Estimates of Baroclinic Tidal Sea Level and Currents from Lagrangian Drifters and Satellite Altimetry

Edward D. Zaron[a] and Shane Elipot[b]

[a] *Oregon State University, Corvallis, Oregon, USA*

[b] *Rosenstiel School of Marine, Atmospheric, and Earth Science, University of Miami, Miami, Florida, USA*

Corresponding author: Edward D. Zaron, edward.d.zaron@oregonstate.edu

ABSTRACT: Internal waves generated by the interaction of the surface tides with topography are known to propagate long distances and lead to observable effects such as sea level variability, ocean currents, and mixing. In an effort to describe and predict these waves, the present work is concerned with using geographically-distributed data from satellite altimeters and drifting buoys to estimate and map the baroclinic sea level associated with the M_2, S_2, N_2, K_1, and O_1 tides. A new mapping methodology is developed, based on a mixed L_1/L_2-norm optimization, and compared with previously-developed methods for tidal estimation from altimeter data. The altimeter and drifter data are considered separately in their roles for estimating tides and for cross-validating estimates obtained with independent data. Estimates obtained from altimetry and drifter data are found to agree remarkably well in regions where the drifter trajectories are spatially dense; however, heterogeneity of the drifter trajectories is a disadvantage when they are considered alone for tidal estimation. When the different data types are combined by using geodetic-mission altimetry to cross-validate estimates obtained with either exact-repeat altimetry or drifter data, and subsequently averaging the latter estimates, the estimates significantly improve on the previously-published HRET8.1 model, as measured by their utility for predicting either sea level anomaly or ocean surface currents in the open ocean. The methodology has been applied to estimate the annual modulations of M_2, which are found to have much smaller amplitudes compared to those reported in HRET8.1, and suggest that the latter estimates of these tides were not reliable.

SIGNIFICANCE STATEMENT: The mechanical and thermodynamic forcing of the ocean occurs primarily at very large scales associated with the gravitational perturbations of the sun and moon (tides), atmospheric wind stress, and solar insolation, but the frictional forces within the ocean act on very small scales. This research addresses the question of how the large-scale tidal forcing is transformed into the smaller-scale motion capable of being influenced by friction. The results show where internal waves are generated, and how they transport energy across ocean basins to eventually be dissipated by friction. The results are useful to scientists interested in mapping the flows of mechanical energy in the ocean and predicting their influences on marine life, ocean temperature, and ocean currents.

1. Introduction

This paper is concerned with estimating and mapping the tidal harmonic constants associated with the sea level anomaly (SLA) of baroclinic tides in the open ocean. It develops an approach based on a model for the SLA consisting of spatial Fourier modes modulated in time by the astronomical gravitational tide-generating potential. It thus extends approaches from the literature in which different forms of a relatively simple kinematic wave model are used to map baroclinic tides (Ray and Cartwright 2001; Zhao et al. 2012; Dushaw 2015; Zaron 2019). This work is broadly motivated by the desire to improve baroclinic tide prediction and to understand the role of the tides in the dynamics of the ocean.

Carrere et al. (2021) compared different models for predicting the baroclinic tides. Broadly, such models could be classified as follows: (1) ad-hoc empirical models in which observed harmonics of the baroclinic tides are smoothly interpolated and extrapolated to yield continuous fields for prediction of the tides at arbitrary locations (Ray and Zaron 2016), (2) kinematic wave models which represent the observed harmonics as a superposition of idealized waves (Dushaw 2015; Zhao et al. 2016; Zaron 2019; Ubelmann et al. 2022), (3) dynamic wave models which solve for the baroclinic tides from the known astronomical tide-generating force (Shriver et al. 2014), and (4) data-assimilative models which assimilate observations into a dynamic wave model, using assumptions about the expected errors of the observations and the model (Egbert and Erofeeva 2014). In approach (2), the kinematic wave models, there are different criteria for identifying the constituent waves and describing their waveforms. In approach (4), the data-assimilative model, the attributes of the dynamical model errors must be quantified, rather than the attributes of the wave field per se, and there is considerable uncertainty about how the dynamical errors should be represented. One might anticipate that a data-assimilative model would be more accurate than a kinematic wave model, because the SLA obtained would unambiguously correspond to some dynamics, even if these differ from the ones specified a priori; however, Carrere et al. (2021) found that the most accurate tidal SLA predictions were obtained from a kinematic wave model, rather than a data-assimilative model. The working hypothesis of this paper is that the systematic pursuit of a descriptive kinematic model will provide insight into the causes of the surprising results of Carrere et al. (2021), and lead to a better understanding of baroclinic wave dynamics in the ocean.

