via

$$h''(\mathbf{x}, t) \equiv h'(\mathbf{x}, t) - h_b(\mathbf{x}, t) = \mathrm{Re}[\sum_{l=1}^{L} \alpha_l(t) G_r * h^l(\mathbf{x})] \simeq \mathrm{Re}[\sum_{l=1}^{L} \alpha_l(t) \beta_r^l h^l(\mathbf{x})] \, . \tag{5}$$

Second, to eliminate the effect of inaccuracies in the geoid (which are quite large) it is useful to consider differences formed between altimetric data at the same location, but at different times. There are numerous ways to do this. Here we focus on crossover differences:

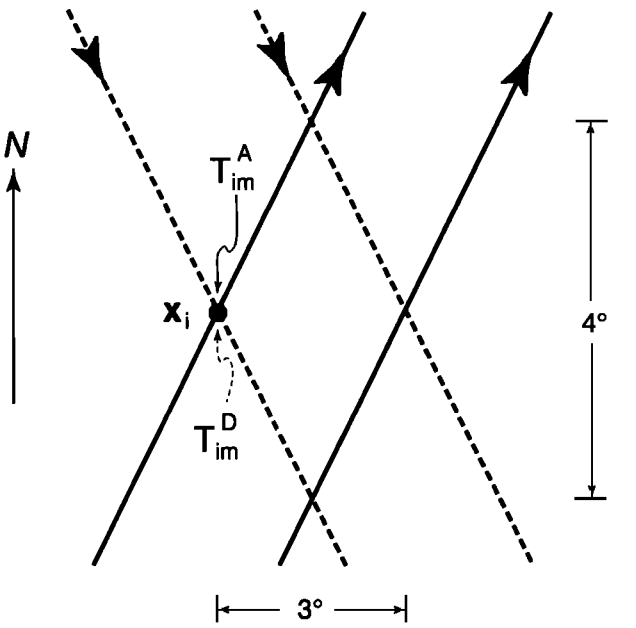

Figure 2. Altimetric crossover differences are computed for orbit cycle m at points $\mathbf{x}_i$ where ascending and descending ground tracks cross, at times T_{im}^A and T_{im}^D, respectively.

see Figure 2. A large number of locations $\mathbf{x}_i$, $(1 \le i \le I)$ are overflown twice during each of M 10-day orbit cycles. In the figure the first overflight occurs on a "descending" orbit during orbit cycle m (the satellite is traveling from north to south) at time T_{im}^D. The second occurs on an "ascending" orbit at time T_{im}^A. The crossover differences, with solid Earth tides removed, can then be represented as

$$d_{im} = h''(\mathbf{x}_i, T_{im}^A) - h''(\mathbf{x}_i, T_{im}^D) = \mathrm{Re}[\sum_{l=1}^{L} a_{iml} h^l(\mathbf{x}_i)] + \varepsilon_{im} \tag{6}$$

where $a_{iml} = \beta_r^l[\alpha_l(T_{im}^A) - \alpha_l(T_{im}^D)]$, and we have added an error term ε_{im} to represent all nontidal sources of signal and noise. We take (6) as the fundamental relation between the altimeter data (crossover differences, corrected for solid Earth body tides) and the state space (complex tidal amplitudes for L major tidal constituents). Other sorts of along-track or crossover differences can clearly be expressed in the same general form.

Note that the model used by CR for harmonic analysis of Geosat altimetry data is qualitatively similar to (6). CR used least squares to estimate the coefficients of an "orthotide" expansion of the tidal admittance [*Groves and Reynolds*, 1975] which best fit all along-track differences in $1^\circ \times 1.5^\circ$ bins. We refer to this model in the following as CR91. However, the approach used by CR did not take advantage of the hydrodynamics or smoothness of the ocean tides in any way. Spatial smoothing of local estimates of tidal amplitude and phase must be accomplished as a separate step [*Cartwright*, 1991]. In contrast, with the inverse approach developed here, no preliminary harmonic analysis is required. Optimal, dynamically consistent (in a sense to be made precise below), spatially smooth complex amplitudes $h^l(\mathbf{x})$ are fit directly to the time domain crossover data using (6)

In this paper we consider both harmonically analyzed tide gauge data and T/P crossover differences. In both cases the data can be viewed as the output of a linear functional acting on elements in the state space τ (or τ_R), with additive measurement errors

$$d_k = L_k[\mathbf{u}] + \varepsilon_k \ . \tag{7}$$

Other forms of tidal data, such as velocities obtained from deep-sea moorings [*Luyten and Stommel*, 1991], sea-floor electric field measurements [*Filloux et al.*, 1991], or reciprocal shooting using acoustic tomography, may also be expressed in this form [*Bennett*, 1985]. For the general development in the remainder of this section we take the full data set to be a vector $\mathbf{d}$ of length K which satisfies

$$\mathbf{d} = L[\mathbf{u}] + \boldsymbol{\varepsilon} \ , \tag{8}$$

without specific reference to the exact form of the data functionals L_k.

The Hydrodynamic Model

To a good approximation the tidal fields $\mathbf{u}^l = (u^l \ v^l \ h^l)^T$ at frequency ω_l satisfy the linearized shallow water equations, corrected for the effects of ocean self-attraction and tidal loading [e.g., *Hendershott*, 1981]:

$$\mathbf{S}_l \mathbf{u}^l = \mathbf{f}_0^l \quad \text{in } O \ , \tag{9}$$

subject to boundary conditions on ∂O. Here $\mathbf{f}_0^l$ is the astronomical forcing for constituent l (corrected for solid Earth tides), and the differential operator $\mathbf{S}_l$ is given by

$$\mathbf{S}_l = \begin{bmatrix} \mathbf{F}_l & gHG^*\nabla \\ \nabla\cdot & i\omega_l \end{bmatrix} \quad \mathbf{F}_l = \begin{bmatrix} i\omega_l+\kappa & -f \\ f & i\omega_l+\kappa \end{bmatrix} . \tag{10}$$

The domain O represents (some portion of) the global oceans, $H = H(\mathbf{x})$ is the ocean depth, f is the Coriolis acceleration, ω_l is the constituent frequency, ∇ and $\nabla\cdot$ represent the two-dimensional gradient and divergence operators on the spherical Earth, and κ represents dissipation. G^* represents convolution with the Green's function for tidal loading and ocean self-attraction [*Hendershott*, 1972]. Here we will adopt a simplified treatment of these effects, replacing the convolution operator G^* with a simple scalar correction factor $\beta = 0.9$ [*Accad and Pekeris*, 1978; *Schwiderski*, 1980a, b]. This approximation is justified by the estimates of dynamical errors given below. Note that G, and the scalar approximation β, are not the same as G_r and β_r used in (5) (which include only elastic deformation of the solid Earth due to water pressure). Note also that this approximation in the dynamics is independent of the approximation to the data functionals in (5). We could use the full convolution in the definition of the altimetric data functionals while still adopting an approximate treatment of load tides in the dynamics. Finally, note that by employing volume transports u^l and v^l, the bathymetry appears only in the momentum equations.

Boundary conditions are no flow normal to the coast:

$$\hat{\mathbf{n}}\cdot\mathbf{u}^l = 0 \quad \text{on } \partial O_c \ , \tag{11}$$

and specified elevations on open boundaries (if appropriate):

$$h^l = h_0^l \quad \text{on } \partial O_o \ . \tag{12}$$

In (11) $\hat{\mathbf{n}} = (n_u \ n_v \ 0)$ is the three vector having as first two components the outward-directed unit normal at the boundary. We will refer to this system of equations and boundary conditions generically as the Laplace tidal equations (LTEs) for constituent l. For the full set of LTEs (including boundary conditions), considering all L constituents simultaneously, we use the abbreviated notation

$$\mathbf{S}\mathbf{u} = \mathbf{f}_0 \ . \tag{13}$$

For the numerical solutions of of the LTEs discussed in this paper, the bathymetry H is taken from the ETOPO5 data base [*National Geophysical Data Center*, 1992] with the boundaries of the ocean domain O defined by the 10 m depth contour. Also, to avoid complications with the singularity of the spherical coordinate system at the poles we restrict O to the latitude range $80°S$ to $70°N$. Thus most of the Arctic Ocean is excluded from our model. Note that the domain includes all of the ocean overflown by the T/P mission, which is restricted to latitudes below 70°. Open boundary conditions (12) are used along the northern edge of the domain in the Arctic ocean, with the specified elevations h_0^l taken from the model "SCH80" of *Schwiderski* [1980a, b, 1981] (see also *Schwiderski and Szeto* [1981]). For the forcing $\mathbf{f}_0$ we used astronomical potentials given by *Cartwright* [1977], with corrections for solid Earth tide potentials calculated using the frequency-dependent Love numbers of *Wahr* [1981].

Dissipation is parameterized simply as

$$\kappa = \kappa_0/\max[H, H_0] \ . \tag{14}$$

To choose κ_0 and H_0 we solved the LTEs numerically for a range of parameter values and compared the elevation component of solutions for the four dominant constituents (M_2, S_2, K_1, and O_1) to tidal constants from the 80 pelagic gauges of Figure 1, and to SCH80. Based on these comparisons, we chose $\kappa_0 = 0.03$ m s^{-1} and $H_0 = 200$ m. Experimentation with other simple parameterizations for dissipation (e.g., the quadratic (in H) expression of *Zahel* [1977]) produced similar results. Although more refined treatments of dissipation are possible [e.g., *Le Provost*, 1977; *Le Provost and Poncet*, 1978], we believe that this simple linear approach is most appropriate for the linear inversion method discussed here, where the dynamics are used only as weak constraints. For the same reason we also omit horizontal viscosity from the tidal equations and use the scalar approximation for tidal loading and self-attraction.

The Penalty Functional

We have two sorts of information concerning the true tidal state $\mathbf{u}_{\text{True}}$: the hydrodynamics and the data. Together these can be expressed as

$$\begin{bmatrix} \mathbf{S} \\ L \end{bmatrix} \mathbf{u} = \begin{bmatrix} \mathbf{f}_0 \\ \mathbf{d} \end{bmatrix} . \tag{15}$$

While it is in theory possible to find an exact (and unique) solution $\mathbf{u}_0$ to the hydrodynamic system $\mathbf{Su} = \mathbf{f}_0$, in general no tidal state will satisfy exactly the full set of equations (15). More formally, the operator formed by combining the data functionals and the hydrodynamic equations is singular. The generalized inverse method essentially amounts to constructing a generalized inverse of this singular operator [*Reid*, 1968].

Because no $\mathbf{u}$ exactly satisfies (15), this inverse calculation must be recast as a fitting problem, in which we try to satisfy both the data constraints and the hydrodynamic constraints "well enough." Observational data such as the T/P crossover differences will always include nontidal signals and noise, so the data should not be fit exactly. Indeed, observing the nontidal signal is the object of the T/P mission; discovering a stable perfect fit to the T/P data using only a tidal state would be unfortunate. Moreover, because of the approximate parameterizations of dissipation and load tides, errors in the bathymetry, and the practical requirement that the system be solved numerically on a discrete grid, we should not expect the true tides to satisfy exactly the assumed hydrodynamic equations. Letting $\mathbf{u}_{\text{True}}$ be the (unknown) true tidal state, we can define

$$\boldsymbol{\varepsilon} = \mathbf{d} - L[\mathbf{u}_{\text{True}}] \qquad \delta\mathbf{f} = \mathbf{f}_0 - \mathbf{S}\mathbf{u}_{\text{True}}$$

to be the measurement errors and the dynamical errors, respectively.

While we do not know and cannot directly measure $\boldsymbol{\varepsilon}$ and $\delta\mathbf{f}$, it is our conviction (hypothesis?) that we can at least approximately characterize these errors in a statistical sense. We thus assume that the errors are realizations of random processes satisfying

$$E\,\boldsymbol{\varepsilon} = 0 \qquad E\,\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^* = \mathbf{C}_\varepsilon \tag{16}$$

$$E\,\delta\mathbf{f} = 0 \qquad E\,\delta\mathbf{f}(\mathbf{x}_1)\delta\mathbf{f}(\mathbf{x}_2)^* = \mathbf{C}_f(\mathbf{x}_1, \mathbf{x}_2) \tag{17}$$

$$E\,\boldsymbol{\varepsilon}\delta\mathbf{f}^* = 0 \tag{18}$$

where E denotes the ensemble average or mathematical expectation, and the asterisk the complex conjugate transpose. Note that the forcing error covariance $\mathbf{C}_f$, is a $3L \times 3L$ complex matrix-valued function of pairs of points $\mathbf{x}_1$, $\mathbf{x}_2$ in, or on the boundary of, the ocean domain O. Note also that $\mathbf{C}_f$ defines a linear operator acting on tidal states $\mathbf{u}$:

$$[\mathbf{C}_f\,\mathbf{u}](\mathbf{x}) = \int_O \mathbf{C}_f(\mathbf{x}, \mathbf{x}')\mathbf{u}(\mathbf{x}')d^2\mathbf{x}'. \tag{19}$$

We will assume that this operator has inverse $\mathbf{C}_f^{-1}$ (ignoring nontrivial mathematical details!). Note that we shall not need actually to calculate this inverse. We defer discussion of the exact form of the assumed covariances to subsequent sections where we consider specific examples. For now note that the covariance operator $\mathbf{C}_f$ will generally act to smooth out or attenuate small-scale features of $\mathbf{u}$, while the inverse $\mathbf{C}_f^{-1}$ will have the opposite effect. Note also that these operators will be self-adjoint so that

$$\int_O \mathbf{v}(\mathbf{x})^*[\mathbf{C}_f\,\mathbf{u}](\mathbf{x})d^2\mathbf{x} = \int_O [\mathbf{C}_f\,\mathbf{v}](\mathbf{x})^*\mathbf{u}(\mathbf{x})d^2\mathbf{x}\,, \tag{20}$$

with a similar relation holding for $\mathbf{C}_f^{-1}$.

The covariances can be used both to normalize the two sources of error and to make precise the notion of fitting the data and the dynamics "well enough." We define the quadratic penalty functional

$$J[\mathbf{u}] = (L[\mathbf{u}]-\mathbf{d})^*\mathbf{C}_\varepsilon^{-1}(L[\mathbf{u}]-\mathbf{d}) + \int_O [\mathbf{Su} - \mathbf{f}_0](\mathbf{x})^*[\mathbf{C}_f^{-1}(\mathbf{Su} - \mathbf{f}_0)](\mathbf{x})d^2\mathbf{x}\,. \tag{21}$$

To solve the generalized inverse problem we minimize J. This solution, which represents a weighted least squares fit to the data and the hydrodynamics, is certainly intuitively appealing. Furthermore, it can be shown [*Bennett*, 1990] that this inverse solution is essentially the minimum variance linear unbiased estimator of $\mathbf{u}$ (i.e., the Gauss-Markov smoother of the tidal data). If both the dynamical errors and the data errors are Gaussian, the estimate is also maximum-likelihood [e.g., *Tarantola*, 1987]. Minimization of J can also be justified in terms of a Bayesian statistical model [e.g., *Tarantola*, 1987; *Backus*, 1988].

In theory minimization of J is straightforward. After discretization, (21) is basically just a large linear least squares problem for which direct solutions are readily available [e.g., *Tarantola*, 1987]. However, if the ocean tides are to be represented with a reasonable resolution, the number of unknown parameters becomes prohibitively large, making a direct application of standard least squares methods unfeasible. For instance, specification of $L = 8$ complex tidal constituents on a 1 degree grid (the resolution of SCH80) requires of the order of 10^6 real parameters. One approach would be to use some sort of descent method (e.g., conjugate gradients) to minimize J iteratively. However, with this approach it is still necessary to invert (the $10^6 \times 10^6$ matrix) $\mathbf{C}_f$. To make this feasible, previous workers have adopted very simple (and almost certainly unrealistic) diagonal form for the dynamical error covariance [*Jourdin et al.*, 1991; *Zahel*, 1991]. Furthermore, iterative minimization of the penalty functional allows only limited control over conditioning and hence little assurance that a stable fit has been obtained. We regard the stability of the fit to the prior information to be at least as important as the closeness-of-fit.

We adopt a different approach, based on the application of some simple Hilbert space theory, which has been used extensively in solid Earth geophysics [e.g., *Backus*, 1971; *Parker*, 1975; *Parker et al.*, 1987]. The method is developed in detail in an oceanographic context by *Bennett* [1992]. Here we only sketch the principal ideas. The approach massively reduces the apparent dimension of the inverse problem to its true dimension, namely, the number of independent data. Note that this reduction is effected prior to numerical discretization.

The Hilbert Space Approach

The dynamic error penalty in (21) is a positive-definite quadratic form which can be used to define an inner product on the state space τ:

$$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \int_O d^2\mathbf{x}\,[\mathbf{Su}_1](\mathbf{x})^*[\mathbf{C}_f^{-1}\mathbf{Su}_2](\mathbf{x})\,, \tag{22}$$

where the superscript asterisk denotes the complex con-

jugate transpose. Under weak assumptions on the covariance $\mathbf{C}_f$, it may be shown (see Appendix A) that our state space τ is a Hilbert space, and that for every reasonable data functional L_k (including pointwise evaluation of tidal components), there exists an element $\mathbf{r}_k$ in the state space τ which satisfies

$$L_k[\mathbf{v}] = \langle \mathbf{r}_k, \mathbf{v} \rangle , \tag{23}$$

for every $\mathbf{v} \in \tau$. The field $\mathbf{r}_k$ is called the representer of the data functional. It can be shown [e.g., *Wahba and Wendelberger*, 1980; *Parker et al.*, 1987; *Bennett*, 1990, 1992] that J attains its global minimum value for a unique element in the state space of the form

$$\hat{\mathbf{u}} = \mathbf{u}_0 + \sum_{k=1}^{K} b_k \mathbf{r}_k . \tag{24}$$

In particular, the entire infinite dimensional null space of unobservable additive corrections $\mathbf{g}$, where $L_k[\mathbf{g}] = \langle \mathbf{g}, \mathbf{r}_k \rangle = 0$, $1 \le k \le K$, is discarded. This is the first, massive step in reducing the apparent dimension of the inverse problem to its true value ($\le K$). The coefficients in the representer expansion $\mathbf{b}$ satisfy the $K \times K$ system of linear equations

$$(\mathbf{R} + \mathbf{C}_\varepsilon)\mathbf{b} = \mathbf{d} - L[\mathbf{u}_0] , \tag{25}$$

where $\mathbf{R}$ is the (Hermitian, positive-definite) representer matrix with elements

$$R_{jk} = \langle \mathbf{r}_j, \mathbf{r}_k \rangle = L_j[\mathbf{r}_k] = L_k[\mathbf{r}_j]^* . \tag{26}$$

Our approach to the inverse tidal problem is based on explicitly calculating the representers for the data functionals, forming the representer matrix, and then solving (25). Further details of these calculations will be given in the following sections. As we shall see, this calculation is still quite formidable. However, even for a very large data set, such as the T/P crossover differences for all orbit cycles, we can arrange things so that K is no more than about 10^5. This is already 1 order of magnitude smaller than the number of parameters needed to describe eight constituents on a 1 degree grid. Furthermore, as we shall show below, the representer analysis allows us to further reduce the size of the inverse problem, to an effective value of K of the order of $10^3 - 10^4$, without significantly sacrificing model resolution or unduly restricting our choice of dynamical error covariances. In addition, the representer theory leads naturally to a complete analysis of the stability and conditioning of the inverse solution. In particular, the posterior error covariance provides us with estimates of the magnitude and spatial structure of errors in the inverse solution [*Bennett*, 1992].

3. Inversion of Harmonically Analyzed Tide Gauge Data

In this section we discuss the somewhat simpler inverse problem appropriate for harmonically analyzed tide gauge data so that we can treat each tidal constituent separately. We take the data to be K point estimates of constituent l, obtained from harmonic analysis of tide gauge data at locations $\mathbf{x}_k$, $k = 1, K$

$$d_k = L_k[\mathbf{u}] + \varepsilon_k = h^l[\mathbf{x}_k] + \varepsilon_k . \tag{27}$$

The focus of this section is on calculation of representers $\mathbf{r}_k$ for the simple single-constituent data functionals of (27). Note that for time domain data we must consider multiple tidal constituents simultaneously, but representers can be constructed essentially as sums of these single-constituent representers. The results of this section thus provide a foundation for development of the altimeter crossover difference inversion which is our ultimate goal. As a more immediate illustration of the representer approach, we use the single-constituent representers to invert the $K = 80$ tidal constants from the pelagic gauges of Figure 1.

Differential Equations and Boundary Conditions for the Representers

As we consider only a single constituent at a time in this section (i.e., we take $L = 1$), we will omit the constituent identifier l from all quantities. To begin, we explicitly separate the inner product which defines the dynamical misfit penalty into interior and boundary terms:

$$\begin{aligned} \langle \mathbf{u}_1, \mathbf{u}_2 \rangle = & \int_O d^2\mathbf{x}\, [\mathbf{S}\mathbf{u}_1](\mathbf{x})^* [\mathbf{C}_i^{-1}\mathbf{S}\mathbf{u}_2](\mathbf{x}) \\ & + \int_{\partial O_c} dl\, \hat{\mathbf{n}}\cdot\mathbf{u}_1(\mathbf{x})^* [\mathbf{C}_c^{-1}\hat{\mathbf{n}}\cdot\mathbf{u}_2](\mathbf{x}) \\ & + \int_{\partial O_o} dl\, h_1(\mathbf{x})^* [\mathbf{C}_o^{-1}h_2](\mathbf{x}) . \end{aligned} \tag{28}$$

Thus the dynamical error covariance function is divided into three distinct parts: $\mathbf{C}_i$ which characterizes errors or inadequacies in the differential equation (9) in the interior of the domain O; $\mathbf{C}_c$ the covariance for errors in the no-flow boundary condition at the coast; and $\mathbf{C}_o$ the covariance for errors in the specified elevations on the open boundaries. Note that in (28) we have implicitly assumed that there is no correlation between the three sorts of errors. Indeed, we have no information about any such correlation.