The approach taken here is to extend previous works in two ways. First, different formulations of the estimators for the kinematic wave model are considered, including the stepwise least-squares approaches of Zhao et al. (2016) and the penalized least-squares of Dushaw (2015), leading to the preferred approach, a type of least-angle regression (Efron et al. 2004). And, second, different independent data sources are used for building the kinematic wave models and for cross-validating them. The independent data are exact-repeat mission altimetry, geodetic mission altimetry, and surface currents estimated with Lagrangian surface drifters. The altimeter data are used in two forms, both as along-track sea level anomaly and as along-track collinear sea level anomaly differences, i.e., sea surface slope. The different kinds of data provide alternate stopping criteria for the iterative estimation algorithms, and reveal the degree to which different criteria lead to different estimates. The different data types were collected over different time periods, they are subject to different sources of instrumental error, and they are subject to different degrees of contamination by non-tidal signals which interfere with the tidal estimation. There is evidence that the baroclinic tides are undergoing long-term secular changes (Zhao 2023), but the focus of the present work is on estimating the time-mean baroclinic tides, the so-called phase-locked or stationary tides, with the hope of returning to the question of long-term variability in the future using the tools developed herein.

The organization of this manuscript is as follows. Section 2 provides an overview of the family of estimators used for analyzing and mapping the diverse data sources mentioned above, and Section 3 describes these data sources in detail. In Section 4, the implementations of the estimators are

TABLE 1. Tides considered. Darwin symbols and corresponding Doodson numbers for the mapped tides are listed. The alias periods of the tides are are shown, in days, for the exact-repeat mission sampling along the Topex/Poseidon/Jason tracks (TX), the Geosat Follow-On tracks (G1), and the ERS-2/Envisat/Saral-AltiKa tracks (E2)

Darwin symbol	Doodson number	Alias periods TX/G1/E2 [day]	Data window L_d [km]
O_1	1 455 554	46/113/75	1000
K_1	1 655 556	173/175/365	1000
N_2	2 456 555	50/52/97	1000
MA_2	2 545 555	75/170/75	1000
M_2	2 555 555	62/318/94	500
MB_2	2 565 555	53/2459/127	1000
S_2	2 735 555	59/169/∞	1000

described and tested, leading to the selection of the formulation which is used subsequently in Section 5. Finally, in Sections 6 and 7, the results are discussed and summarized.

2. Mapping Methodology, Part I: Overview

a. A kinematic wave model

The tides are unique among ocean phenomena in that their temporal structure is well-defined by the known gravitational dynamics of the Sun, Earth, and Moon. The baroclinic tides are represented here by the SLA that is phase-locked with the astronomical gravitational tidal potential,

$$\eta(\theta,\phi,t) = \sum_j Re[\zeta_j(\theta,\phi) f_j(t) \exp(-i(\omega_j t + u_j(t)))], \tag{1}$$

where j indexes the partial tides which are here denoted by their Darwin symbols, e.g., M_2, with corresponding frequencies ω_j obtained from the Doodson numbers enumerated in Table 1; (θ,ϕ) are spherical-polar spatial coordinates; $\zeta_j(\theta,\phi)$ is the complex-valued field giving the spatial structure of the j-th partial tide; and $f_j(t)$ and $u_j(t)$ are given functions which account for the