The linear functionals of potential interest are averages over the interior domain O or as a point measurement on either the interior or boundary (tide gauges). These can all be expressed in the general form

$$L[\mathbf{u}] = \int_O d^2\mathbf{x}\, \boldsymbol{\mu}^*\cdot\mathbf{u} + \int_{\partial O} dl\, \boldsymbol{\nu}^*\cdot\mathbf{u} , \tag{29}$$

where we allow the interior and boundary measurement kernels $\boldsymbol{\mu} = (\mu_u, \mu_v, \mu_h)^T$ and $\boldsymbol{\nu} = (\nu_u, \nu_v, \nu_h)^T$ to be delta-type functions. Note that we explicitly allow separate interior ($\boldsymbol{\mu}$) and boundary ($\boldsymbol{\nu}$) parts for the linear functionals. This is consistent with the special treatment of the boundary in the definition of the penalty functional. Note also that with this general formulation we could use the full convolution form of the radial deformation Greens function (4) in our definition of the altimeter data functionals. We now derive an explicit form for the differential equations and boundary conditions which determine the representer $\mathbf{r} = (r_u\ r_v\ r_h)^T$ for a linear data functional of the general form (29).

We require for all $\mathbf{u} = (u\ v\ h)^T$

$$L[\mathbf{u}] =$$

$$\int_O d^2\mathbf{x}\, \mu(x)^* \cdot \mathbf{u}(\mathbf{x}) + \int_{\partial O} dl\ \nu(\mathbf{x})^* \cdot \mathbf{u}(\mathbf{x}) = \langle \mathbf{r}, \mathbf{u} \rangle$$

$$= \int_O d^2\mathbf{x}\, [\mathbf{S}\mathbf{r}](\mathbf{x})^* [\mathcal{C}_i^{-1}\mathbf{S}\mathbf{u}](\mathbf{x}) + \int_{\partial O_c} dl\ \hat{\mathbf{n}}\cdot\mathbf{r}(\mathbf{x})^* [\mathcal{C}_c^{-1}\hat{\mathbf{n}}\cdot\mathbf{u}](\mathbf{x})$$

$$+ \int_{\partial O_o} dl\ r_h(\mathbf{x})^* [\mathcal{C}_o^{-1} h](\mathbf{x}) . \qquad (30)$$

The last equality follows from the definition of the inner product.

Letting

$$\boldsymbol{\eta} = (\eta\ \chi\ \zeta)^T = \mathcal{C}_i^{-1}\mathbf{S}\mathbf{r} , \qquad (31)$$

using (20), and integrating by parts, the first term in (30) is

$$\int_O d^2\mathbf{x}\, [\mathbf{S}\mathbf{r}](\mathbf{x})^* [\mathcal{C}_i^{-1}\mathbf{S}\mathbf{u}](\mathbf{x}) = \int_O d^2\mathbf{x}\, \boldsymbol{\eta}(\mathbf{x})^* [\mathbf{S}\mathbf{u}](\mathbf{x}) \qquad (32)$$

$$= \int_O d^2\mathbf{x}\, [\mathbf{S}^\dagger\boldsymbol{\eta}](\mathbf{x})^* \cdot \mathbf{u}(\mathbf{x}) + \int_{\partial O} dl\ [\beta g H \hat{\mathbf{n}}\cdot\boldsymbol{\eta}^* h + \zeta^* \hat{\mathbf{n}}\cdot\mathbf{u}] .$$

Here

$$\mathbf{S}^\dagger = \begin{bmatrix} \mathbf{F}^\dagger & -\nabla \\ -\beta g \nabla\cdot H & -i\omega \end{bmatrix}$$

defines the adjoint of $\mathbf{S}$ where

$$\mathbf{F}^\dagger = \begin{bmatrix} -i\omega+\kappa & f \\ -f & -i\omega+\kappa \end{bmatrix} . \qquad (33)$$

Substituting (31) into (30) we have

$$\int_O d\mathbf{x}^2\, \boldsymbol{\mu}^* \cdot \mathbf{u} + \int_{\partial O} dl\ \nu^* \cdot \mathbf{u} = \int_O d^2\mathbf{x}\, (\mathbf{S}^\dagger\boldsymbol{\eta})^* \cdot \mathbf{u}$$

$$+ \int_{\partial O_c} \cdots + \int_{\partial O_o} \cdots + \int_{\partial O} dl\ [\beta g H \hat{\mathbf{n}}\cdot\boldsymbol{\eta}^* h + \zeta^* \hat{\mathbf{n}}\cdot\mathbf{u}] . \qquad (34)$$

Noting (31) it may be seen that (34) will hold identically for all $\mathbf{u}$, provided $\mathbf{r}$ and $\boldsymbol{\eta}$ satisfy the partial differential equations

$$\mathbf{S}^\dagger\boldsymbol{\eta} = \boldsymbol{\mu} \qquad (35)$$

$$\mathbf{S}\mathbf{r} = \mathcal{C}_f \boldsymbol{\eta} , \qquad (36)$$

subject to boundary conditions chosen to annihilate the boundary integrals in (34). Using the fact that the inverse of the covariance operators are self-adjoint, this condition requires that the integrand of

$$\int_{\partial O} dl\, [\beta g H \hat{\mathbf{n}}\cdot\boldsymbol{\eta}^* h + \zeta^* \hat{\mathbf{n}}\cdot\mathbf{u}] - \nu^* \cdot \mathbf{u} +$$

$$\int_{\partial O_c} dl\, [\mathcal{C}_c^{-1}\hat{\mathbf{n}}\cdot\mathbf{r}]^* \hat{\mathbf{n}}\cdot\mathbf{u} + \int_{\partial O_o} dl\, [\mathcal{C}_o^{-1} r_h]^* h \qquad (37)$$

vanish identically for all $\mathbf{u}$. On the rigid part of the boundary (∂O_c) this implies

$$\hat{\mathbf{n}}\cdot\boldsymbol{\eta} = \nu_h/\beta g H , \qquad (38)$$

$$\hat{\mathbf{n}}\cdot\mathbf{r} = \mathcal{C}_c(\hat{\mathbf{n}}\cdot\mathbf{v} - \zeta) , \qquad (39)$$

as boundary conditions for the backward and forward problems, respectively. On the open part of the boundary (∂O_o) the conditions are

$$\zeta = 0 , \qquad (40)$$

$$r_h = \mathcal{C}_o[\nu_h - \beta g H \hat{\mathbf{n}}\cdot\boldsymbol{\eta}] . \qquad (41)$$

The calculation of the representers is thus accomplished by first solving the adjoint, or backward, system (35) subject to the boundary conditions (38) and (40), for coastal and open boundary segments, respectively. Note that in the case of coastal tide gauges the rigid boundary conditions for the adjoint system are inhomogeneous and the forcing (on the interior) is zero, since for a tide gauge on the boundary at $\mathbf{x}_k$ the boundary kernel is $\nu_h(\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}_k)$, and the interior kernel is $\mu\equiv 0$. For most other cases of interest (pelagic tide gauges, altimetric data) the boundary conditions for the adjoint system are homogeneous, and the forcing in (35) is restricted to the adjoint variable ζ (e.g., for a pelagic tide gauge at $\mathbf{x}_k$, the interior and boundary kernels are $\mu_h(\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}_k)$, $\nu\equiv 0$, respectively). The above mentioned first step results in a solution $\boldsymbol{\eta} = (\eta, \chi, \zeta)$, which is then smoothed by the dynamical error covariances, yielding the inhomogeneous boundary conditions of (39) and (41) and also the forcing on the right hand side of (36). The resulting system of equations and boundary conditions (the forward equations) can then be solved to yield the representer $\mathbf{r}$.

We solve the system of equations (35-36) numerically on an Arakawa "C" grid by adopting a periodic forcing (of frequency ω), and then explicitly time-stepping forward from homogeneous initial conditions. A very simple harmonic analysis of the time dependent asymptotic solution yields the complex harmonic amplitudes [*McIntosh and Bennett*, 1984]. The solution to the adjoint equations is found in essentially the same way, but with the time-stepping proceeding backward from homogeneous final conditions. The linearity of the dynamics and the presence of (linear) drag assure that the periodic equilibrium states are independent of the choice of initial or final states. Note that we have chosen the coefficients in the finite difference approximations for these systems so that the discretized backward equations are the exact discrete adjoint of the discretized forward equations. This insures that the representer matrix will be Hermitian and positive definite, to within machine accuracy.

The representer calculation requires solution of both the forward and backward equations repeatedly, in principle once for each independent datum. However, the time-stepping algorithm is well suited to massively parallel processing. Hence we have implemented the algorithm in Fortran-90 on a Thinking Machines Corporation CM-200, to allow us to handle the very large T/P crossover data set.

Dynamical Errors

The error vector $\delta\mathbf{f}$ represents inadequacies in our hydrodynamic model and boundary conditions and in the numerical approximation of these equations on a discrete grid. Prior information about the magnitude and spatial

characteristics of these errors is provided through the dynamical or forcing error covariance $\mathbf{C}_f$. In conjunction with the data error covariance $\mathbf{C}_\varepsilon$, $\mathbf{C}_f$ determines the trade-off between fitting the observational data, and satisfying the hydrodynamic model. To some degree, $\mathbf{C}_f$ also controls the smoothness of the inverse solution. Inevitably, our model for the dynamical error covariance must represent a compromise between fidelity to the true situation (which is at best incompletely understood) and computational tractability. When in doubt, we opt for simplicity. In particular, as noted already, we take errors in the boundary conditions to be independent of errors in the dynamical equations. We thus consider $\mathbf{C}_i$, $\mathbf{C}_c$ and $\mathbf{C}_o$ separately in the following. Also, for now we treat only a single constituent at a time, ignoring any possible interconstituent correlation of dynamical errors. We will extend our dynamical error covariance model to allow for this complication in the next section, where we treat the more general multiple-constituent tidal inverse problem.

Dynamical Error Covariance $\mathbf{C}_i$

Because a more complicated model would be difficult to establish, we assume that the components of the dynamical error (two for the momentum equations, and one for the continuity equation) are uncorrelated. The dynamical error covariances for each component characterize both the expected magnitude and the smoothness, or spatial correlation, of the dynamical errors. Both of these characteristics of $\delta\mathbf{f}$ might realistically be expected to be spatially varying. As we show below, an analysis of approximations and errors inherent in the dynamical equations can be used to estimate the spatial dependence of the dynamical error variances. Comparable estimates of spatially varying (or even average) decorrelation length scales are not so obvious. We thus adopt a dynamical error covariance function which combines a simple spatially homogeneous and isotropic correlation with an inhomogeneous variance. We assume the covariance between $\delta\mathbf{f}(\mathbf{x})$ and $\delta\mathbf{f}(\mathbf{x}')$ has the form

$$[\mathbf{C}_i(\mathbf{x}, \mathbf{x}')]_{jj'} = \delta_{jj'}\sigma_j(\mathbf{x})\sigma_{j'}(\mathbf{x}')\psi(\cos\lambda) , \qquad (42)$$

where the index j (=1, 2, 3) refers to the component (u, v, or h) of the tidal fields, $\sigma_j^2(\mathbf{x})$ is the spatially varying dynamical error variance for component j, and $\psi(\cos\lambda)$ is the spatial correlation (on the sphere) of components of $\delta\mathbf{f}_i$ with angular separation λ. The correlation function ψ, which has a decorrelation length scale of approximately 5° is plotted in Figure 3.

It is difficult to provide a rigorous justification for the chosen decorrelation length scale, much less the exact form of ψ. However, we have to assume a certain degree of smoothness in the dynamical errors to make our inverse problem formally well posed [*Yosida*, 1980; *Bennett*, 1990] (Appendix A). This requirement is more than a mathematical technicality. In experiments with completely uncorrelated dynamical errors, we have found that the inverse solutions for tidal elevations contain localized peaks or holes at some tide gauge locations. This reflects the local singular behavior of the functions $\mathbf{r}_k$ computed with uncorrelated dynamical errors for point evaluation functionals [*Bennett*, 1990]. To suppress this unphysical behavior in the T/P crossover difference inversion, we have chosen the decorrelation length scale for the dynamical errors to be comparable to the typical crossover difference separation. This choice has about as mild an impact as is acceptable.

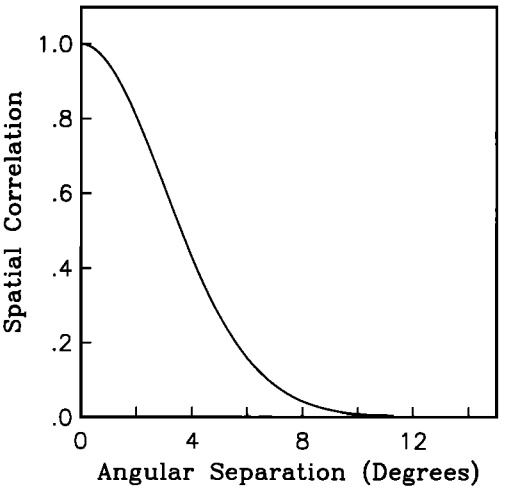

Figure 3. Spatial correlation ψ of dynamical errors as a function of angular separation on the sphere. The same correlation structure is assumed for all components of the dynamical errors, and all harmonic constituents. Smoothing with ψ is accomplished efficiently by repeated application of a local smoother (see Appendix C for details).

The functional form of ψ has been dictated by computational considerations. The covariance is used to smooth and scale the backward solution $\boldsymbol{\eta}$ (see (36)). Given our spatially homogeneous correlation, this requires a combination of scalar multiplication and convolution with the correlation function on the projected spherical grid. For numerical efficiency we use an iterative scheme which repeatedly smooths $\boldsymbol{\eta}$ locally, using a latitude-dependent set of weights. The result is identical to smoothing with a covariance which is essentially of the form given in (42). Note that this iterative implementation of the covariance smoother leads to some small deviations from exact spatial homogeneity of the spatial correlation near the coasts. This complication and computational details of our implementation of the covariance smoother on the CM-200 are discussed in Appendix C.

We turn now to a specification of the spatially inhomogeneous dynamical error variances. By defining the discretized variables (u, v) and h to represent, respectively, the total mass flux across the edges of a numerical grid cell, and the sea surface elevation (relative to the bottom) averaged over that cell, the exact conservation of mass is assured for the discretized system. We adopt this point of view, and take the forcing error covariance for the continuity equation to be identically zero. In making this assumption we are of course neglecting some small errors associated with stratification and compressibility of the ocean, and with the small-amplitude approximation. However, these errors are much smaller than other sources of error in the discretized time separated momentum conservation equations [*McIntosh and Bennett*, 1984]. For the zonal component

the relevant differential equation is

$$(i\omega + \kappa)u - fv + \beta gH\partial_x h = f_{u0} \qquad (43)$$

There are several distinct sources of error here. First, (43) approximates the physics in three important ways: (1) by replacing the full three-dimensional Navier-Stokes equations by the linearized shallow water equations, (2) by the crude parameterization of dissipative effects, and (3) by the approximate treatment of tidal loading and self-attraction. Next, the bathymetry $H(\mathbf{x})$ is only approximately known. Finally, we must solve (43) numerically on a discrete grid. For our analysis of these numerical errors we consider calculations on a 512×256 grid (≈78×65 km at the equator).

The approximate modeling of dissipation by (14) is undoubtedly the most serious source of error [*McIntosh and Bennett*, 1984]. A wide range of linear drag coefficients have been used in previous tidal modeling studies, with the preferred value depending on the scale of, and the range of depths encountered in, the model under consideration. See *Pekeris and Accad* [1969] for a discussion of this issue. *McIntosh and Bennett* [1984] assumed a drag coefficient of the form $\kappa = 0.002/H$ for a model of Bass Strait. For their regional scale model, with ocean depths of the order of 100 m, the major energy loss occurred through the open boundaries. The linear drag coefficient used by *Schwiderski* [1978] for the SCH80 global model, while rather complicated in detail, is generally much larger, being nominally of the form $\kappa \approx 0.01/H$. Our drag coefficient, which is even larger (except in shallow water; see (14)), is more comparable to that used by *Pekeris and Accad* [1969], who restricted their tidal model to the deep oceans ($H \geq 1000$ m). Based on this great variation in reported parameters, together with our conviction that the overall scale of dissipation used in our forward model is reasonably correct on average, we assume 100% errors in the dissipation term of (43). Thus calling this error δf_κ, we have

$$\mathrm{Var}(\delta f_\kappa) = |\kappa U|^2 = \left[\frac{\kappa_0 |U|}{\max[H(\mathbf{x}), H_0]}\right]^2, \qquad (44)$$

where U is a transport scale. This variance can be estimated numerically (as a function of position $\mathbf{x}$) using for U a prior estimate of the volume transports (e.g., the transport components of the prior model $\mathbf{u}_0$). Compared to a typical term in (43) (i.e., ωu), δf_κ represents approximately a 10% error in the dynamics in the deep ocean, increasing to 100% errors on the shelves (where $\kappa = 1.5\times10^{-4}\ \mathrm{s}^{-1} \approx \omega$).

The analysis of *McIntosh and Bennett* [1984] for regional scale tidal modeling shows that errors resulting from the linearized shallow water approximation are generally much smaller than likely errors in the parameterization of dissipative effects. This result also holds for the global scale model considered here. Again following *McIntosh and Bennett* [1984], relative errors from the numerical discretization of $\partial_x h$ should be $O(k^2\Delta x^2)$, where $k \approx \omega(gH)^{-1/2}$ is the characteristic tidal wavelength, and Δx is the numerical grid spacing. For our grid the relative error from numerical discretization of gradients is negligible in the deep ocean ($\approx 10^{-4}$) but significant on the continental shelf (where $k^{-1} \approx 200$ km and the relative error is ≈ 0.15). However, smaller spatial scales may be introduced into the tidal fields by roughness in $H(\mathbf{x})$ or coastal boundaries. Perhaps most seriously there are significant subgrid scale variations in topography (e.g., small islands and seamounts) which lead to scattering and dissipation unaccounted in the discretized LTEs.

To examine this possibility more closely, we used the 5′×5′ gridded values from the ETOPO5 database to compute H_{avg} and H_{min}, which are the average and minimum depths in each cell in the 512×256 grid, respectively. In our discrete numerical approximation of the LTEs we of course use H_{avg} to approximate $H(\mathbf{x})$. Over most of the ocean $H_{min} \approx H_{avg}$. However, there are numerous grid cells, particularly along volcanic arcs and hot spot tracks, where $H_{min} \ll H_{avg}$. Here there are clearly significant subgrid scale topographic features which are completely ignored in our numerical treatment. To approximately allow for these effects in our estimate of the dynamical error variance, we modified (44) slightly, replacing $H(\mathbf{x})$ by H_{min}. Note that we apply this modification only to our estimate of dynamical error variance and not to the definition of the dissipation parameter κ. With this approximation we effectively assume 100% errors in the dynamics, for all grid cells for which any part of the bathymetry (at the ≈10 km scale of ETOPO5) is shallow (i.e., 200 m or less).

The error in the momentum balance equations due to our approximate treatment of tidal loading and ocean self-attraction is

$$\delta f_\beta = gH\partial_x[G^*h - \beta h]. \qquad (45)$$

The magnitude of δf_β can be estimated using the tidal loading calculations reported by *Francis and Mazzega* [1990]. For the M_2 constituent of the the SCH80 model, they find that $G^*h - \beta h$ (with $\beta = 0.9$) is generally very small (a few millimeters at most), except near the coast where amplitudes can reach several centimeters [*Francis and Mazzega*, 1990; Figure 7a]. Gradients of $G^*h - \beta h$ are significant only in the direction normal to the coast. Typical values for these normal gradients near the coast are of the order of 1-2 cm per several hundred kilometers. If we use this estimate with $H \approx 1000$m in (45) we then have

$$\mathrm{Var}(\delta f_\beta) \approx 1\text{–}5\times10^{-7}\mathrm{m}^4\ \mathrm{s}^{-4} \ll |\omega u|^2 \approx 10^{-4}\text{–}10^{-6}\ \mathrm{m}^4\ \mathrm{s}^{-4}. \qquad (46)$$

Thus δf_β is negligible except perhaps for the normal component of momentum balance within a few hundred kilometers of the coast. Even here, δf_β will generally be much smaller than δf_κ.

Finally, errors in the bathymetry ΔH result in errors in the momentum equation of the form

$$\delta f_H = \beta g \Delta H \partial_x h. \qquad (47)$$

Using a prior estimate of h, $\mathrm{Var}(\delta f_H)$ can be estimated as a function of position if we have an estimate of typical scales of the bathymetry errors ΔH. Based on a comparison of ETOPO5 with more recent detailed compilations of data in the north Pacific [*Youtsey*, 1993; *Smith*, 1993], we estimate that the relative errors in the

ETOPO5 data base are of the order of 5% in the deep ocean, and approximately 10 m in water shallower than 200 m. Over much of the ocean δf_H is roughly an order of magnitude smaller than δf_κ. However, along the edges of the ocean basins where tidal gradients are large, amplitudes of these two sources of error can be comparable.

The considerations of the preceding paragraphs provide rough estimates of typical amplitudes of possible sources of dynamical errors. These estimates depend on $\mathbf{x}$ through H, u, v, $\partial_x h$ and $\partial_y h$, which can be estimated directly from a prior tidal model (e.g., $\mathbf{u}_0 = \mathbf{S}^{-1}\mathbf{f}_0$). In fact, these fields are comparatively rough, and exhibit many small-scale features which may be spurious. To avoid reinforcing these features (indirectly, through the covariance) in the inverse solution, we smooth the spatial dependence of the estimated error variances as we construct the dynamical error variances $\sigma_u(\mathbf{x})$ and $\sigma_v(\mathbf{x})$. After computing an inverse solution, variance estimates can be refined using improved estimates of u, v, $\partial_x h$, and $\partial_y h$. We applied this bootstrap approach (using results from the 80-site pelagic data inversion discussed below) to compute estimates of $\sigma_j(\mathbf{x})$, $j = 1, 2$ (see Figure 4). Combined with the spatial correlation of Figure 3, these variances define the dynamical error covariance $\mathbf{C}_t$ which we use for all further inverse calculations discussed in this paper.

Boundary Condition Covariances $\mathbf{C}_c$ and $\mathbf{C}_o$

The error $\delta f_{\hat{\mathbf{n}}\mathbf{u}}$ in the rigid boundary condition at the coast represents flow across the discrete numerical boundary, which does not generally coincide with the actual coast. Consider a $\Delta x \times \Delta y$ cell in the C grid adjacent to a coast running north-south. One side of the cell corresponds to the numerical boundary, where the volume flux is taken to be zero. The total flux across the opposite side is $u\Delta y$. Adopting a simple model, where the true coast (where the volume flux really is zero) is displaced a random distance between $-\Delta x/2$ and $\Delta x/2$ from the numerical coast, and assuming that u varies linearly across the grid cell, leads to the estimate

$$\mathrm{Var}(\delta f_{\Delta y \hat{\mathbf{n}}\cdot\mathbf{u}}) \approx 0.10 \times \Delta y^2 |\hat{\mathbf{n}}\cdot\mathbf{u}|^2 . \qquad (48)$$

Assuming further that adjacent to the coast the continuity equation is dominated by the across-shore divergence, we can express the normal velocity component in (48) in terms of h and the grid spacing Δx:

$$\hat{\mathbf{n}}\cdot\mathbf{u} \approx i\omega h \Delta x . \qquad (49)$$

Thus we have the estimate

$$\mathrm{Var}(\delta f_{\Delta y \hat{\mathbf{n}}\cdot\mathbf{u}}) \approx 0.10 \times \Delta y^2 \omega^2 \Delta x^2 h^2 . \qquad (50)$$

Again we can use a prior, or refined, estimate of tidal elevations to compute a spatially varying estimate of the coastal boundary condition error variance. Note that the variance of (50) depends on the grid spacing, consistent with our assumption that the primary source of error originates with the discrete numerical grid. Note also that this sort of discretization error will result in errors in the boundary conditions which are at least approximately uncorrelated between distinct nodes in the numerical boundary. We thus take the rigid boundary condition error to be a white noise process, with "variance" (or spectral density) $0.10 \times \Delta y\, \omega^2 \Delta x^2 h^2$, consistent with the estimate (for variance of volume flux across the numerical boundary segment) given in (48).

With the assumption of a non-zero covariance for errors in the no-flow boundary condition at the coast, we allow a periodic variation of the total mass inside the modeled domain. Thus our inverse solution will not generally conserve mass exactly. It is possible that for some applications in geophysics or oceanography (e.g., studies of crustal deformation by tidal loading, or tidally induced variations in Earth rotation) this may present

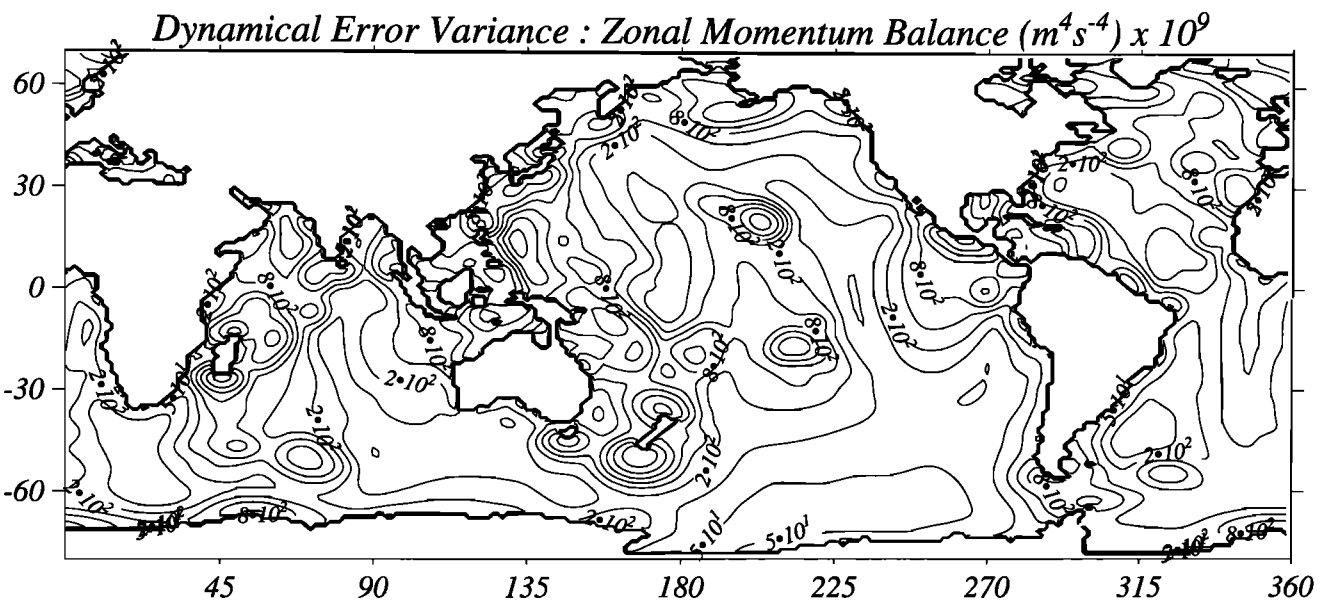

Figure 4. Spatial pattern of dynamical error variance for the zonal momentum balance. The estimates are computed by combining the estimates given in the text with prior estimates of tidal elevations and transports. Units are $m^4\ s^{-4}$. Note that the contours are evenly spaced on a logarithmic scale.

some difficulties. If exact mass conservation were deemed necessary, this could be guaranteed by simply setting the rigid boundary covariance $\mathbf{C}_c$ to zero, since we have already assumed that the continuity equation is a hard constraint. However, for many other purposes, including detiding of altimetry data, global conservation of mass is of no importance. Indeed, for these applications exactly enforcing the boundary conditions at the numerical boundary may seriously degrade the solution near the coast. Consistent with our general philosophy that the magnitude of the errors in the dynamics, as numerically modeled, should be used to determine optimal weights, we thus adopt (50) for the rigid boundary condition error covariance.

For the open boundary conditions we use the SCH80 models. Based on the comparison to ground truth data given by CR, we estimate that the error in the complex tidal constants for the M_2 constituent of this model has an amplitude of 6 cm, about 15% of the global RMS amplitude for this constituent (40 cm). We use this estimate of relative error (i.e., 15%), together with the RMS amplitude of the sea surface elevation in the SCH80 model along the open northern boundary, to estimate open boundary condition error variances for each constituent. Here we take the decorrelation length scale for the open boundary errors to be 5°, as for the dynamical errors.

An Example: Inversion of 80 Pelagic Tide Gauges

We illustrate the approach outlined above with a modest example: inversion of complex tidal constants for four constituents M_2, S_2, K_1, and O_1, from the 80 open-

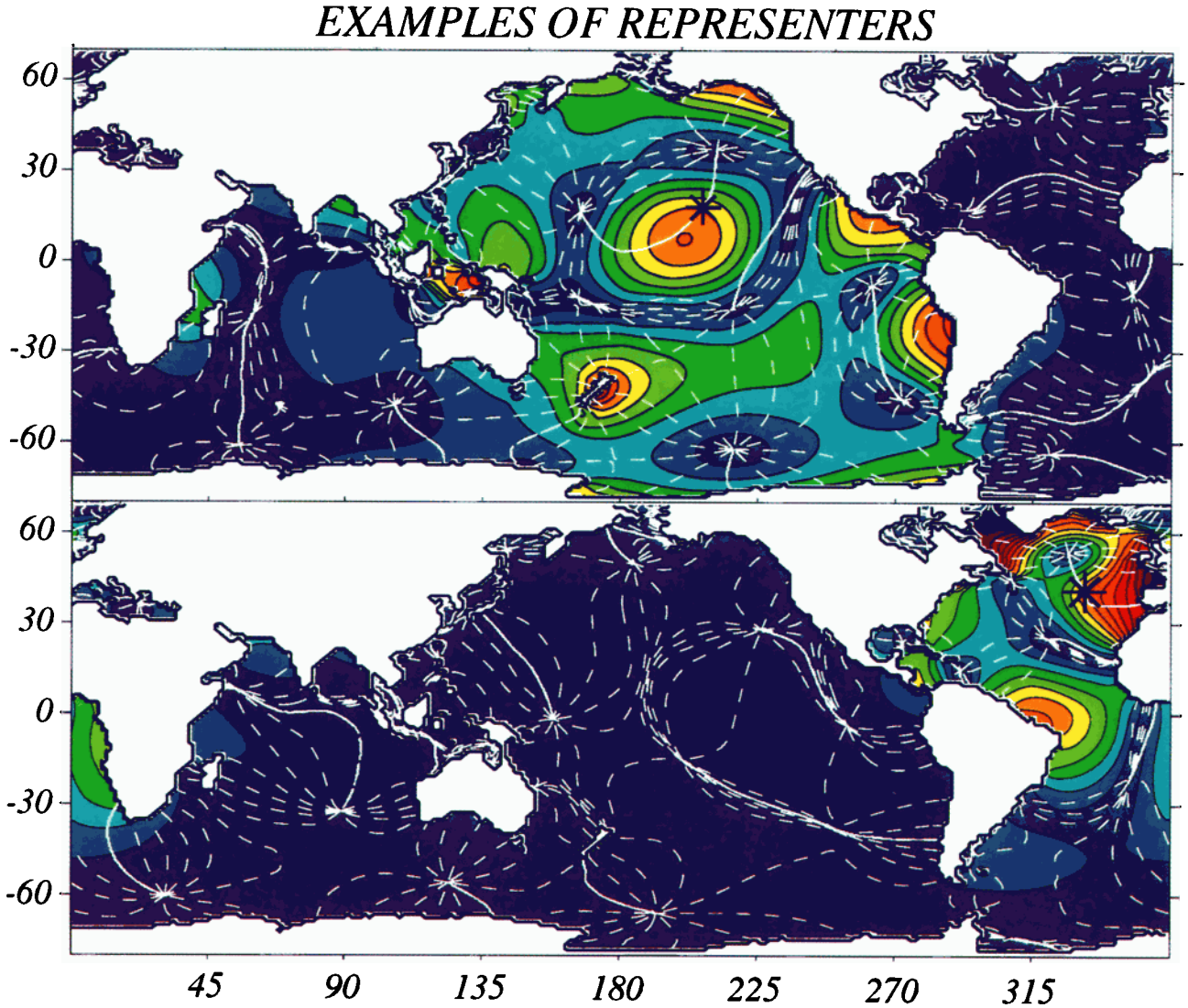

Plate 1. Examples of representers $\mathbf{r}_k$ for the M_2 constituent measured by tide gauges (a) in the central Pacific and (b) the North Atlantic. Amplitudes of the complex elevation component (h) are indicated by the color filled contours, and phase isolines are dashed, except for the line of zero phase. The locations of the measurement points $\mathbf{x}_k$ are marked by the black asterisks. Note that each $\mathbf{r}_k$ is a field of complex triples, with transport components in addition to the plotted sea level. The representers are broadly peaked in the vicinity of (but not exactly at) the observation points and have large amplitudes throughout the basin. Note also that $\mathbf{r}_k$ is real at $\mathbf{x}_k$, consistent with the Hermitian property of the representer matrix.

ocean locations of Figure 1 (either pelagic gauges or gauges on small islands [*Ray*, 1993]). This data set is a slightly expanded version of the ground truth data used for comparing tidal models in CR. For now we treat each constituent separately and solve four separate inverse problems. The 80 representers for each constituent were computed as outlined above on a 256×128 numerical grid. Examples of the h component of these representers are given in Plate 1 for tide gauges at two sites. The representers are broadly peaked in the vicinity of the tide gauge location, but vary smoothly and have large amplitudes over an entire ocean basin. Note that each representer is a complex vector field with transport components r_u and r_v in addition to the elevation component r_h plotted in Plate 1.

Singular Value Decomposition of the Representer Matrix

In the absence of measurement noise ($\mathbf{C}_\varepsilon = 0$), the global minimizer of the penalty functional is the linear combination of the prior model and the representers $\hat{\mathbf{u}}(\mathbf{x}) = \mathbf{u}_0 + \sum b_k \mathbf{r}_k$ for which the tide gauge data are exactly fit. In terms of the $K \times K$ representer matrix $\mathbf{R}$ and the representer coefficients $\mathbf{b}$, this condition is

$$\mathbf{Rb} = \mathbf{d} - L[\mathbf{u}_0] \,. \tag{51}$$

As the representers are dominated by smooth, basin-scale features, we expect the representers from nearby tide gauges to be grossly similar, so that this system of equations will be ill-conditioned. To explore this issue, and to fit the data in a stable fashion, we compute the singular value decomposition (SVD) of the Hermitian positive-definite representer matrix, $\mathbf{R} = \mathbf{V}\Lambda\mathbf{V}^*$. Here $\mathbf{V}$ is the unitary matrix of eigenvectors of $\mathbf{R}$, and Λ is the real diagonal matrix of eigenvalues ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$), plotted for the M_2 and K_1 constituents in Figure 5. The stability of the solution to the system (51) is determined by the spread of these eigenvalues. Even for the small-scale problem considered here, the inversion is moderately unstable. Exactly fitting the pelagic data (which are certainly contaminated by some noise) would thus be ill-advised.

The matrix $\mathbf{R}$ may be shown [*Bennett*, 1992] to be the covariance of $L[\mathbf{u}_{\text{True}} - \mathbf{u}_0]$ ($= \mathbf{d} - L[\mathbf{u}_0]$ in the absence of measurement errors). Thus $\mathbf{R}$ tells us how the dynamical errors affect deviations of observations from the predictions of the prior model. The SVD can be viewed as a rotation of the data vectors and the measurement functionals

$$d'_k = \sum_{k'} v_{k'k} d_{k'} \qquad L'_k = \sum_{k'} v_{k'k} L_{k'} \tag{52}$$

into a coordinate system for which $\mathbf{R}$ becomes a diagonal matrix (Λ). The eigenvalues λ_k give the variance of the rotated datum $d'_k = \sum_{k'} v_{k'k} d_{k'}$ (due to dynamical error). Allowing for measurement errors, with $\mathbf{C}_\varepsilon = \sigma^2\mathbf{I}$ for simplicity, the total variance (dynamical error plus measurement errors) of the rotated datum d'_k is

$$\text{Var}(d'_k) = \sigma^2 + \lambda_k \tag{53}$$

Obviously, if $\sigma^2 \gg \lambda_k$, d'_k is dominated by noise and offers little information about the tides. Using the SVD, a straightforward calculation [e.g., *Parker*, 1975; *Bennett*, 1992] shows that the inverse solution may be expressed as

$$\hat{\mathbf{u}}(\mathbf{x}) = \mathbf{u}_0(\mathbf{x}) + \sum_{k=1}^{K} \frac{d'_k}{\lambda_k + \sigma^2} \mathbf{r}'_k(\mathbf{x}) \tag{54}$$

where

$$\mathbf{r}'_k(\mathbf{x}) = \sum_{k'} v^*_{k'k} \mathbf{r}_{k'}(\mathbf{x}) \tag{55}$$

are the representers of the rotated data functionals L'_k, or "array modes" [*Bennett*, 1985].

The fitted data can also be expressed conveniently in this coordinate system as

$$\hat{d}_k = L'_k[\hat{\mathbf{u}}] = L'_k[\mathbf{u}_0] + \frac{\lambda_k}{\lambda_k + \sigma^2} d'_k \tag{56}$$

Thus if $\lambda_k \ll \sigma^2$ the rotated datum d'_k is essentially ignored. This means [e.g., *Parker*, 1975] that provided $\lambda_k \ll \sigma^2$ for $k > K'$, the summation of (54) can be truncated

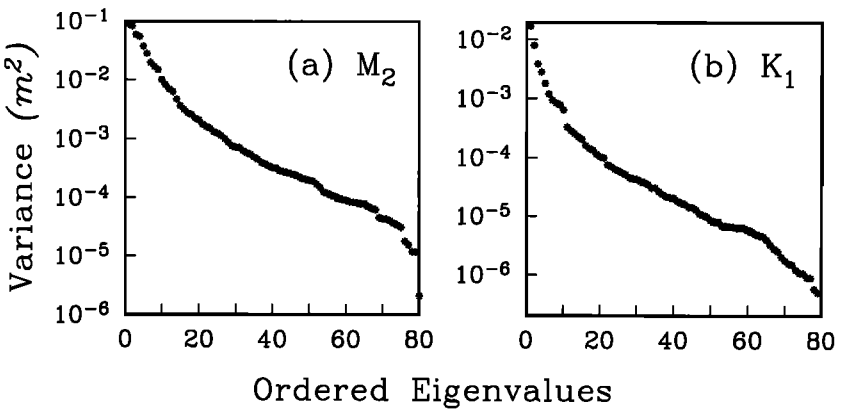

Figure 5. Ordered eigenvalues of representer matrices for the 80–site pelagic tide gauge inversion: (a) M_2; (b) K_1. Eigenvalues can be interpreted as predicted variances (due to dynamical errors) of linear combinations of the observable data. The eigenvalue spectra decay rapidly, indicating that most of the tidal signal can be explained by a small number of array modes.

Table 1. Model Misfits and Prediction Errors

	M_2	S_2	K_1	O_1
Misfit $\mathbf{u}_0$ (256×128 grid)	15.60	6.57	4.77	3.94
Misfit $\mathbf{u}_0$ (512×256 grid)	13.80	5.88	3.95	3.30
Misfit $\mathbf{u}$	1.97	1.07	0.65	0.72
RMSPE $\mathbf{u}$	4.45	2.09	1.72	1.71
σ	2.45	1.00	0.67	0.89

Summary of RMS misfits and prediction errors for 80–site pelagic tide gauge data set. RMS misfits (of real and imaginary parts separately) are given in centimeters for prior models $\mathbf{u}_0$ computed on two grids and for the inverse solution $\mathbf{u}$. The last two rows give, repsectively, the RMS prediction error estimated from the cross validation analysis and the inferred measurement error standard deviation (i.e., the value of σ for which RMSPE is minimized).

at $K'<K$ without significantly affecting the fit to the data. More precisely, the fraction of the (expected) tidal signal variance in the data which may be fit using only the first K' of the array modes r'_k is

$$\sum_{k=1}^{K'}\lambda_k / \sum_{k=1}^{K}\lambda_k \tag{57}$$

The eigenvalue spectra of Figure 5 fall off rapidly, suggesting that most of the tidal signal in $\mathbf{d} - L[\mathbf{u}_0]$ can be fit with a small number (≈15–20) of the array mode basis functions. The representer formulation allows us to restrict our search for the inverse solution to a space of finite dimension K, by discarding the infinite dimensional null space of the data functionals [e.g., *Bennett*, 1992]. The SVD effectively allows us to reduce further the dimension of the space of resolvable parameters. For the pelagic tide gauge inversion, where K=80, this reduction in state space dimension is not particularly important. However, for the inversion of crossover data considered in the next section this sort of reduction will be of critical importance.

So far our analysis of the tidal inverse problem has not used, and indeed does not depend upon, the actual values of the observations. The representer matrix depends on the dynamics, on the assumed properties of the dynamical errors, and the locations of the tide gauges. We now consider fitting the actual data. For our prior solution $\mathbf{u}_0$ we use the solution of the system (9)-(12) computed numerically on a 512×256 grid. The RMS misfits of the 80 pelagic constants to $\mathbf{u}_0$ are summarized in Table 1. In Figure 6 we plot the RMS error for the M_2 and K_1 constituents as a function of the number K' of array modes $\mathbf{r}'_k$ fit to the data. By fitting the leading 15 array modes, the RMS misfits for M_2 are reduced from 13.8 cm to ≈4 cm, comparable to RMS misfits to the same data set achieved by the SCH80 and CR91 models [*Ray*, 1993]. The variance reduction achieved with these 15 modes is 92%, roughly consistent with (57) and the eigenvalue spectra of Figure 5. Results for the principal diurnal constituent K_1 are qualitatively similar, with magnitudes of all quantities smaller by a factor of roughly 2.