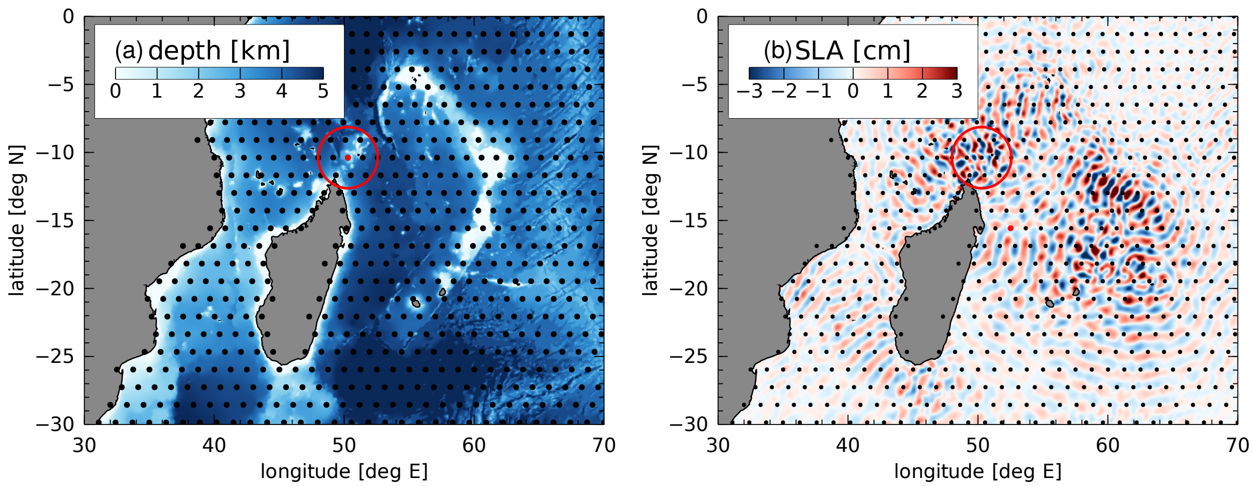


FIG. 1. Centers of tangent-planes used for mapping M_2 are indicated with black dots overlaid on (a) water depth and (b) the M_2 baroclinic sea level anomaly in the region around Madagascar in the West Indian Ocean. The red dot and red circle indicate the center of a tangent-plane and the data disk of a representative D_m. The SLA shown in panel (b) is the E^* estimate discussed later in the text.

modulations associated with the 18.6-yr precession of the node of the lunar orbit (Foreman et al. 2009).

The spatial structure of the tides, $\zeta_j(\mathbf{x})$, is represented with a kinematic wave model (as distinct from a dynamical wave model) comprised of a linear combination of idealized propagating waveforms. For computational considerations, the properties of the waves are assumed to be independent within a patchwork of locally-defined tangent-planes, D_m, overlapping around the globe. The centers of the tangent planes used for mapping M_2 are shown in Figure 1. The spacing between D_m centers is approximately one-fourth the radius of a data disk, L_d, at the center of each tangent-plane, to be described later.

Let $\zeta_{jm}(\mathbf{x})$ be the estimate of ζ_j obtained for $\mathbf{x} \in D_m$. Within D_m, the spatial structure of ζ_{jm} is assumed to be,

$$\zeta_{jm}(\mathbf{x}) = \sum_{p,q=0}^{p+q=P} \sum_{\mathbf{k}} a_{jmpq\mathbf{k}} x^p y^q \exp(i\mathbf{k}\cdot\mathbf{x}), \tag{2}$$

which can be regarded as a sum of propagating plane-waves with vector wavenumbers, $\mathbf{k}$, modulated by a polynomial amplitude envelope, $x^p y^q$, for $0 \le p+q \le P$. Note that the order-P polynomial amplitude envelope provides $N_p = (P+1)(P+2)/2$ polynomial coefficients per wavenumber component, and a quadratic envelope is used here ($P = 2$; $N_p = 6$). Within each tangent-plane, local Cartesion coordinates are used, $\mathbf{x} = (x, y)$, relative to the center (θ_m, ϕ_m), and the wavenumbers $\mathbf{k} = (k, l)$ are taken as the set of discrete Fourier wavenumbers at $\Delta x = 6$ km spatial resolution, the approximate spatial resolution of the along-track altimetry data. Note that the coefficients, $a_{jm00\mathbf{k}}$, are the discrete Fourier transform coefficients of the local plane-wave representation of ζ_{jm}; the other coefficients, $a_{jmpq\mathbf{k}}$ for $p, q \neq 0$, represent non-plane-wave features of ζ_{jm}.

Let $\hat{\mathbf{a}}_{jmpq} = \{a_{jmpq\mathbf{k}}\}$ denote the collection of coefficients for all the discrete Fourier wavenumbers, $\mathbf{k}$, and let F and $F^{\dagger}$ denote the unitary discrete Fourier transform pair. A compact representation of equation (2) is given by,

$$\zeta_{jm}(\mathbf{x}) = \sum_{p,q=0}^{p+q=P} x^p y^q F^{\dagger} \hat{\mathbf{a}}_{jmpq}, \tag{3}$$

which is the core of the computationally-efficient implementation of the estimators presented below. For notational convenience, equation (3) shall be written as a linear system,

$$\zeta_{jm} = \mathbf{F}\mathbf{a}_{jm}, \tag{4}$$

where the vector $\mathbf{a}_{jm} = \{\hat{\mathbf{a}}_{jmpq}\}$ collects all the unknown coefficients, and $\mathbf{F}$ is a linear operator assembled from the modified Fourier operators, $x^p y^q F^{\dagger}$. The vector, $\mathbf{a}_{jm}$, will be referred to as the vector of generalized Fourier coefficients, since each element of the vector is the coefficient of a plane-wave component with a discrete Fourier wavenumber, multiplied by $x^p y^q$. In equation (4), ζ_{jm} is a vector of gridded harmonic constants for tide-j within the Cartesian tangent-plane D_m corresponding to the function, $\zeta_{jm}(\mathbf{x})$, in equation (2). To be explicit, on a square domain containing $M \times M$ grid-points, the dimension of the ζ_{jm} vector is $M^2 \times 1$, the dimension of the $\mathbf{F}$ matrix is $M^2 \times N_p M^2$, and the dimension of the $\mathbf{a}_{jm}$ vector is $N_p M^2 \times 1$.