Cross Validation

The trade-off between fitting the data and fitting the dynamics is controlled by the dynamical and measurement error covariances $\mathbf{C}_f$ and $\mathbf{C}_e$. In fact, our proposed model for $\mathbf{C}_f$ involves many uncertainties and should be viewed as a hypothesis, which may be challenged by the data. Measurement errors in the pelagic tidal constants are also somewhat difficult to assess. It is thus difficult to prescribe a priori a consistent level of fit to the data. To explore this issue further we assume $\mathbf{C}_e = \sigma^2\mathbf{I}$ and consider inverse solutions for a range of σ^2. As σ^2 varies, we trade off between fitting the data and satisfying the dynamics. For $\sigma^2 = 0$, the data are fit exactly, while for $\sigma^2 = \infty$, the data are ignored and the dynamics are exactly satisfied (so $\hat{\mathbf{u}} = \mathbf{u}_0$). To choose a reasonable value of σ^2 we use cross validation (CV [e.g., *Wahba*, 1990]) to assess the ability of the inverse solution to smooth between the available tide gauges (as a function of the assumed σ^2). In its conceptually simplest form this is accomplished by estimating the RMS prediction error (RMSPE, the RMS error in predicting data for one additional tide gauge) by repeating the inverse calcula-

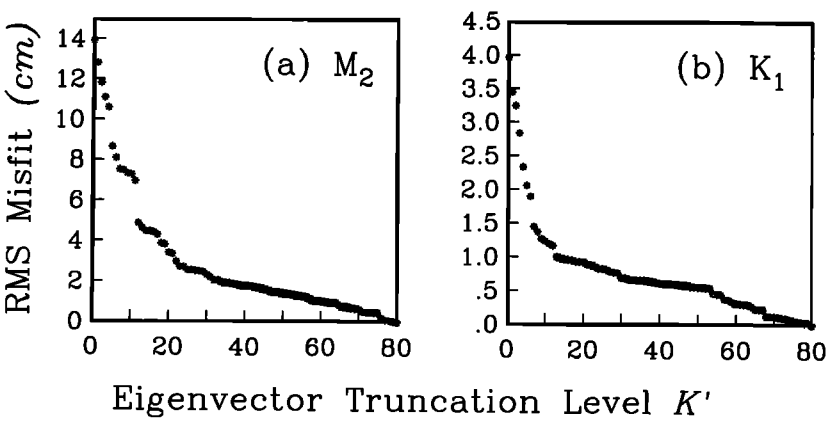

Figure 6. Residual RMS error for (a) M_2, and (b) K_1 constituents as a function of the number K' of array modes fit to the data. For M_2, fitting 15 array modes reduces the RMS misfit to ≈4 cm; this corresponds to a 92% reduction in variance. Similar reductions of misfit (from 4 cm to 1 cm, a 95% variance reduction) are achieved by fitting 13 array modes for K_1.

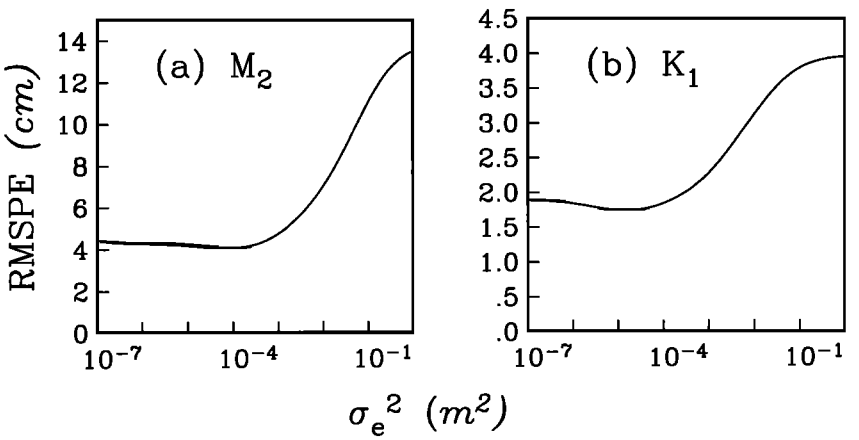

Figure 7. Estimated RMS prediction error computed by cross validation for (a) M_2 and (b) K_1 constituents as a function of the assumed measurement error variance σ^2. The value of σ^2 for computation of the inverse solution is chosen to be the largest value of σ^2 for which the broad minimum RMSPE is achieved (6×10^{-4} m^2, 5×10^{-5} m^2 for M_2 and K_1, respectively). Note that the minimum RMSPE for the two constituents (4.45 and 1.72 cm) and the inferred measurement error standard deviations (2.5 and 0.7 cm) do not agree exactly due to inadequacies in our specification of the dynamical error covariances. These results suggest that the dynamical error variances are somewhat larger than we have assumed.

tion K times, omitting each data point in turn and computing

$$\mathrm{RMSPE}(\sigma^2) = [K^{-1}\sum_{k=1}^{K} | d_k - L_k[\hat{\mathbf{u}}_{(k)}(\sigma^2)] |^2]^{1/2} . \quad (58)$$

Here $\hat{\mathbf{u}}_{(k)}(\sigma^2)$ is the inverse solution of (25)-(26) calculated (under the assumption $\mathbf{C}_\varepsilon = \sigma^2\mathbf{I}$) with the kth datum omitted. For M_2, as σ^2 is reduced from 1 m^2 to $\approx10^{-3}m^2$, RMSPE is reduced from a maximum of 13.8 cm (the prior RMS misfit) to a minimum of $\approx$4.0 cm (Figure 7). Further decreases of σ^2 result in small increases in RMSPE. There are several points worth noting here. First, the CV analysis implies that models based on interpolation of these 80 tide gauges will be at best accurate to $\approx$4 cm (for M_2) in the open ocean, where almost all of the 80 gauges are located; the reliability of the model is probably much poorer on the shelves. This does not necessarily mean that the errors in the tidal constants are actually this large - only that the predictability of the tide from nearby gauges and the simplified hydrodynamics assumed here is limited to 4 cm. With a denser set of tide gauges a smaller minimum RMSPE might be achievable.

To optimize the predictive power of the inverse solution, we should choose σ^2 as the argument which minimizes the prediction error estimates plotted in Figure 7. To compute the final inverse solution for each of the four constituents, we thus use the largest value of σ^2 for which the broad minimum in RMSPE is achieved. For all constituents the estimates of measurement error variances obtained by this cross validation approach (see Table 1) are smaller than the estimated prediction errors. For instance, for M_2 the optimal value of σ^2 (6×10^{-4}) corresponds to a measurement error standard deviation of $\sigma = 2.4$ cm, while the RMSPE is $\approx$4 cm. The cross-validation analysis is only sensitive to the relative amplitudes of dynamical and measurement errors: doubling the dynamical error variances would lead to a doubling of the optimal measurement error variance. To make σ agree with the estimated RMSPE, we would have to increase the magnitude of the dynamical error variances. The inconsistency between RMSPE and the inferred measurement error variances thus suggests that the dynamical error variances may be somewhat larger than our analysis of the dynamical errors suggests.

Calculation of the Inverse Solution on a Refined Grid

Once the representer coefficients have been determined by solving (25), the inverse solution can be constructed by directly computing the sum of representer fields (24) at each point on the numerical grid. With this approach the resolution of our final model would be limited by the grid used for the representer calculation (256×128 for the example discussed here). This limitation can be overcome with a slightly different approach.

The inverse solution has the form $\hat{\mathbf{u}} = \mathbf{u}_0 + \delta\mathbf{u} = \mathbf{u}_0 + \sum b_k \mathbf{r}_k$, where each $\mathbf{r}_k$ satisfies

$$\mathbf{S}\mathbf{r}_k = \mathbf{C}_f \boldsymbol{\eta}_k \qquad \mathbf{S}^\dagger\boldsymbol{\eta}_k = \boldsymbol{\Delta}_k \quad (59)$$

and $\boldsymbol{\Delta}_k = (0, 0, \delta(\mathbf{x}-\mathbf{x}_k))^T$ represents a unit magnitude impulse forcing of the elevation component at $\mathbf{x}_k$. But then, by linearity, $\delta\mathbf{u}$ must satisfy

$$\mathbf{S}\delta\mathbf{u} = \mathbf{C}_f\boldsymbol{\eta} \qquad \mathbf{S}^\dagger\boldsymbol{\eta} = \sum b_k\boldsymbol{\Delta}_k . \quad (60)$$

Thus $\delta\mathbf{u}$ may be computed by repeating the representer calculation once more with the backward system forced by the scaled comb of delta functions $\sum b_k\boldsymbol{\Delta}_k$.

There is nothing to stop us from doing this final step on a grid finer than that used for the representer calculation (e.g., on the 512×256 grid used for the prior model calculation). This would yield the fine-grid representers $\mathbf{r}_k$ with "coarse-grid" coefficients b_k. However there is no guarantee that the result will really minimize the fine-grid discrete version of the penalty functional. Indeed, an inverse solution computed in this way may not even fit the data adequately. However, as we now show, it is possible to refine iteratively the fine-grid

inverse solution until the true (fine grid) minimizer of (21) is found.

Assume that there are two numerical grids, a coarse one on which the representer matrix $\mathbf{R}$ has been computed, and a finer grid on which the prior model and final inverse solution are to be calculated. Let $\mathbf{R}_F$ be the representer matrix which would be calculated on the fine grid. We would like to find the solution to

$$(\mathbf{R}_F + \mathbf{C}_\varepsilon)\mathbf{b} = \mathbf{d} - L[\mathbf{u}_0] . \quad (61)$$

Notice that while we do not compute the full matrix $\mathbf{R}_F$ we can evaluate $\mathbf{R}_F\mathbf{a}$ for any complex vector $\mathbf{a}$. To do this, we solve the coupled system (60) on the fine grid, with the backward system forced by $\sum a_k \Delta_k$. Let $\delta h(\mathbf{x})$ be the elevation component of the resulting solution $\delta\mathbf{u}$. Then it is readily verified that

$$\delta h(\mathbf{x}_k) = L_k[\delta\mathbf{u}] = (\mathbf{R}_F\mathbf{a})_k . \quad (62)$$

Thus we can compute $(\mathbf{R}_F + \mathbf{C}_\varepsilon)\mathbf{a}$ by doing the equivalent of one representer calculation on the fine grid, evaluating the solution at the measurement points $\mathbf{x}_k$, and adding $\mathbf{C}_\varepsilon\mathbf{a}$ to the result. This calculation is exactly what is needed to solve the Hermitian system (61) iteratively using a conjugate gradient approach [e.g., *Press et al.*, 1986], without actually calculating all of $\mathbf{R}_F$. The convergence of such a scheme, while guaranteed, will generally be very slow if $\mathbf{R}_F + \mathbf{C}_\varepsilon$ is poorly conditioned. However, we can use the coarse-grid representer matrix to calculate a very good initial estimate of the fine grid representer coefficients $\mathbf{b}$, and to improve the conditioning of the linear system. Specifically, letting $\mathbf{R} + \mathbf{C}_\varepsilon = \mathbf{U}\Lambda\mathbf{U}^*$. We replace (61) by

$$\Lambda^{-1/2}\mathbf{U}^*(\mathbf{R}_F + \mathbf{C}_\varepsilon)\mathbf{U}\Lambda^{-1/2}\mathbf{b}' = \Lambda^{-1/2}\mathbf{U}^*(\mathbf{d} - L[\mathbf{u}_0])$$
$$\mathbf{b}' = \Lambda^{1/2}\mathbf{U}^*\mathbf{b} . \quad (63)$$

If $\mathbf{R}_F \approx \mathbf{R}$, which should hold as long as the initial grid is not too coarse, the matrix on the left-hand side of (63) will be very nearly the identity, and convergence of the conjugate gradient algorithm will be rapid. For our 80 tide gauge inversion example, the conjugate gradient scheme converged, with a relative error

$$\|(\mathbf{R}_F + \mathbf{C}_\varepsilon)\mathbf{b} - (\mathbf{d} - L[\mathbf{u}_0])\| / \|\mathbf{d} - L[\mathbf{u}_0]\| < 0.001 ,$$

in only six steps.

The final result of this iterative scheme is an inverse solution, on a 512×256 grid, for each of the four constituents (M_2, S_2, K_1, O_1). Misfit statistics for these models are summarized in Table 1. Amplitude and phase of the sea surface elevation for the inverse solution $\hat{\mathbf{u}}$ are given in Plate 2 for the K_1 constituent, with the prior solution $h_0(\mathbf{x})$, and SCH80 presented for comparison. While qualitatively similar to SCH80, the inverse solution is noticeably smoother. For the other three constituents the inverse solutions are very similar to, but smoother than, the corresponding SCH80 tides.

4. Direct Inversion of Altimetry Crossover Differences

We are now ready to consider direct inversion of the T/P crossover difference data. To do this we must work with all L constituents simultaneously. We thus return to the notation used in our initial definition of the state space (equation (1)), which combines the complex three-dimensional vector fields for single constituents (e.g., tidal elevation and transport fields, representers, dynamical errors) into $3L$-dimensional complex vector fields. At the end of this section we apply the methods developed in this paper to crossover differences from M=38 orbit cycles of TOPEX data.

Representers for Multiple Constituents

In the previous section we calculated the representers for evaluation of a single constituent l at $\mathbf{x}_i$. The assumption that dynamical errors were not correlated between constituents was implicit in this calculation. In fact, tidal elevations and transports vary smoothly with forcing frequency, and it seems more reasonable to assume that forcing errors will also, since they represent inadequacies in parameterizations of dissipation and tidal loading, and in numerical approximations of the equations. This expectation is confirmed by results from our inversion of the 80 pelagic sites. We estimated the correlation between dynamical errors for pairs of constituents, by averaging cross products of the estimated errors (i.e., $\mathbf{C}_f\boldsymbol{\eta}$ from (60)) over the ocean (Table 2). These globally averaged sample correlations are high for constituents all of the same species (i.e., both diurnal or both semidiurnal), and close to zero for constituents of different species. To capture this interconstituent correlation in a simple manner we extend the single-constituent dynamical error covariance model of (42) to

$$\mathrm{Cov}[f_j^l(\mathbf{x}), f_{j'}^{l'}(\mathbf{x}')] = \delta_{jj'}\sigma_j^l(\mathbf{x})\sigma_{j'}^{l'}(\mathbf{x}')\psi(\lambda)\rho_{ll'} . \quad (64)$$

Here $\rho_{ll'}$ is the (globally averaged) correlation between forcing errors for constituents l and l', which is assumed to be the same for all components j (i.e., u, v, h). For our initial application of the four-constituent model to the T/P crossover data we use the simplified correlation model given in parentheses in Table 2.

This more complex dynamical error covariance forces us to treat all constituents simultaneously. The representer calculation is now slightly more complicated even for harmonically analyzed data. We partition the full covariance operator, and the forcing error vectors for the L constituents, respectively, as

$$\mathbf{C}_f = \begin{bmatrix} \mathbf{C}_f^{11} & \cdots & \mathbf{C}_f^{1L} \\ \vdots & & \vdots \\ \mathbf{C}_f^{L1} & \cdots & \mathbf{C}_f^{LL} \end{bmatrix} \qquad \delta\mathbf{f} = \begin{bmatrix} \delta\mathbf{f}^1 \\ \vdots \\ \delta\mathbf{f}^L \end{bmatrix} , \quad (65)$$

so that the covariance between different constituent error vectors is

$$\mathrm{Cov}[\delta\mathbf{f}^l, \delta\mathbf{f}^{l'}] = \mathbf{C}_f^{ll'} . \quad (66)$$

Note that covariances for boundary conditions, for which we assume the same interconstituent correlation, are implicitly included in the appropriate blocks of the partitioned covariance operator of (65).

Next, we formally define for $1 \le l, l' \le L$, $1 \le i \le I$

$$\mathbf{p}_{il'}^l = \mathbf{S}_l^{-1}\mathbf{C}_f^{ll'}\mathbf{S}_{l'}^{-\dagger}\Delta_i , \quad (67)$$

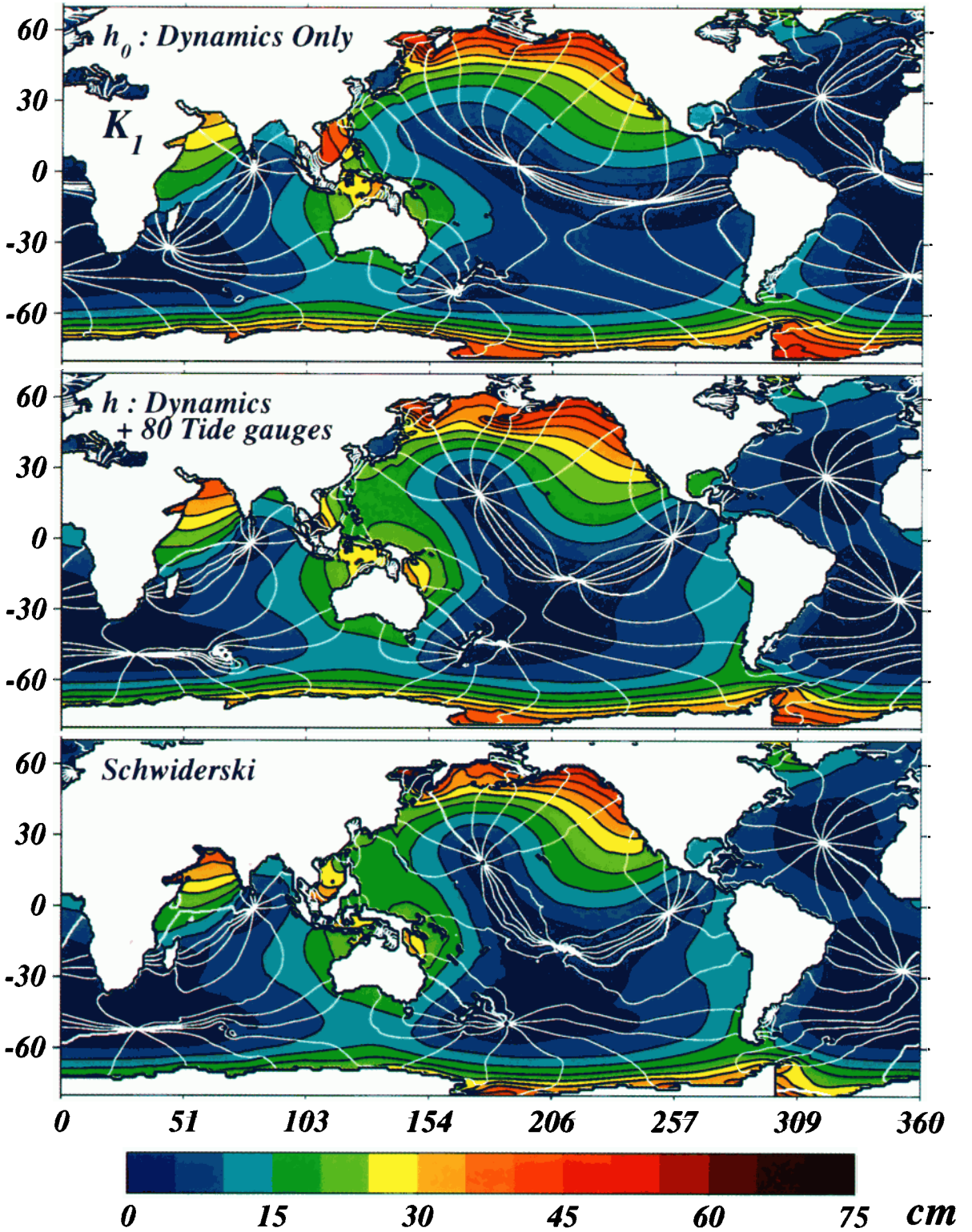

Plate 2. Global models for sea surface elevation $h(\mathbf{x})$ for the K_1 tidal constituent. (a) prior model $\mathbf{u}_0$; (b) inverse solution $\mathbf{u}$ for the 80 pelagic tide gauges of Figure 1, and (c) SCH80. Phase isolines are plotted in white over color filled contours of amplitude. Contour interval is 5 cm for amplitude, 30° for phase. Adding data from the 80 tide gauges makes the prior model much more like the SCH80 solution. Note that while grossly similar to SCH80 the inversion result is noticeably smoother.

Table 2. Interconstituent Dynamical Error Correlation

	M_2	S_2	K_1	O_1
M_2	1.00+0.00i	(0.71+0.37i)	(0.00+0.00i)	(0.00+0.00i)
S_2	0.71−0.37i	1.00+0.00i	(0.00+0.00i)	(0.00+0.00i)
K_1	0.10+0.14i	0.04+0.09i	1.00+0.00i	(0.56−0.36i)
O_1	0.09+0.14i	0.07+0.17i	0.56+0.36i	1.00+0.00i

Interconstituent correlation for dynamical errors estimated from 80–site pelagic tide gauge data set. Actual sample correlations, computed by averaging interconstituent cross products of the estimated dynamical errors over the ocean, are given below the diagonal. The complex correlation matrix is Hermitian. The entries above the diagonal (in parentheses) give the simplified correlation structure used for the full multiple constituent dyanamical error covariance.

where Δ_i is an impulsive forcing at $\mathbf{x}_i$. Recall that Δ_i is a field of complex triples, while $\mathbf{S}_l$, $\mathbf{C}_f^{ll'}$ and $\mathbf{S}_{l'}^{\dagger}$ are complex 3×3 matrix operators. The actual calculation of $\mathbf{p}_{il'}^{l}$ proceeds almost exactly as described in the previous section for the single-constituent representers. Indeed, the single-constituent representers are just a special case of the fields defined in (67) with $l = l'$. For the more general case with $l \neq l'$, different frequencies are used for time-stepping the forward (ω_l) and backward ($\omega_{l'}$) equations, and the off-diagonal covariance $\mathbf{C}_f^{ll'}$ is used for smoothing the backward solution $\boldsymbol{\eta}$ (see Appendix C). For the general case, where we allow for correlated dynamical errors, the representer field (with respect to the state space τ of $3L$-dimensional complex fields) for the evaluation functional for constituent l at $\mathbf{x}_i$ can now be given as

$$\mathbf{p}_{il} = \begin{bmatrix} \mathbf{p}_{il}^{1} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{p}_{il}^{L} \end{bmatrix} . \tag{68}$$

The full complex, $IL \times IL$ representer matrix for L constituents of harmonically analyzed data at the I crossover points is then

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \cdots & \mathbf{P}_{1I} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \mathbf{P}_{I1} & \cdots & \mathbf{P}_{II} \end{bmatrix} . \tag{69}$$

where $P_{ii'}^{ll'}$ is just the elevation component of $\mathbf{p}_{i'l'}^{l}$ evaluated at $\mathbf{x}_i$. Recall that the elevation impulse is at frequency $\omega_{l'}$ at point $\mathbf{x}_{i'}$, while the measurement is taken at frequency ω_l at point $\mathbf{x}_i$. The generalized L constituent representers of (68) are the building blocks from which we can construct representer fields for time domain data, such as the T/P crossover differences.

Representers for Time Domain Data

Our goal is to invert altimetry differences from I crossover points, each sampled during some or all of M orbit cycles (see Figure 2). As above, the total number of data points will be denoted by $K \leq IM$ (in general, some data will be discarded or missing). However, for most purposes it will be more convenient to use the double subscript im to refer to individual crossover differences. Thus crossover differences, which are related to the unknown complex tidal amplitudes for L constituents through (6), are denoted explicitly as d_{im}.

The functional L_{im} (defined implicitly in (6)) which relates the time domain datum d_{im} and the complex tidal state $\mathbf{u} \in \tau$ is not strictly linear. Consider the real-valued functional L_{im}: $L_{im}[c\,\mathbf{u}] \neq cL_{im}[\mathbf{u}]$ for all complex constants c. Of course, (6) is linear in the real and imaginary parts of $\mathbf{u}$. To make the time domain data functionals linear, it is necessary to treat the tidal state (and representers, dynamical errors, etc.) as real $6L$-dimensional vector fields in the real Hilbert space τ_R. This presents us with only minor technical complications, which we do not wish to dwell on. However, there is a close connection between the complex representers for evaluation of a single constituent, and the real representers for time domain data. To allow us to give these relationships correctly and succinctly in the following discussion, some additional notation is necessary. For any complex n-dimensional vector field $\mathbf{v}$, we denote by $\bar{\mathbf{v}}$ the real $2n$-dimensional vector constructed by interlacing the real and imaginary parts of the components of $\mathbf{v}$. Similarly, given an $N \times M$ complex matrix $\mathbf{X}$ we can construct a $2N \times 2M$ real matrix $\bar{\bar{\mathbf{X}}}$, by interlacing the real and imaginary parts in both the rows and the columns, as discussed in Appendix D. Representers and matrices needed for inversion of time domain data are essentially linear combinations of the the corresponding frequency domain quantities given above. The basic idea behind these calculations can be readily understood without reference to the exact nature of the correspondence between real and complex vectors and matrices. The reader who is interested in a more precise treatment of these technical details is referred to Appendix D.

Using this notation, a simple calculation (see Appendix D) shows that the representer for the crossover difference linear function L_{im} is the real field $\bar{\mathbf{r}}_{im}(\mathbf{x})$ where

$$\mathbf{r}_{im} = \begin{bmatrix} \sum_{l'=1}^{L} a_{iml'}^{*}\, \mathbf{p}_{il'}^{1} \\ \cdot \\ \cdot \\ \cdot \\ \sum_{l'=1}^{L} a_{iml'}^{*}\, \mathbf{p}_{il'}^{L} \end{bmatrix} . \tag{70}$$

The representer matrix for the crossover differences can be expressed in terms of the representer matrix $\mathbf{P}$ of (70). The elements of $\mathbf{R}$ are given by

$$R_{(im)(i'm')} = \sum_{l,l'=1}^{L} \mathrm{Re}[a_{iml}\, a_{i'm'l'}^{*}\, P_{ii'}^{ll'}] . \tag{71}$$

The relationship between $\mathbf{R}$ and $\mathbf{P}$ can also be given in matrix notation. Let $\mathbf{A}$ be the $K \times 2IL$ real matrix with columns indexed by the pair (im)=(crossover point, orbit cycle), and rows by the triple $(ji'l)$= (real or imaginary part, crossover point, constituent)

$$A_{(im)(jl')} = \delta_{ll'} \text{ Re } a_{iml} \qquad j = 1$$
$$A_{(im)(jl')} = -\delta_{ll'} \text{ Im } a_{iml} \qquad j = 2 \qquad (72)$$

Then, (71) can be written

$$\mathbf{R} = \mathbf{A}\bar{\bar{\mathbf{P}}}\mathbf{A}^T . \qquad (73)$$

In a sense, (72) provides a neat division between the spatial and temporal aspects of the crossover data. Spatial aspects (locations of crossover points, spatial structure implied by the dynamics) are encoded in **P**, temporal aspects (times of the orbit crossings) are in **A**. **P** can be calculated once and for all, independent of the ultimate number of orbit cycles. **A**, on the other hand, grows in size as the number of orbit cycles (M) grows.

Since the crossover representers are all linear combinations of the IL complex fields $\mathbf{p}_{il}(\mathbf{x})$, the crossover difference inverse solution (which will be a linear combination of the crossover representers and the prior solution) must be of the form

$$\hat{\mathbf{u}} = \mathbf{u}_0 + \sum_{i=1}^{I}\sum_{l=1}^{L} b_{il}\mathbf{r}_{il} . \qquad (74)$$

The inverse solution is thus completely characterized by the $K' = 2IL$ real numbers in the complex coefficient vector $\mathbf{b}$ ($K' \approx 8\times10^4$ for a model with eight constituents). The total number of crossover differences from all orbit cycles expected during the T/P mission ($M \approx 100$ or more) will be considerably larger ($K \approx IM = \approx 5\times10^5$). Thus (as can already be seen from (73)) the real $K\times K$ crossover difference data representer matrix **R** defined in (71) cannot be of full rank. Some minor modifications to the general inversion theory outlined in section 2 are required to treat this case.

The situation is simplest if the measurement error covariance is "white" and isotropic ($\mathbf{C}_\varepsilon = \sigma^2\mathbf{I}$). In theory we may always transform the data vector $\mathbf{d} \rightarrow \mathbf{C}_\varepsilon^{-1/2}\mathbf{d}$ and the matrix $\mathbf{A} \rightarrow \mathbf{C}_\varepsilon^{-1/2}\mathbf{A}$ so that this is so [e.g., *Parker*, 1975]. However, for the crossover inversion, K will be too large for this to be feasible unless the form of $\mathbf{C}_\varepsilon$ is restricted to some degree. Fortunately, crossover differences are probably at least approximately uncorrelated between distinct orbit cycles. Any long-period signals or noise in the altimetric data will be strongly attenuated in the differences (typically separated in time by ≈3–7 days) of ascending and descending orbits. With this assumption, $\mathbf{C}_\varepsilon$ is block diagonal, with blocks no larger than the number of crossover data in a single orbit cycle (≈5000). For the initial crossover inversion results presented here, we will assume a very simple diagonal form for $\mathbf{C}_\varepsilon$, with constant measurement error variance. In future we intend to refine our treatment of the measurement error covariance, allowing for spatial variability of oceanographic signal, and for correlation due to orbit errors, but retaining the assumption of independence between orbit cycles. For now, assume we have reduced to the simple case $\mathbf{C}_\varepsilon = \sigma^2\mathbf{I}$.

We can express the penalty functional directly in terms of the real coefficient vector $\bar{\mathbf{b}}$ as

$$J[\bar{\mathbf{b}},\mathbf{d}] = \sigma^{-2}(\mathbf{d}-[L[\bar{\mathbf{u}}_0] + \mathbf{A}\bar{\bar{\mathbf{P}}}\bar{\mathbf{b}}])^T(\mathbf{d}-[L[\bar{\mathbf{u}}_0] + \mathbf{A}\bar{\bar{\mathbf{P}}}\bar{\mathbf{b}}]) + \bar{\mathbf{b}}^T\bar{\bar{\mathbf{P}}}\bar{\mathbf{b}} . \qquad (75)$$

Let $\mathbf{A} = \mathbf{QU}$ be the Q-R decomposition of **A** into the product of a $K\times K'$ orthonormal matrix **Q**, and a $K'\times K'$ upper triangular matrix **U**. Note that $\mathbf{QQ}^T$ is a matrix which projects orthonormally onto the span of the columns of **A**, that is, onto the K' dimensional subspace of the data space which can be fit by any L constituent tidal state. The part of the crossover data in the orthogonal complement of this space, namely $(\mathbf{d} - \mathbf{QQ}^T\mathbf{d})$, is unaffected by any possible changes in tidal state. Note that **A** depends only on the crossover times. As a consequence $\mathbf{QQ}^T\mathbf{d}$ (which is closely related to a harmonic analysis of the differences at each crossover point into L constituents) is completely independent of our hydrodynamic assumptions. Thus $M^2_{\min} = \|\mathbf{d}-\mathbf{QQ}^T\mathbf{d}\|^2$ is the minimum achievable total squared data misfit for any tidal state. Let

$$\mathbf{d}' = \mathbf{Q}^T(\mathbf{d}-L[\bar{\mathbf{u}}_0]) \qquad \bar{\mathbf{b}}' = \mathbf{U}^{-T}\bar{\mathbf{b}}$$
$$\mathbf{R}' = \mathbf{U}\bar{\bar{\mathbf{P}}}\mathbf{U}^T . \qquad (76)$$

Substituting $\mathbf{QU} = \mathbf{A}$ into (75) and using (76) to simplify, we find

$$J[\bar{\mathbf{b}},\mathbf{d}] = \qquad (77)$$
$$\sigma^{-2}[M^2_{\min} + (\mathbf{d}'-\mathbf{R}'\bar{\mathbf{b}}')^T(\mathbf{d}'-\mathbf{R}'\bar{\mathbf{b}}')] + \bar{\mathbf{b}}'^T\mathbf{R}'\bar{\mathbf{b}}' ,$$

which is minimized for the coefficient vector

$$\bar{\mathbf{b}} = \mathbf{U}^T(\mathbf{R}' + \sigma^2\mathbf{I})^{-1}\mathbf{d}' . \qquad (78)$$

By taking linear combinations of the crossover differences, we have reduced the inverse problem to the standard form treated in previous sections. In the process the vector of all crossover differences **d** of dimension K has been reduced to a data vector $\mathbf{d}'$ of significantly lower dimension $K' = 2IL$, for which the corresponding data functionals and representers are linearly independent. The matrix $\mathbf{R}'$ defined in (76) is just the representer matrix for this reduced data set. All of the analysis and formulae discussed in previous sections (cross validation, iterative refinement of the inverse solution on a finer grid) are thus directly applicable.

Practical Strategies for Large Data Sets

In theory, (76) and (77) provide a complete solution to the inverse problem for crossover differences. Unfortunately, $\mathbf{R}'$ is still too large for a direct application of these equations to be practical. Simply storing $\mathbf{R}'$ for the T/P crossover differences would require of the order of 2.5×10^{10} bytes (25 Gb) of computer memory. More seriously, calculation of all elements of the matrix **P** requires solving the full (multiple-constituent) forward and backward systems IL times each. Even with a supercomputer, such as the CM-200 used for this project, computing representers at all crossover points is not really practical.

In fact, it seems a priori unlikely that the calculation of representers at each and every crossover point is truly necessary. As discussed above, and illustrated in Plate 1, the representers are dominated by basin-scale features which are generally very similar for nearby measurement functionals. The redundancy of representers for the tidal inverse problem is clearly illustrated by the SVD analysis of the pelagic tide gauge representer matrix.

The globally distributed set of 80 harmonic constants could be fit quite well (e.g., 92% of the variance explained for M_2) using only the dominant 15 array modes. In a very real sense, the full set of representers is not necessary for a reasonable solution to the inverse problem.

Parker and Shure [1982] suggested a general approach to reducing the size of an ill-conditioned inverse problem such as this. The idea is to solve a slightly modified problem, in which the minimizer of the penalty functional is sought in a restricted subspace of the full representer span. For the generic inverse problem, we seek $\hat{\mathbf{u}}$ which minimizes J over

$$\left\{ \mathbf{u}(\mathbf{x}) = \sum_{n=1}^{N} b_n \mathbf{r}'_n(\mathbf{x}) \right\} \tag{79}$$

where the functions $\mathbf{r}'_n$, $1 \le n \le N$ are linear combinations of the full set of representers $\mathbf{r}_k$, $1 \le k \le K$

$$\mathbf{r}'_n = \sum_{k=1}^{K} v_{nk} \mathbf{r}_k \ , \tag{80}$$

and $N \ll K$. If the subspace defined by the linear combinations of (80) is carefully chosen, the result should be a good approximation to the full inverse solution. For instance, we could choose the v_{nk} as in (55) (i.e., from the elements of the dominant eigenvectors of $\mathbf{R}$), so that the basis functions $\mathbf{r}'_k$ are just the array modes. As discussed above, the expansion of the inverse solution in array modes may be truncated without significantly affecting the final result (see (54)-(57)). A key idea of *Parker and Shure* [1982] is that there are often simpler ways to achieve this reduction in the effective dimension of the state space.

To discuss specifics for the time domain crossover difference inversion it is most useful to describe the state space, and possible subspace, in terms of the representers for harmonically analyzed data (i.e., the complex vector fields $\mathbf{p}_{il}$). Compared to the dense array of crossover points (approximately 3-4 degree spacing at middle latitudes; see Figures 1 and 8), the 80 pelagic tide gauge sites are widely spaced. We thus expect a much greater degree of redundancy in the full set of representers for the crossover data. Calculation of the dominant array modes by computing the SVD of $\mathbf{P}$ should thus result in significant reductions in the practical dimension of the state space. However, this approach would still require the calculation of all IL representers $\mathbf{p}_{il}$, and this is not practical. A simpler way to reduce the size of the inverse problem is to simply omit some of the representers, so that $v_{nk} = \delta_{nk}$ in (80). This is the approach taken by *Parker and Shure* [1982], for inversion of Magsat data for magnetic fields at the surface of the Earth's core.

As a preliminary reduction of the problem, we first winnowed the densely spaced high-latitude crossovers so that a minimum spacing of approximately 1 degree in latitude was maintained. Crossover points over land, or over seas not connected (on our numerical grid) to the open ocean were also eliminated from the inversion at this stage. This gave us a grid of $I = 6355$ crossover points, (the "full set") for which differences will be fit (see Figure 8). From these we then chose an approximately uniform subgrid of $I' = 986$ points (Figure 8), for which we calculated representers for evaluation of harmonically analyzed constituents ($\mathbf{p}_{i'l}$; $1 \le i' \le I'$; $1 \le l \le L$). For this calculation we took L to be four (i.e., constituents M_2, S_2, K_1, O_1), and used the interconstituent dynamical error correlation suggested by Table 2 (so that cross representers (defined in (67)) between the pairs M_2, S_2 and K_1, O_1 were calculated). Note that we assume that the full set of I crossover locations is

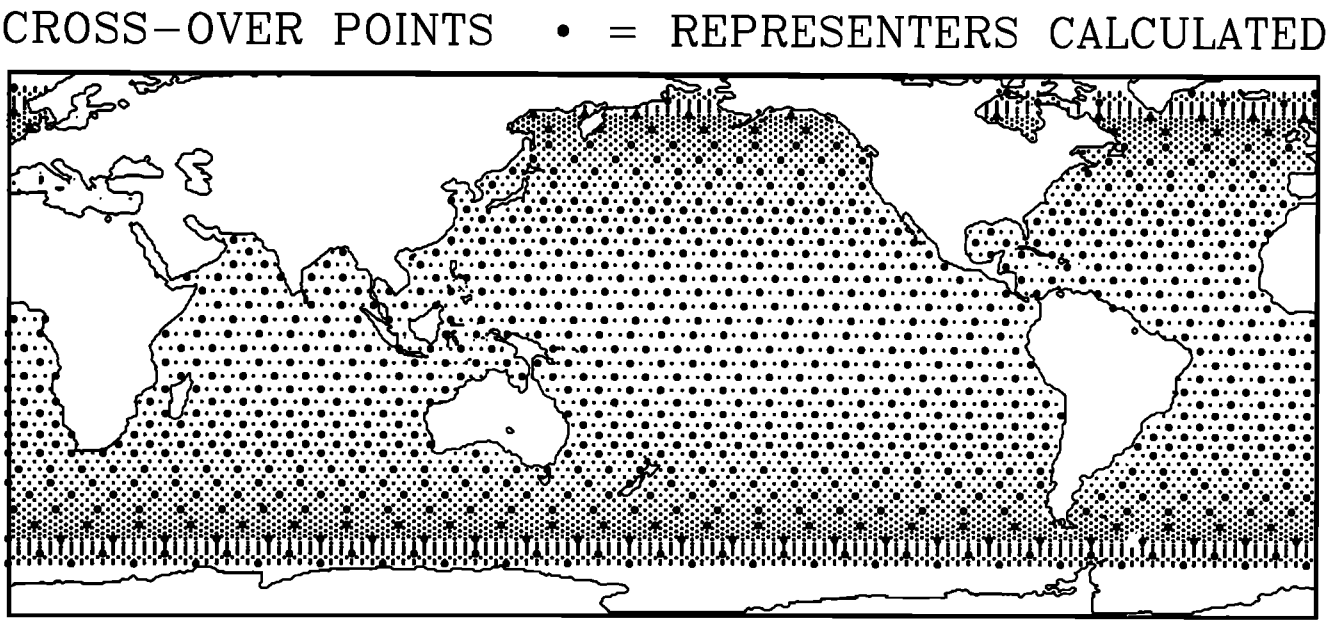

Figure 8. TOPEX/POSEIDON ocean crossover points used for this study. Representers were calculated only for the winnowed subset of 986 crossover points denoted by the large solid circles. Differences from all 6355 crossover points (small dots and large solid circles) were included in the data misfit penalty.

ordered so that the first I' constitute the subgrid used for the representer calculation.

For some purposes it may be useful to further reduce the set of $I'L(=3944)$ calculated representers. To discuss how this might best be done, it is useful to first partition the full $IL \times IL$ harmonically analyzed complex representer matrix $\mathbf{P}$ as

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} . \quad (81)$$

Here $\mathbf{P}_1$ contains the first I' columns of $\mathbf{P}$, and $\mathbf{P}_{11}$ contains the first I' rows of $\mathbf{P}_1$. Note that from the reduced set of calculated representers we can compute $\mathbf{P}_1$, but not $\mathbf{P}_2$. With the inverse solution restricted to the reduced set, the predicted data vector $L[\hat{\mathbf{u}}]$ must be a linear combination of $L[\mathbf{u}_0]$, and the columns of $\mathbf{P}_1$. By computing the SVD of this matrix: $\mathbf{P}_1 = \mathbf{Y}\Lambda\mathbf{W}^*$, we effectively find the linear combinations of these columns, and the corresponding linear combinations of representers

$$\mathbf{r}'_n = \sum_{i=1}^{I'} \sum_{l=1}^{L} w^*_{n(il)} \mathbf{p}_{il} \quad (82)$$

which are most efficient at approximating the tidal signal in the data (i.e., the part of the signal observable at the full set of crossover points). For our problem the singular-value spectrum of $\mathbf{P}_1$ falls off rapidly (Figure 9), implying that most of the signal in the crossover differences can be explained with a state space of relatively low dimension (say $N = 1000$ or less). It thus seems reasonable to seek a solution to the inverse problem in the space spanned by the reduced set of rotated representers $\mathbf{r}'_n$, $n = 1, N$ which correspond to the largest singular values of $\mathbf{P}_1$. In terms of the general formulation of (80), our two-stage reduction of the state space dimension amounts to chosing

$$\begin{aligned} v_{n(il)} &= w_{n(il)} && i \le I' \\ v_{n(il)} &= 0 && i > I' , \end{aligned} \quad (83)$$

for $n = 1, N$; $i = 1, I$; $l = 1, L$. We use $\mathbf{V}$ to denote the $N \times IL$ matrix of coefficients $v_{n(il)}$.

For the time-domain crossover data, with states restricted to the space defined by (80) and (83), the predicted data vector can be expressed in terms of the $2N$-dimensional real vector of representer coefficients $\bar{\mathbf{b}}$, the real matrices $\bar{\bar{\mathbf{V}}}$, $\bar{\bar{\mathbf{P}}}$, and $\mathbf{A}$, and the prior solution $\mathbf{u}_0$ as

$$L[\hat{\mathbf{u}}] = L[\mathbf{u}_0] + \mathbf{A}\bar{\bar{\mathbf{P}}}\bar{\bar{\mathbf{V}}}\bar{\mathbf{b}} , \quad (84)$$

while the dynamical error penalty is

$$\bar{\mathbf{b}}^T\bar{\bar{\mathbf{V}}}^T\bar{\bar{\mathbf{P}}}\bar{\bar{\mathbf{V}}}\bar{\mathbf{b}} . \quad (85)$$

Then, assuming an isotropic data error covariance, and proceeding as in (75)-(76) we can reduce the penalty functional to the standard form (77). Specifically, we first form and Q-R decompose the matrix

$$\mathbf{A}\bar{\bar{\mathbf{P}}}\bar{\bar{\mathbf{V}}}[\bar{\bar{\mathbf{V}}}^T\bar{\bar{\mathbf{P}}}\bar{\bar{\mathbf{V}}}]^{-1} = \mathbf{QU} . \quad (86)$$

As in (76), we set

$$\begin{aligned} \mathbf{d}' &= \mathbf{Q}^T(\mathbf{d} - L[\mathbf{u}_0]) \qquad \bar{\mathbf{b}}' = \mathbf{U}^{-T}\bar{\mathbf{b}} \\ \mathbf{R}' &= \mathbf{U}\bar{\bar{\mathbf{V}}}^T\bar{\bar{\mathbf{P}}}\bar{\bar{\mathbf{V}}}\mathbf{U}^T \qquad M^2_{\min} = \|\mathbf{d} - \mathbf{QQ}^T\mathbf{d}\|^2 . \end{aligned} \quad (87)$$

Note that by using row update methods for the Q-R decomposition [*Golub and Van Loan*, 1989] the computations of (86) and (87) can be efficiently carried out for essentially any number of rows in $\mathbf{A}$. Substituting $\mathbf{d}'$, $\mathbf{b}'$, $\mathbf{R}'$, and $M^2_{\min}$ into (84) and (85) and simplifying, it is easy to check that the resulting penalty functional J is exactly of the form (77). We may thus proceed with calculation of the inverse solution (including iterative refinement of $\delta\mathbf{u}$ on a finer grid, etc.) exactly as above.

In summary, we have found linear combinations of the crossover differences for which the inverse problem is reduced to the "standard form" (i.e., the same number of data functionals as representer coefficients). Now, however, the dimension of the reduced data vector $\mathbf{d}'$ is much smaller than the actual number of linearly independent data. This additional reduction is accomplished by throwing out the portion of the state space which we expect to have only a weak effect on any observable data. The reduction is accomplished in two steps. First, we winnow the full set of $I = 6355$ ocean crossovers to an evenly spaced set of $I' = 986$ crossovers, at each of which representers for $L = 4$ constituents are calculated. Second, using the SVD of the resulting representer matrix we select $N \approx 1000$ dominant array modes as a basis for the subspace in which to seek the inverse solution. The full data vector (I crossovers, M orbit cycles) and the dynamics are then fit using the $2N$ degrees of freedom selected in these two steps.

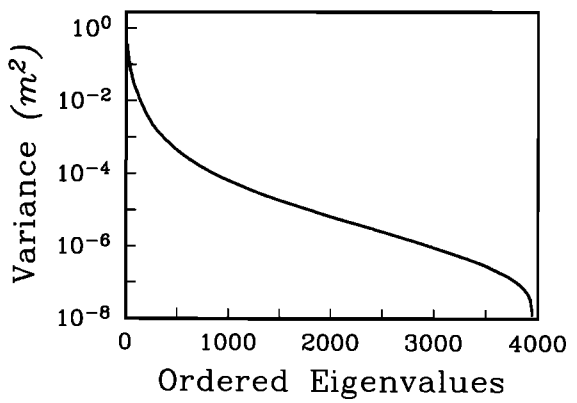

Figure 9. Singular-value spectrum for the rectangular matrix $\mathbf{P}_1$ computed for $I' = 986$ crossover points, and $L = 4$ constituents, with the interconstituent correlation structure of Table 2.

Initial Crossover Inversion Results

We now give results of an initial application of our inversion scheme to T/P data. In many respects the results given here are preliminary. Numerous refinements are currently underway: increasing L from 4 to 8, allowing for nonuniform and correlated data errors, computing the final solution on an even finer (1024×512) grid, and including coastal tide gauges in the inversion. For the inversion results discussed here we used TOPEX GDR data from orbit cycles 1-40. No data from the POSEIDON altimeter were used; in particular, all of orbit cycles 20 and 31 were excluded. Before forming the crossover differences, standard corrections were

applied (solid Earth and pole tide, electromagnetic bias, and inverse barometer). As a simple check for large outliers, crossover differences were then compared to the SCH80 tidal correction on the GDR. Differences which deviated from the SCH80 prediction by more than 1.0 m were rejected, leaving a total of K = 179,541 data for the inversion. After the initial screening, data were corrected for long-period constituents M_m, M_f, and S_{sa} (which are not included in our inversion) using the standard equilibrium forms. The resulting crossover data set has an RMS amplitude of 0.4574 m. Subtracting the prediction of the prior model ($\boldsymbol{L}[\mathbf{u}_0]$ where $\mathbf{S}\mathbf{u}_0 = \mathbf{f}_0$) reduces the RMS crossover difference to 0.2330 m.

Using the methods given above, we have computed inverse solutions for three different array mode truncation levels: N = 512, 1024, and 2048. This allows at least a partial check on the validity of our rather severe state space truncation. Eigenvalue spectra for the $2N \times 2N$ real reduced representer matrices $\mathbf{R}'$ (see (87)) are given for these three truncation levels in Figure 10a. The three eigenvalue spectra are very similar for the largest $\approx$ 400 eigenvalues. Beyond this point the eigenvalues for N = 256 diverge, and decrease relatively more rapidly. The spectra for N = 1024 and N = 2048 remain similar up to approximately eigenvalue number 1000 before a similar divergence is exhibited. As with the simpler case of harmonically analyzed data discussed in the previous section, eigenvalues $\lambda_k \ll \sigma^2$, where σ^2 is the "noise" (i.e., nontidal) variance, correspond to linear combinations of the data with poor signal-to-noise ratios. These data are not important to the inverse solution. To assess the significance of the differences in eigenvalue spectra seen in Figure 10a, we thus need an estimate of σ^2.

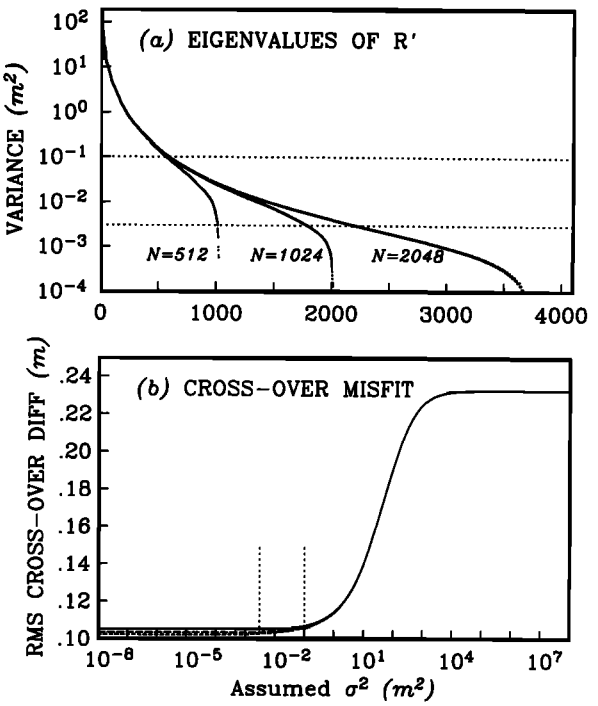

Figure 10. By restricting the final model fit to the dominant "array modes" we make the computational problem more tractable and stablize the inversion. Here we compare truncation of the ≈4000 calculated complex representers to N = 512, 1024, and 2048 array modes, resulting in reduced (real Hermitian) representer matrices $\mathbf{R}'$ (see(87)) of dimension 1024, 2048, and 4096. For these three cases we plot (a) eigenvalue spectra and (b) the variation in RMS misfit achieved when the assumed value for the crossover data error variance (σ^2) is varied. In both plots the range of plausible values for σ^2 is demarked by the dashed lines. Comparing the divergence of the three spectra in Figure 10a to the likely range of σ^2 suggests that inversion results will be relatively insensitive to the exact degree of truncation, beyond $\approx N$ = 1024. The small change in minimum RMS crossover difference achieved by increasing N is verified in Figure 10b. Note that the minimum misfit is very nearly achieved for a wide range of plausible values for σ^2. Note also that for our preferred values of σ^2 of $\approx 3\times10^{-2}$ m^2, the effective number of parameters fit is approximately 1000.

The minimum achievable RMS misfit of the crossover data for all choices of truncation level is approximately 0.1 m (0.1055 m for N = 512; 0.1040 m for N = 1024; 0.1032 m for N = 2048; see Figure 10b). This is consistent with RMS crossover residuals for T/P data obtained with other tidal corrections (e.g., SCH80, CR91), [*Schrama and Ray*, this issue; *Ma et al.*, this issue; *Le Provost et al.*, this issue), and suggests $\sigma^2 \approx 0.01$ m^2. This value is supported by by the flattening of the RMS misfit curves near this value (Figure 10b), and by cross-validation analysis of the reduced data vector and representer matrix. Based on this evidence (and allowing for possible misspecification of dynamical errors), we take 3×10^{-3}–10^{-1} m^{-2} as a reasonable range for σ^2. The divergence of the eigenvalue spectrum for N = 512 occurs at a variance of $\approx 2\times10^{-1}$m^2, at least somewhat above the likely range of σ^2. This implies that truncation at N = 512 array modes omits at least some useful signal from the inverse model and is thus almost certainly too severe. The divergence between the spectra for N = 1024 and N = 2048 occurs at a variance of 5×10^{-3}m^{-2}, near the lower end of the likely range for σ^2. This suggests that the difference in inverse solutions between these two truncation levels should be minimal. This conclusion is supported by the relatively small change in minimum achievable RMS crossover difference, and by comparison of inversion results for the two cases. For further results and comparisons discussed here, we focus on inversions done with the truncations set at N = 1024 array modes, and a value of $\sigma^2 = 3\times10^{-2}$ m. Experiments with other values of σ^2 in the preferred range 3×10^{-3}–10^{-1} m^{-2}, show only small changes in the RMS crossover residual (Figure 10b), or in the resulting inverse solutions.

The actual inverse solution (which we will refer to subsequently as TPXO.1) was computed by solving (78) iteratively on a 512×256 grid, for the four principal tidal constituents M_2, S_2, K_1, and O_1. Amplitude and phase for the M_2 constituent of the inverse solution are given in Plate 3. The result is qualitatively very similar to previously available tidal models and clearly exhibits the well-known dominant features of the M_2 tide (e.g., the equatorial Pacific 3/2 wave [*Platzman et al.*, 1981]). Locations of amphidromes are generally consistent with previous empirical determinations. For example, in TPXO.1 the California amphidrome, which is associated

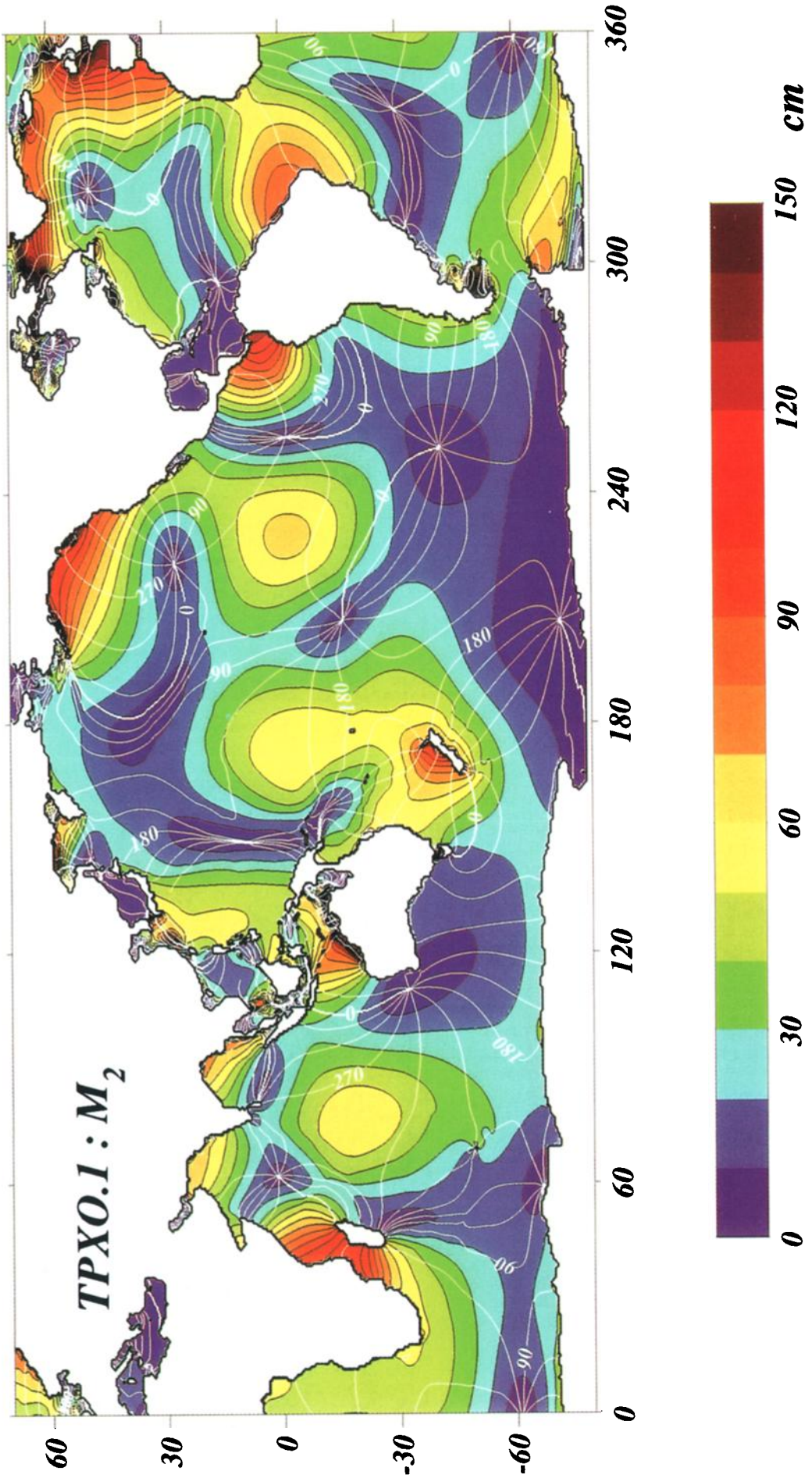

Plate 3. The M_2 constituent of TPXO.1. Contour interval is 10 cm for amplitude, 30° for phase. The model is grossly similar to SCH80 and CR91 but is noticeably smoother.

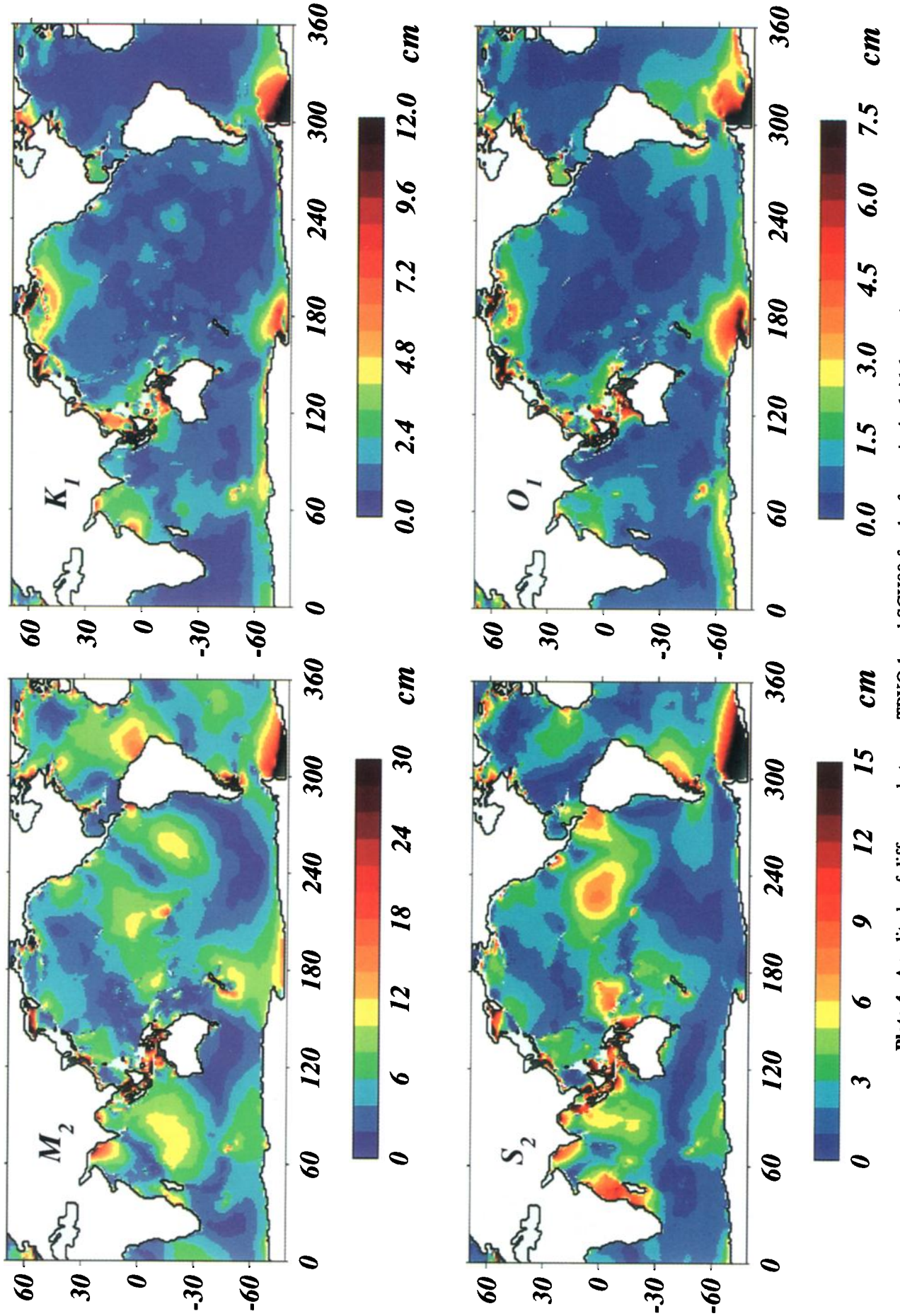

Plate 4. Amplitude of difference between TPXO.1 and SCH80 for the four principal tidal constituents, M_2, S_2, K_1, and O_1.

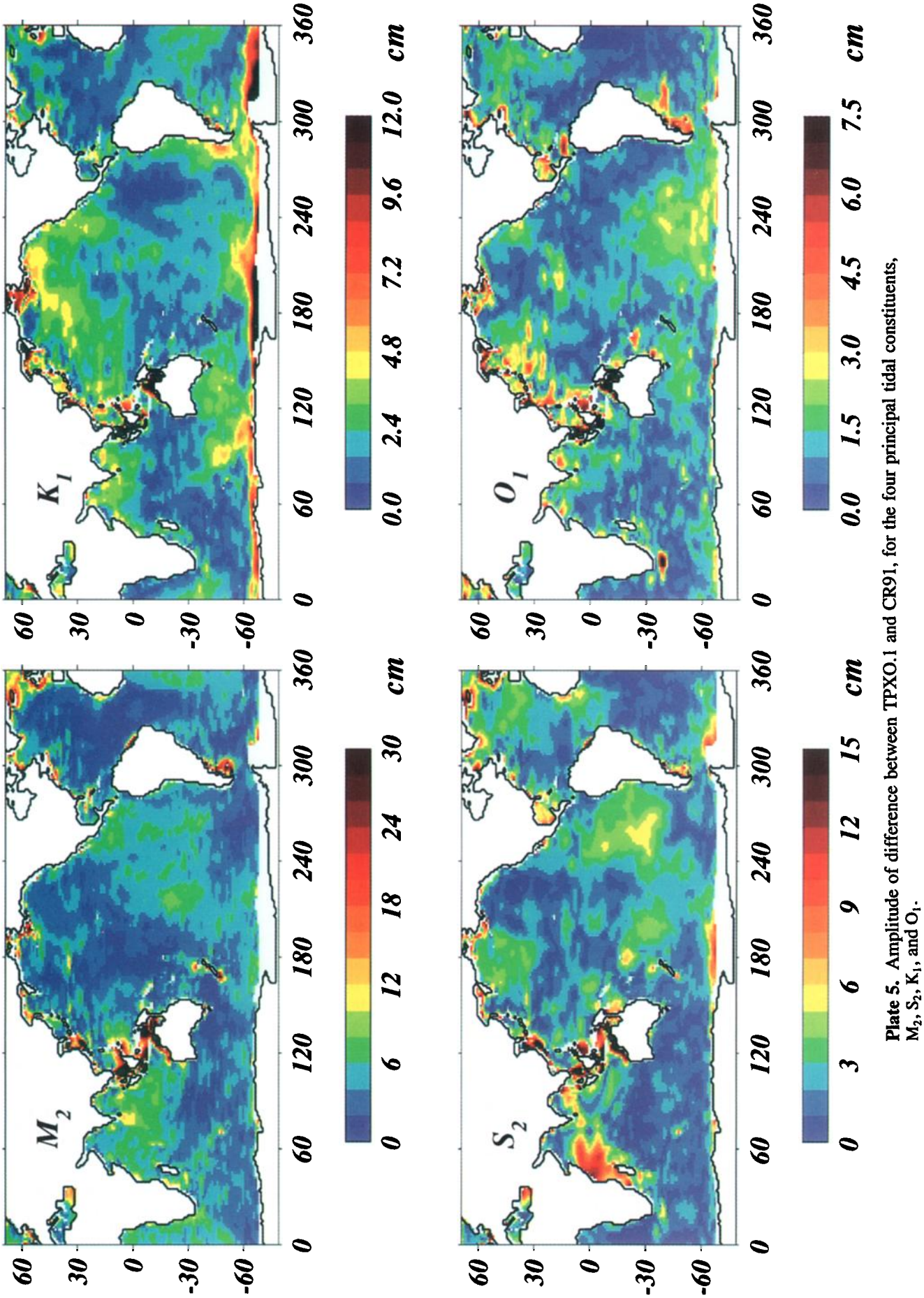

Plate 5. Amplitude of difference between TPXO.1 and CR91, for the four principal tidal constituents, M_2, S_2, K_1, and O_1.

with northward propagation of the tide along the west coast of North America, is located at 27.7°N 222.2°E. This is right in the middle of four previous determinations of the amphidrome location reviewed by *Platzman et al.* [1981]. TPXO.1 is very smooth, particularly when compared to the purely empirical (and unsmoothed) CR91, but also relative to SCH80. Further differences between TPXO.1 and previous proposed models are more easily seen in difference plots.

In Plate 4 we plot amplitudes of differences between TPXO.1 and SCH80 for the four constituents. The corresponding results for CR91 are given in Plate 5. The light grey areas in these plots correspond to areas where either SCH80 or CR91 are not defined. In general there are large differences between TPXO.1 and the other models around Antarctica. Since there are no T/P data to constrain our inverse solution in this area, TPXO.1 must be considered suspect in this area. In the future we intend to incorporate such tide gauges as are available to help constrain our solution here. Note also that the relative roughness of CR91 can be clearly seen in the rough and "patchy" nature of the the difference amplitudes of Plate 5.

For M_2 the biggest difference between TPXO.1 and SCH80 ($\approx$ 15 cm) is in the North Atlantic between South America and Africa (Plate 4). Other large areas with significant differences ($\approx$ 10 cm) are found in the South Atlantic, the Eastern Pacific, and over most of the Indian Ocean. For CR91 the differences for the M_2 constituent are generally somewhat smaller (Plate 5). However, significant differences (5-10 cm) still occur over much of the Indian Ocean, large areas in the South Pacific, and in the Atlantic off the coast of Africa. There are also many small areas where TPXO.1 differs from one or both of the other models by 20 cm or more. Examples include the areas around New Zealand, the Kerguelen Plateau, the Patagonian shelf, the North Atlantic near Europe, and throughout the Indonesian Archipelago. These are all areas where the tides are spatially complex. It is likely that all current models have some shortcomings in these areas.

In contrast to M_2, for the K_1 constituent TPXO.1 appears to be in slightly better agreement with SCH80 (Plate 4) than with CR91 (Plate 5). Differences with both models of order 3-5 cm are seen in the North Pacific, where diurnal tides are large. For CR91, the area of disagreement covers almost the whole of the Pacific north of the equator, including areas where diurnal tides are not particularly large. Significant differences between CR91 and TPXO.1 are also seen throughout the Southern Ocean and in the South Atlantic. Patterns of differences for O_1 (Plates 4 and 5) are broadly similar to those for K_1.

For S_2 (Plates 4 and 5) discrepancies with both models are greatest in the Indian Ocean, especially near the Somali Current. Other significant differences for the S_2 constituent occur in the equatorial Eastern Pacific and the South Atlantic off the coast of South America for SCH80, and for CR91 in scattered patches across much of the South Pacific.

Reductions in residual crossover variance achieved by TPXO.1, SCH80, and CR91 are summarized in Table 3. Note that only crossovers available for all models (177,939 of the 179,541 data points used for the inversion) are included in this comparison, so the results are slightly different from those quoted above. Relative to SCH80 and CR91, TPXO.1 achieves crossover variance reductions of 29% and 20% respectively. The last two lines of Table 3 give results for two variants on TPXO.1. TPXO.S is the inverse solution computed using SCH80 as the prior model $\mathbf{u}_0$, and inverting for only the the four largest constituents. For TPXO.S2 the prior was taken, as for TPXO.1, to be the solution to the LTEs $\mathbf{S}\mathbf{u}_0 = \mathbf{f}_0$. However, in this case only the four principal constituents were fit in the inversion. Rather than use the admittance approach to account for secondary constituents, these were now taken directly from SCH80. These variants provide modest, but nonnegligible, reductions in the residual crossover variance.

Table 3. Comparison of Tidal Models: T/P Crossovers

	RMS, m	Variance, m^2
Raw Crossovers	0.4524	2.047×10^{-1}
Prior ($\mathbf{u}_0$)	0.2279	5.192×10^{-2}
SCH80	0.1199	1.418×10^{-2}
CR91	0.1120	1.254×10^{-2}
TPXO.1	0.1003	1.005×10^{-2}
TPXO.S2	0.0993	0.986×10^{-2}
TPXO.S	0.0984	0.968×10^{-2}

Summary of residual T/P crossover difference RMS and variance for raw crossovers and various models. Only crossovers available for all models are used for the comparison. Relative to SCH80 and CR91, TPXO.1 achieves variance reductions of 29% and 20%, respectively. TPXO.S2 and TPXO.S are variants on the inverse solution TPXO.1. For TPXO.S2 SCH80 was used to correct for secondary constituents, and only the four largest constituents were inverted for. TPXO.S is the inverse solution computed by inverting for only the four largest constituents, while using SCH80 as the prior model. These veraiants provide modest reductions in the residual crossover variance.

To further validate the inverse solution we compare TPXO.1 to harmonic constants from the pelagic and island tide gauges of Figure 1 (i.e., the "ground truth" data of *Ray* [1993]). Note that two sites (IAPSO 1.1.42 and Campbell Island), which have been identified as erroneous (C. Le Provost, personal communication, 1994), are omitted from this comparison, which is summarized in Table 4. Relative to CR91, TPXO.1 significantly improves the fit to the tide gauges for all constituents. Compared to SCH80, TPXO.1 reduces the misfit for all constituents except S_2. TPXO.S2 and TPXO.S provide even better fits to the validation data set, particularly for the semidiurnal species. The improvement in fit is particularly noticeable for S_2, which Table 4 suggests is the most poorly resolved of the four constituents in TPXO.1. A more detailed comparison to the pelagic gauges reveals that the largest discrepancies for the S_2 constituent of TPXO.1 occur at equatorial latitudes. It is possible that we have failed to resolve adequately S_2 at low latitudes with the inversion due to limitations of the within-cycle crossover difference data set. An analysis of phase shifts of tidal constituents between ascending and descending tracks at

Table 4. Comparison of Tidal Models: RMS Misfits to Ground Truth Data

	M_2	S_2	O_1	K_1	Total
SCH80	0.0411	0.0186	0.0150	0.0178	0.0508
CR91	0.0372	0.0263	0.0133	0.0196	0.0496
TPXO.1	0.0301	0.0211	0.0123	0.0155	0.0418
TPXO.S2	0.0280	0.0188	0.0124	0.0155	0.0392
TPXO.S	0.0276	0.0161	0.0120	0.0132	0.0366

RMS misfits for the "ground truth" pelagic tide gauge data set of *Ray* [1993] for SCH80, CR91, and three TPXO tidal models (units are meters). TPXO.1 is the full inverse solution in which all secondary constituents are accounted for by interpolation of the four largest constituents. Note that for the TPXO tidal solutions two sites (IAPSO 1.1.42 and Campbell Island) which have been identified as erroneous (C. LeProvost, personal communication, 1994) have been omitted.

the crossover points [*Schrama and Ray*, this issue] shows that the S_2 signal in the crossovers will generally be very weak at low latitudes. Including additional data (e.g., between cycle crossovers) may well significantly improve the inverse solution. It is also possible that the difficulties with S_2 lie with the calibration gauges. A more careful comparison to other tide gauge data, and to tidal models obtained from the T/P data with other empirical approaches is warranted.

Note that while the analysis of *Cartwright and Amin* [1986] suggests the harmonic constants should generally have a precision of 1 cm or better, closer examination often reveals significantly larger variations in reported values from a single gauge [e.g., *Ray*, 1993]. Indeed, in the data set we used, amplitudes were only given to the nearest centimeter. Considering that some of the validation gauges are on islands where local effects may not be negligible, and that pelagic gauges require some assumptions about the water column to convert measured pressure variations into sea level, it is not at all clear how well a reasonable global scale model should fit these validation gauges.

The relatively better fit of TPXO.S2 and TPXO.S to both the crossover and the tide gauge validation data raises several important issues. While TPXO.1 is completely independent of the tide gauge data, SCH80 is not (some of the validation gauges were used as boundary conditions for this model). Thus TPXO.S and (to a lesser extent) TPXO.S2 are also intertwined with the tide gauge validation data set in a manner which is difficult to unravel. It is thus not clear that the results of Table 4 should be taken to mean that TPXO.S is closer to the truth than TPXO.1. After all, directly inverting these pelagic gauges resulted in a much better fit than is achieved by any of the models of Table 4 (see Table 1). Note that this same complication exists for validating and comparing other tidal models estimated as corrections to SCH80 [e.g., *Schrama and Ray*, this issue].

Interpretation of the differences in RMS crossover residuals seen in Table 3 is perhaps clearer. These results suggest that TPXO.S and TPXO.S2 should provide a slightly more accurate tidal correction for the T/P altimetry data and point to a significant shortcoming of TPXO.1. So far we have only directly included M_2, S_2, K_1, and O_1 in our inversion, with all of the remaining tidal constituents accounted for by straight-line interpolation in frequency (see Appendix B). The reduction in RMS crossover difference, when the admittance approach is not used to account for these secondary constituents (model TPXO.S2), strongly suggests that this approach is too crude. This is perhaps not too surprising, when one considers that there are resonant frequencies in, or at least very near the frequency band we are interpolating through [*Platzman et al.*, 1981]. Indeed, for the semidiurnal constituents the differences between TPXO.1 and TPXO.S bear a very strong resemblance to the two nearly semidiurnal frequency modes 34 and 35 of *Platzman et al.* [1981]. We will almost certainly be able to significantly improve our inverse solution by including more constituents in the state space, thus allowing for a more complex variation with frequency of the admittance curve.

While the comparisons of Tables 3 and 4 suggest that tidal models which make use of SCH80 as a starting point might be more accurate than the "pure" inverse model TPXO.1, it is far from clear that such models would provide a better tidal correction for the T/P GDR. TPXO.1 is demonstrably smoother than either of TPXO.S or TPXO.S2. Any roughness in SCH80 (see Plate 2 for example) will remain in the final solution. Since it is the gradient of the surface elevation which is ultimately diagnostic of the ocean circulation, smoothness of the tidal correction is critical. For this reason TPXO.1 is our preferred model for now.

5. Conclusions

We have presented a general and practical approach to inversion of very large large data sets for global ocean tides. The method allows us to combine rationally both dynamical information and direct observational data in a single tidal solution. All efforts to use altimetry data to improve empirically our knowledge of the open-ocean tides depend critically on the long-known temporal properties of the tides: at a fixed location the temporal variation of the tides is well approximated as a linear combination of a very small number of sines and cosines. This property, which derives from the nature of the forcing, says nothing about spatial variations of the tides. By using in addition information about the nature of the dynamics, we can derive a rational, dynamically consistent spatio-temporal set of basis functions for fitting the tides. In effect, this is what we have accomplished with the representer approach developed here. Using the dynamics, and estimates of the inadequacies in these dynamics, we have a priori reduced the free parameters

which are fit to the altimetry data to a very small number. For TPXO.1, only 2048 real parameters (i.e., the $2N$ representer coefficients of (84) were fit to the data. In fact, eigenvalues of $\mathbf{R}'$ below the assumed measurement error variance ($\sigma^2 = 2\times10^{-2}\text{m}^2$ was used for TPXO.1) correspond to array modes which are essentially not fit to the data. Thus for TPXO.1 the effective number of parameters fit is approximately 1000 (see Figure 10). This corresponds to only 125 complex numbers per constituent for the full global solution. Because we fit so few free parameters, and because the spatiotemporal basis functions used have been "tuned" for the appropriate tidal dynamics, we minimize the possibility that nontidal sources of oceanographic signal (the true focus of the T/P mission) will be aliased into the tidal corrections. Specifically, we are rejecting shallow water reduced gravity disturbances, having either nontidal frequencies or else phase speeds other than the external speed.

There are several further advantages to our inversion approach, which we can only list here. In addition to sea level, our solutions provide direct estimates of tidal currents. We can provide realistic maps of prior and posterior error covariances for tidal elevations (and currents), and we can assess the value of particular additional observations in improving our knowledge of the ocean tides. In a similar vein, prior and posterior data error covariances can be calculated to more quantitatively assess the degree to which our inversion removes nonoceanographic signal. We can also use the dynamical residuals (a byproduct of the inversion which we have not discussed in detail) to improve our understanding of inadequacies in the hydrodynamic model and to map dissipation in the ocean. Finally, additional sources of data (in particular, coastal and pelagic tide gauges) can be readily incorporated into our inverse solution. These issues will be addressed more fully in future work.

Our initial inverse solution TPXO.1 is very smooth. Based on the reduction of RMS crossover differences and comparison to pelagic gauges, we believe that this solution represents a significant improvement over the tidal models SCH80 and CR91 which are currently on the GDR. With refinements currently underway (increasing the number of modeled constituents, including tide gauge data and additional altimetry, calculations on higher resolution grids), further improvements are expected in the near future.

Note added in proof. A new eight constituent version of the OSU inverse solution (TPXO.2) has now been completed, and is available from the authors via anonymous ftp (ftp.oce.orst.edu). Relative to TPXO.1, the new solution significantly improves fits to T/P crossover data and to validation tide gauges. For more information send e-mail to tides@oce.orst.edu.

Appendix A. The Tidal State Space: Mathematical Details

Here we give a more precise definition of the state space τ and briefly sketch a proof that τ is a reproducing kernel Hilbert space. Note that all functions here are vector or tensor valued and that the domain of these functions includes separate interior and boundary portions. We do not explicitly discuss these technical complications here.

Because the dynamical error covariance is positive semidefinite we may decompose the covariance $\mathbf{C}_f(\mathbf{x}_1, \mathbf{x}_2) = \langle \mathbf{C}_f(\mathbf{x}_1, \cdot)^{1/2}, \mathbf{C}_f(\cdot, \mathbf{x}_2)^{1/2} \rangle_2$, where $\langle \cdot, \cdot \rangle_2$ denotes the usual L_2 inner product on the ocean domain and boundary O. For each $\mathbf{x} \in \in O$, let $\mathbf{T}(\cdot, \mathbf{x})$ be the solution to the LTEs with forcing $\mathbf{C}_f^{1/2}(\cdot, \mathbf{x})$. We generate $\mathbf{C}_f^{1/2}$ by solving a diffusion equation (see Appendix C), so the forcing for $\mathbf{T}$ is a C^∞ function in both its arguments. By assuming a smooth coastline and smooth bathymetry, we can guarantee that the solutions to the LTEs (i.e., $\mathbf{T}(\cdot, \mathbf{x})$) also have nice properties [*Gilbarg and Trudinger*, 1988]. The definitions of "smooth" and "nice" may in practice be adjusted to meet our requirements. For instance, taking the coastal boundary and the bathymetry to be C^∞ functions, we can guarantee that $\mathbf{T}$ is C^∞ in both arguments [*Gilbarg and Trudinger*, 1988, Chapter 6]. In fact, we need only assume enough regularity (of $\mathbf{C}_f$, bathymetry, coastline) so that $\mathbf{T} \in L_2[O \times O]$. Note that this condition would be violated if a white noise covariance on the interior of the domain were assumed [*Bennett*, 1990].

Defining the operator

$$\mathcal{T}: L_2[O] \rightarrow L_2[O]$$

$$\mathcal{T}\mathbf{u}(\mathbf{x}) = \langle \mathbf{T}(\mathbf{x}, \cdot), \mathbf{u} \rangle_2 , \tag{A1}$$

we then define our state space to be the image of $L_2[O]$ under $\mathcal{T}$:

$$\tau = \mathcal{T}[L_2[O]] . \tag{A2}$$

Suppose that $\mathcal{T}$ is one-to-one, so that an inverse $\mathcal{T}^{-1}: \tau \rightarrow L_2[O]$ can be defined. Then for $\mathbf{u}, \mathbf{v} \in \tau$ we can define the inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathcal{T}^{-1}\mathbf{u}, \mathcal{T}^{-1}\mathbf{v} \rangle_2 . \tag{A3}$$

It is trivial to check that with this inner product, τ is a Hilbert space (completeness is "inherited" from L_2). It is also straightforward to verify that the inner product $\langle \cdot, \cdot \rangle$ given in (A3) is equivalent to that defined in (22).

The requirement that $\mathcal{T}$ be one-to-one deserves a brief comment. Uniqueness of the solution to the (dissipative!) LTEs implies that $\mathcal{T}$ will be one-to-one provided convolution with $\mathbf{C}_f^{1/2}$ (equivalently $\mathbf{C}_f$) is also one-to-one (or invertible). Clearly, this condition is also necessary for the inner product defined in (22) to make sense. In fact, by making the constraints on the continuity equation hard, we violate this condition. However, since the (assumed exact) continuity equation determines h once the transports u and v are given, we can eliminate the h component from the state vectors (and from the LTEs). For this modified state space, our covariance function is positive definite, and the corresponding convolution operator is invertible. Note that this elimination is necessary only in this formal argument: because we never need the inverse of the covariance, this complication does not arise in actual computations.

Finally, it is straightforward to verify that pointwise evaluation is a bounded linear functional on τ with the inner product defined in (A3). Let $L_\mathbf{x}$ denote evaluation at $\mathbf{x}$. For every $\mathbf{u} \in \tau$ there is $\mathbf{v} \in L_2[O]$ such that

$\mathbf{u} = \mathcal{T}\mathbf{v}$. But then by the definition of $\mathcal{T}$ in (A1) we have

$$L_{\mathbf{x}}\mathbf{u} = \langle \mathbf{T}(\mathbf{x}, \cdot), \mathbf{v} \rangle \tag{A4}$$

Suppose for some $\{ \mathbf{u}_n \in \tau : n=1, \infty \}$ $\langle \mathbf{u}_n, \mathbf{u}_n \rangle \to 0$. Then we have $\mathbf{v}_n = \mathcal{T}^{-1}\mathbf{u}_n$, with $\langle \mathbf{v}_n, \mathbf{v}_n \rangle_2 \to 0$, and $L_{\mathbf{x}}\mathbf{u}_n = \langle \mathbf{T}(\mathbf{x}, \cdot), \mathbf{v}_n \rangle_2 \to 0$, since we have assumed $\mathbf{T}$ is in L_2. This establishes that $L_{\mathbf{x}}$ is bounded. Together with the Riesz representation theorem [e.g., *Yosida*, 1980], this guarantees that τ is a reproducing kernel Hilbert space, so that of the representers derived in section 3 are indeed well defined.

Appendix B. Incorporating Nodal Corrections and Secondary Constituents

In this appendix we give the exact form of the function $\alpha_l(t)$, which incorporates the nodal corrections, and allows for secondary tidal constituents by interpolating the admittance.

For each constituent l, the partial tide with nodal corrections can be expressed as $h^l(\mathbf{x})\alpha_l(t)$, where

$$\alpha_l(t) = (1 + n_f^l(t))\exp[i(\omega_l(t-t_0) + V_l(t_0) + n_u^l(t))], \tag{B1}$$

with $n_f^l(t)$ and n_u^l providing the modulation of amplitude and phase [e.g., *Cartwright and Tayler*, 1971; *Schureman*, 1940].

To allow for secondary tidal constituents in the simplest fashion we consider a single tidal species (diurnal or semidiurnal) and take the two dominant constituents ((K_1, O_1) or (M_2, S_2)) in this species. Here we denote the frequencies for these two dominant constituents by ω_1 and ω_2. Suppose there are L' constituents in the species under consideration, and for $1 \le l' \le L'$ let

$$w_{1l'} = \frac{(\omega_2 - \omega_{l'})e_{l'}}{(\omega_2 - \omega_1)e_1} \qquad w_{2l'} = \frac{(\omega_{l'} - \omega_1)e_{l'}}{(\omega_2 - \omega_1)e_2} \tag{B2}$$

where $e_{l'}$ are the equilibrium tide amplitudes. By definition, the tidal admittance at frequency ω_l is $\mathbf{u}^l(\mathbf{x})/e_l$, where $\mathbf{u}^l(\mathbf{x})$ gives the tidal fields for constituent l. Then if we linearly interpolate (and possibly extrapolate) the tidal admittance between frequencies ω_1 and ω_2, we find that the tidal fields for a constituent at frequency $\omega_{l'}$ are approximately

$$\mathbf{u}^{l'} = w_{1l'}\mathbf{u}^1 + w_{2l'}\mathbf{u}^2 \ . \tag{B3}$$

Thus the contribution of this species to sea level time series can be approximated

$$\sum_{l'=1}^{L'} \alpha_{l'}(t)h^{l'}(\mathbf{x}) = \sum_{l'=1}^{L'} \alpha_{l'}[w_{1l'}h^1(\mathbf{x}) + w_{2l'}h^2(\mathbf{x})]$$
$$= \sum_{l=1}^{2}\left[\sum_{l'=1}^{L'} w_{ll'}\alpha_{l'}(t)\right]h^l(\mathbf{x}) \ . \tag{B4}$$

The terms in brackets are the modified forms for the functions $\alpha_l(t)$ of (3) and (5). In fact, it would be desirable to use a more complicated interpolation scheme to include minor constituents [*Le Provost et al.*, 1991], but this will not be possible until we include more than two constituents per species in the state space. Clearly, more complex schemes for interpolation of secondary constituents (e.g., the orthotide expansion of *Groves and Reynolds* [1975]) can be accommodated within the general framework outlined here.

Appendix C. Dynamical Error Covariance: Numerical Implementation

In this appendix we discuss implementation of the dynamical error covariance smoother on the CM-200 computer. As in the main text, we consider first the case of a single constituent of harmonically analyzed data, with the covariance given by (42). Estimation of the component variance functions $\sigma_f^2(\mathbf{x})$ is discussed in detail in section 3. We take these functions as given for our discussion here. Note that while the inverse covariance operator appears in the definition of the penalty functional, with the representer approach we never need construct this inverse. Such is not the case for direct schemes (gradient descent, simulated annealing) for minimizing the penalty functional. Indeed, we do not even need to construct the full covariance function $\mathbf{C}_f(\mathbf{x}, \mathbf{x}')$, all that is necessary is to be able to calculate $\mathbf{C}_f\eta$ for any η (see (36)). For the covariance model (42) this calculation can be accomplished in three steps: (1) rescale $\eta(\mathbf{x})$ by multiplying by $\sigma_j(\mathbf{x})$ (pointwise), (2) smooth the result by convolving with the correlation ψ, and (3) rescale again by multiplying the smoothed result of step 2 by σ_j. Only the smoothing of step 2 needs further elaboration.

For (42) to be a valid spatial covariance, it is necessary and sufficient that the real spatial correlation ψ be a positive semidefinite function (i.e., for every square integrable f, $\int f(\mathbf{x})d^2\mathbf{x}\int\psi(\mathbf{x}, \mathbf{x}')f(\mathbf{x}')d^2\mathbf{x}' \ge 0$), which is symmetric in the arguments $\mathbf{x}$ and $\mathbf{x}'$ [e.g., *Ripley*, 1981]. Beyond this, we have little information about the correct form for the spatial covariance of dynamical errors. Our choice of ψ is thus largely dictated by computational considerations.

Consider the diffusion equation on the unit sphere

$$\partial_t u - \gamma\nabla^2 u = 0 \tag{C1}$$

where ∇^2 is the two-dimensional surface Laplacian, and γ is the (spatially uniform) diffusion parameter. Let $G(t; \hat{\mathbf{x}}, \hat{\mathbf{x}}')$ be the time-dependent Greens function for this equation

$$\partial_t G - \gamma\nabla^2 G = 0 \qquad G(0; \hat{\mathbf{x}}, \hat{\mathbf{x}}') = \delta(\hat{\mathbf{x}}\cdot\hat{\mathbf{x}}') \ . \tag{C2}$$

For any pseudo-time t, $G(t; \hat{\mathbf{x}}, \hat{\mathbf{x}}')$ can be shown to be symmetric in $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$. Furthermore by expressing (C1) in spherical harmonics, it is readily demonstrated that G is positive-definite. For t near zero and $\mathbf{x}$ near $\mathbf{x}'$ it can be shown [e.g., *Watson*, 1983] that

$$G(t; \hat{\mathbf{x}}, \hat{\mathbf{x}}') \approx (4\pi\gamma t)^{-1/2}\exp\left[-\frac{\|\hat{\mathbf{x}}-\hat{\mathbf{x}}'\|^2}{4\gamma t}\right] \ . \tag{C3}$$

$G(t; \mathbf{x}, \mathbf{x}')$ is thus a spherical analogue of the usual Gaussian density function, which is often used as a spatial covariance function in planar geometry.

The spherical correlation function

$$\psi_{t_0}(\mathbf{x}, \mathbf{x}') = G(t_0; \hat{\mathbf{x}}, \hat{\mathbf{x}}')/G_0 \tag{C4}$$

where $G_0 = G(0; \mathbf{x}, \mathbf{x})$ is a normalization constant, is very well suited to our problem. As suggested by (C3),

the decorrelation length scale can be adjusted by varying t_0. Numerical computation of a convolution such as

$$\psi_{t_0} * u = \int \psi_{t_0}(\mathbf{x}, \mathbf{x}') u(\mathbf{x}') d^2\mathbf{x}' \qquad \text{(C5)}$$

is simple, and particularly well suited to the massively parallel architecture of the CM-200 used for our representer calculation. Letting $\mathbf{D}$ be the finite difference operator for ∇^2 on the grid of u nodes, computation of ψ_{t_0} can be accomplished by time-stepping

$$\psi_{t+1} = \psi_t + \gamma \mathbf{D} \psi_t \qquad \text{(C6)}$$

from an impulsive initial condition (scaled by G_0^{-1}) to the solution at time t_0. Clearly, the convolution of (C5) can be accomplished by doing the same time-stepping of (C6), but now starting from initial conditions of uG_0^{-1}.

Each time step of (C6) represents a local smoothing based on averages over adjacent grid points, with positive weights summing to 1. The full convolution with ψ is thus achieved by repeated application of a local smoother. The repeated smoothing may be represented in matrix notation as

$$\psi_{t_0} * u = G_0^{-1} (\mathbf{I} + \gamma \mathbf{D})^N u \ , \qquad \text{(C7)}$$

where N and γ are chosen to achieve the desired spatial decorrelation length scale. For the $\approx 5^\circ$ correlation function of Figure 3, we used $\gamma = 0.01$ and $N = 250$. Note that the relatively small value of γ is required to maintain stability of the iterative scheme at high latitudes where grid cells are very narrow.

So far we have ignored the irregular shape of the ocean domain to which the dynamical error convolution is confined. In terms of the diffusion equation (C2) we must add boundary conditions to accommodate this more complex domain. For our implementation of the covariance smoother we have assumed homogeneous Neumann boundary conditions ($\partial_n u = 0$). In terms of the local smoother on the discrete grid, this amounts to replacing the average over all adjacent nodes, with an average over only ocean nodes. The presence of continents makes our correlation function spatially inhomogeneous, particularly near the coasts. This does not seem to be a serious issue, given our limited knowledge of the spatial structure of dynamical errors. However, this means that the normalization factor G_0, now depends upon $\mathbf{x}$ and must be computed for each grid point (but only once).

With the simple form of interconstituent correlation assumed in (64), the off-diagonal smoothers $C_f^{ll'}$, $l \neq l'$ are also easily implemented. Indeed, if Σ_j^l is the diagonal matrix of dynamical error standard deviations $\sigma_j^l(\mathbf{x})$ on the numerical grid, and $\mathbf{G}_0$ is the diagonal matrix of normalization constants for all nodes in the grid, the matrix representation of the generalized interconstituent smoother $C_{fj}^{ll'}$ for component j on the discrete numerical grid is

$$\psi_{t_0} * u = \rho_{ll'} \Sigma_j^l \mathbf{G}_0^{-1/2} (\mathbf{I} + \gamma \mathbf{D})^N \mathbf{G}_0^{-1/2} \Sigma_j^{l'} u \ . \qquad \text{(C8)}$$

Since each matrix multiply in (C8) is local, involving only neighboring nodes, the convolution smoother is very fast on the CM-200.

Appendix D. Representers for Time Domain Data

The purpose of this appendix is to give a more precise definition of the notation used in section 4 to relate real and imaginary vectors and matrices. We also derive (70) and (71) which relate the real representers and representer matrix for time domain data to the complex quantities appropriate for harmonically analyzed data. To facilitate the translation between real and complex vector fields we define, for any complex n-dimensional vector field $\mathbf{v}$, the real $2n$-dimensional vector field

$$\bar{\mathbf{v}} = (\mathrm{Re}\ v_1\ \mathrm{Im}\ v_1\ \ldots\ \mathrm{Re}\ v_n\ \mathrm{Im}\ v_n)^T \ . \qquad \text{(D1)}$$

As above, the real and complex versions of the state space are denoted by τ and τ_R, respectively. Given an inner product $\langle .,. \rangle$ defined on τ, we can define a real inner product on τ_R by

$$(\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2) = \mathrm{Re} \langle \mathbf{v}_1, \mathbf{v}_2 \rangle \ . \qquad \text{(D2)}$$

It is easily verified that (D2) defines an inner product on τ_R. Clearly the norms induced by the real and complex inner products are the same, in the sense that $\|\bar{\mathbf{v}}\|_{\tau_R} = \|\mathbf{v}\|_\tau$. Thus the dynamic error penalty of (21) can be expressed in terms of either inner product. Finally, for any $N \times M$ complex matrix $\mathbf{X}$ we define the $2N \times 2M$ real matrix

$$\bar{\bar{\mathbf{X}}} = (\bar{\mathbf{x}}_1\ \overline{[i\mathbf{x}_1]}\ \cdots\ \bar{\mathbf{x}}_M\ \overline{[i\mathbf{x}_M]}) \ . \qquad \text{(D3)}$$

As in section 4 we denote the complex representer (relative to the $3L$-dimensional state space) of evaluation of h^l at $\mathbf{x}_i$ by $\mathbf{p}_{il}$, we let L_{im} be the linear functional for crossover datum d_{im}, and we take a_{iml} to be the complex constants defined in (6). Then, using the notation defined above, a simple calculation shows

$$\Big(\sum_{l=1}^{L} \overline{a_{iml}^* \mathbf{p}_{il}}, \bar{\mathbf{u}}\Big) = \mathrm{Re} \Big\langle \sum_{l=1}^{L} a_{iml}^* \mathbf{p}_{il}, \mathbf{u} \Big\rangle$$

$$= \mathrm{Re} \sum_{l=1}^{L} \langle \mathbf{p}_{il}, a_{iml} \mathbf{u} \rangle = \mathrm{Re} \sum_{l=1}^{L} a_{iml} h^l(\mathbf{x}_i) \ . \qquad \text{(D4)}$$

Comparing (6) and (D4) verifies (70).

The representer matrix for the full set of crossover differences d_{im}, $1 \le i \le I$, $1 \le m \le M$ is also easily calculated:

$$R_{(im)(i'm')} = (\overline{\mathbf{r}_{im}}, \overline{\mathbf{r}_{i'm'}}) = \mathrm{Re} \langle \mathbf{r}_{im}, \mathbf{r}_{i'm'} \rangle =$$

$$\mathrm{Re} \Big\langle \sum_{l=1}^{L} a_{iml}^* \mathbf{p}_{il}, \sum_{l'=1}^{L} a_{i'm'l'}^* \mathbf{p}_{i'l'} \Big\rangle = \mathrm{Re} \sum_{l,l'=1}^{L} a_{iml} a_{i'm'l'}^* \langle \mathbf{p}_{il}, \mathbf{p}_{i'l'} \rangle$$

$$= \mathrm{Re} \sum_{l,l'=1}^{L} a_{iml} a_{i'm'l'}^* P_{ii'}^{ll'} \ . \qquad \text{(D5)}$$

The last step follows from the fact that $P_{ii'}^{ll'}$ is defined to be the h component of $\mathbf{p}_{il'}^l(\mathbf{x}_i)$, while $\mathbf{p}_{il}$ is the representer for this evaluation functional. This verifies (71).

Acknowledgments. Calculations were performed on the COAS CM-200 and CM-5 computers, which are supported by NASA-EOS. We thank R. James and P. Palmer for assistance with programming, and D. Chelton, M. Schlax, and G. Jacobs for

providing us with T/P crossover data. The manuscript was improved by the helpful comments of G. W. Platzman, R. D. Ray and two anonymous reviewers. G.D.E. and A.F.B. were supported in part by NASA Grant NAGW 2715.

References

Accad, Y., and C. L. Pekeris, Solution of the tidal equations for the M_2 and S_2 tides in the world oceans from a knowledge of the tidal potential alone, *Philos. Trans. R. Soc. London, 290*, 235-266, 1978.

Backus, G., Bayesian inference in geomagnetism, *Geophys. J. R. Astron. Soc., 92*, 125-142, 1988.

Backus, G., Inference from inadequate and inaccurate data, *Lect. Appl. Math., 14*, pp. 1-105, Am. Math. Soc., Providence, R.I., 1971.

Bennett, A. F., Array design by inverse methods, *Progr. Oceanogr., 15*, 129-156, 1985.

Bennett, A. F., Inverse methods for assessing Ship-of-Opportunity networks and estimating circulation and winds from tropical expendable bathythermograph data, *J. Geophys. Res., 95*, 16,111-16,148, 1990.

Bennett, A. F., *Inverse Methods in Physical Oceanography, Monographs on Mechanics and Applied Mathematics*, 346ppp, Cambridge University Press, New York, 1992.

Bennett, A. F., and P. C. McIntosh, Open ocean modelling as an inverse problem: Tidal theory, *J. Phys. Oceanogr., 12*, 1004-1018, 1982.

Cartwright, D. E., Oceanic tides, *Rep. Prog. Phys., 40*, 666-708, 1977.

Cartwright, D. E., Detection of tides from artificial satellites, in *Advances in Tidal Hydrodynamics*, edited by B. Parker, pp. 547-568, John Wiley, New York, 1991.

Cartwright, D. E., and M. Amin, The variances of tidal harmonics, *Dtsch. Hydrogr. Z., 39*, 235-253, 1986.

Cartwright, D. E., and R. D. Ray, Oceanic tides from Geosat altimetry, *J. Geophys. Res., 95* 3069-3090, 1990.

Cartwright, D. E., and R. D. Ray, Energetics of global ocean tides from Geosat altimetry, *J. Geophys. Res., 96*, 16,897-16,912, 1991.

Cartwright, D. E., and R. J. Tayler, New computations of the tide-generating potential, *Geophys. J. R. Astron. Soc., 23*, 45-74, 1971.

Cartwright, D. E., R. Spencer, J. M. Vassie, and P. L. Woodworth, The tides of the Atlantic Ocean, 60° N to 30° S, *Philos. Trans. R. Soc. London, 324*, 513-563, 1988.

Farrell, W. E., Deformation of the Earth by surface loads, *Rev. Geophys., 10*, 761-797, 1972.

Filloux, J. H., D. S. Luther, and A. D. Chave, Update on seafloor pressure and electric field observations from the north-central and northeastern Pacific: Tides, infratidal fluctuation, and barotropic flow, in *Advances in Tidal Hydrodynamics*, edited by B. Parker, pp. 617-640, John Wiley, New York, 1991.

Francis, O., and P. Mazzega, Global charts of ocean tide loading effects, *J. Geophys. Res., 95*, 11,411-11,424, 1990.

Golub, G. H., and C. F. Van Loan, *Matrix Computation*, 2nd ed., 642 pp., Johns Hopkins University Press, Baltimore, Md., 1989.

Gilbarg, D. and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd ed., 512 pp., Springer-Verlag, New York, 1988.

Groves, G. W., and R. W. Reynolds, An orthogonalized convolution method of tide prediction, *J. Geophys. Res., 80*, 4131-4138, 1975.

Hendershott, M., The effects of solid Earth deformation on global ocean tides, *Geophys. J. R. Astron. Soc., 29*, 389-403, 1972.

Hendershott, M., Numerical models of ocean tides, in *The Sea* vol. 6, *Marine Modelling*, edited by E. D. Goldberg, I. N. McCave, J. J. O'Brien, and J. H. Steele, pp. 47-95, John Wiley, New York, 1977.

Hendershott, M., Long waves and ocean tides, in *Evolution of Physical Oceanography*, edited by B. Warren and C. Wunsch, MIT Press, Cambridge, Mass., 1981.

Jourdin, F. O., Francis, P. Vincent, and P. Mazzega, Some results of heterogeneous data inversions for ocean tides, *J. Geophys. Res., 96*, 20,267-20,288, 1991.

Koblinsky, C. J., P. Gaspar, and G. Lagerloef, *The Future of Space-borne Altimetry: Oceans and Climate Change* Joint Oceanographic Institutions, Washington, D. C., 1992.

Le Provost, C., Sur une novelle methode numerique pour calculen les marees oceaniques et littorales, *C. R. Acad. Sci., Ser. B*, 285, 349-352, 1977.

Le Provost, C., and F. Lyard, Towards a detailed knowledge of the world ocean tides: The example of Kerguelen Plateau, *Geophys. Res. Lett., 20*, 1519-1522, 1993.

Le Provost, C., and A. Poncet, Finite element method for spectral modelling of tides, *Int. J. Numer. Methods Eng., 12*, 853-871, 1978.

Le Provost, C., and P. Vincent, Finite element method for modelling ocean tides, in *Advances in Tidal Hydrodynamics*, edited by B. Parker, pp. 617-640, John Wiley, New York, 1991.

Le Provost, C., F. Lyard, and J.-M. Molines, Improving ocean tide predictions by using additional semidiurnal constituents from spline interpolation in the frequency domain, *Geophys. Res. Lett., 18*, 845-848, 1991.

Le Provost, C. L., M. L. Genco, F. Lyard, P. Vincent, and P. Canceil, Spectroscopy of the world ocean tides from a hydrodynamic finite element model, *J. Geophys. Res.*, this issue.

Luyten, J. R., and H. M. Stommel, Comparison of M_2 tidal currents observed by some deep moored current meters with those of Schwiderski and Laplace models, *Deep Sea Res., 38, suppl. 1*, S573-S589, 1991.

Ma, X. C., C. K. Shum, R. J. Eanes, and B. D. Tapley, Determination of ocean tides from the first year of TOPEX/POSEIDON altimeter measurements, *J. Geophys. Res.*, this issue.

McIntosh, P. C., and A. F. Bennett, Open ocean modeling as an inverse problem: M_2 tides in Bass Strait, *J. Phys. Oceanogr., 14*, 601-614, 1984.

Munk, W., and D. E. Cartwright, Tidal spectroscopy and prediction, *Philos. Trans. R. Soc. London, 259*, 533-581, 1966.

National Geophysical Data Center, GEODAS CD-ROM worldwide marine geophysical data, *Data Announce. 92-MGG-02*, Natl. Oceanic and Atmos. Admin. U.S. Dep. Commer., Boulder, Colo., 1992.

Parker, R. L., Understanding inverse theory, *Annu. Rev. Earth Planet. Sci., 5*, 35-64, 1975.

Parker, R. L., L. Shure, and J. A., Hildebrand, The application of inverse theory to seamount magnetism, *Rev. Geophys., 25*, 17-40, 1987.

Parker, R. L., and L. Shure, Efficient modelling of the Earth's magnetic field with harmonic splines, *Geophys. Res. Lett., 9*, 812-815, 1982.

Pekeris, C. L., and Y. Accad, Solution of Laplace's equations for the M_2 tide in the world oceans, *Philos. Trans. R. Soc. London A, 265*, 413-436, 1969.

Platzman, G. W., G. A. Curtis, K. S. Hansen, and R. D. Slater, Normal modes of the world ocean, II: Description of modes in the period range 8 to 80 hours, *J. Phys. Oceanogr., 11*, 579-603, 1981.

Press, W. H., B. P. Flanerry, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, 818 pp., Cambridge University Press, New York, 1986.

Ray, R. D., Global ocean tide models on the eve of TOPEX/POSEIDON, *IEEE Trans. Geosci. Remote Sens.*, *31*, 355-364, 1993.

Ray, R. D., and B. V. Sanchez, Radial deformation of the Earth by oceanic tidal loading, *NASA Tech. Memor., 1000743*, 1989.

Ray, R. D., C. J. Koblinsky and B. D. Beckley, On the effectiveness of Geosat altimeter corrections, *Int. J. Remote Sens., 12*, 1979-1984, 1991.

Reid, W. T., Generalized inverses of differential and integral operators, in *Theory and Applications of Generalized Inverses of Matrices*, edited by T. L. Boullion and P. L. Odell, pp. 1-25, Texas Technical College, Lubbock, 1968.

Ripley, B., *Spatial Statistics*, John Wiley, New York, 1981.

Schureman, *Manual of Harmonic Analysis and Prediction of Tides*, U.S. Department of Commerce, Coast and Geodetic Survey, Washington D. C., 1940.

Schrama, E. O. J., and R. D. Ray, A preliminary tidal analysis of TOPEX/POSEIDON altimetry, *J. Geophys. Res.*, this issue.

Schwiderski, E. W., Global ocean tides, I, A detailed hydrodynamical interpolation model, *Rep. NSWC/DL TR-3866*, Nav. Surf. Weapons Cent., Dahlgren, Va., 1978.

Schwiderski, E. W., Ocean tides, I, Global ocean tidal equations, *Mar. Geod., 3*, 161-217, 1980a.

Schwiderski, E. W., Ocean tides, II, A hydrodynamic interpolation model, *Mar. Geod., 3*, 219-255, 1980b.

Schwiderski, E. W., Global ocean tides, IV, The diurnal lunisolar declination tide (K_1), atlas of tidal charts and maps, *Rep. NSWC TR 81-142*, Nav. Surf. Weapons Cent., Dahlgren, Va., 1981.

Schwiderski, E. W., and L. T. Szeto, The NSWC Global Ocean Tide Data Tape (GOTD), its features and applications, *Rep. NSWC TR 81-264*, Nav. Surf. Weapons Cent., Dahlgren, Va., 1981.

Smith, W. H. F., On the accuracy of digital bathymetric data, *J. Geophys. Res., 98*, 9591-9603, 1993.

Tarantola, A., *Inverse Problem Theory. Methods for Data Fitting and Model Parameter Estimation*, 613 pp., Elsevier, New York, 1987.

Wahba, G., *Spline Methods for Observational Data*, 169 pp., Society for Industrial and Applied Mathematics, Philadelphia, 1990.

Wahba, G., and J. Wendelberger, Some new mathematical methods for variational objective analysis using splines and cross validation, *Mon. Weather Rev., 108*, 1122-1143, 1980.

Wahr, J. M., Body tides on an elliptical, rotating, elastic, oceanless Earth, *Geophys. J. R. Astron. Soc., 64*, 677-703, 1981.

Watson, G. S., *Statistics on Spheres*, 283 pp., John Wiley, New York, 1983.

Yosida, K., *Functional Analysis*, 6th ed., 500pp., Springer-Verlag, New York, 1980.

Youtsey, W. J., Report detailing modifications to the 1/8° global bathymetry, *Rep. NRL/MR/732393-7023*, Nav. Res. Lab., Stennis, Miss., 1993.

Zahel, W., *Proceedings of the IRIA International Colloquim on Numerical Methods of Science and Technical Computation*, Springer-Verlag, New York, 1977.

Zahel, W., Modeling ocean tides with and without assimilating data, *J. Geophys. Res., 96*, 20,379-20,391, 1991.

A. F. Bennett and G. D. Egbert, College of Oceanic and Atmospheric Sciences, Oregon State University, Oceanography Admin. Bldg. 104, Corvallis, OR, 97331-5503. (e-mail: bennett@oce.orst.edu; egbert@oce.orst.edu)

M. G. G. Foreman, Institute of Ocean Sciences, 9860 W. Saanich Rd., Sidney, BC. V8L 4B2, Canada. (e-mail: mforeman@mike.ios.bc.ca)

(Received March 21, 1994; revised June 16, 1994;
accepted June 20, 1994.)