b. The local Cartesian planes and their blending

Estimates of ζ_{jm} are found using data within a circular patch of radius L_d contained within each D_m. Each D_m is a square with side length, $L = \sqrt{2} L_d$, which provides a data-free region around the data patch. The domains, D_m, are overlapping and staggered so that the centers are offset by approximately $L_d/2$. The data-free region prevents the periodic boundary condition of the Fourier basis from unduly influencing ζ_{jm} estimates within the data disk, while the overlapping tangent planes insure that at least two independent estimates of the harmonic constants are made at any location, except near coastlines. The radius of the data window, L_d, will be specified, below.

A continuously-differentiable representation of $\zeta_j(\mathbf{x})$ is computed as a weighted average,

$$\zeta_j(\mathbf{x}) = N_j^{-1}(\mathbf{x}) \int_D \sum_m K(||\mathbf{x}-\mathbf{x}_m||/L_d)\zeta_{jm}(\mathbf{x})d\mathbf{x}, \tag{5}$$

where the normalization factor is given by,

$$N_j(\mathbf{x}) = \int_D \sum_m K(||\mathbf{x}-\mathbf{x}_m||/L_d)d\mathbf{x}, \tag{6}$$

for the global domain, D. The averaging kernel, $K(r)$, is a radial basis function with compact support (Wendland 1995),

$$K(r) = (1-r)^3(3r+1), \tag{7}$$

for $0 \le r \le 1$, and zero otherwise. Explicitly, the distance function is defined by $||\mathbf{x}-\mathbf{x}_m||^2 = [(\theta-\theta_m)^2 + ((\phi-\phi_m)\cos\theta_m))^2]r_e^2$, where r_e is Earth's radius.

c. A dynamical relationship between surface velocity and η

One goal of the present work is to estimate ocean surface currents due to baroclinic tides. Here, currents are estimated with the horizontal momentum equation,

$$-i\omega_j\mathbf{u}_j + f\hat{k}\times\mathbf{u}_j = -g\nabla\eta_j, \tag{8}$$

where $\mathbf{u}_j = (u_j, v_j)$ is the horizontal velocity vector associated with the j-th partial tide, f is the Coriolis frequency, and g is gravitational acceleration. This equation is an approximation which neglects nonlinear advection and turbulent stresses. In keeping with the kinematic nature of the wave model for η_j, these approximations are accepted as part of the descriptive nature of the model, and they can be partly justified a posteriori. Note that the $\mathbf{u}_j$ and f in equation (8) should not be confused with the nodal factors, u_j and f_j, appearing in equation (1), above.

When $\omega_j = f$, equation (8) is singular and cannot be inverted for $\mathbf{u}_j$. Initially, a Rayleigh damping term was added to regularize the inversion, but experimentation revealed that no damping was necessary so long as $|\omega_j - f|$ is larger than machine precision at the gridpoints of the local tangent plane. In principle, a physical model for damping could be justified (Savva and Vanneste 2018; Kafiabad et al. 2019; Dong et al. 2020; Kelly et al. 2021); however, at this stage it is unclear if the data are sufficient to distinguish among plausible alternative models.

d. Estimators for $\mathbf{a}_{jm}$

As described above, the baroclinic tidal sea level anomaly is represented in terms of generalized Fourier coefficients, $\mathbf{a}_{jm}$. This section introduces a family of estimators for $\mathbf{a}_{jm}$ capable of incorporating both observed data (measurements of SLA and surface velocity) and constraints on allowable wavenumbers, $\mathbf{k}$, inferred from the dispersion relation. Recall that the j and m subscripts on $\mathbf{a}_{jm}$ refer to the tidal frequency (ω_j) and the local tangent plane (D_m). For convenience in this section, let $\mathbf{a}_m = \{\mathbf{a}_{jm}\}$ denote a single vector of generalized Fourier coefficients for all the partial tides; furthermore, subscript m shall be omitted since the generalized Fourier coefficients will be estimated independently within each tangent plane.

One family of estimators for $\mathbf{a}$ are minimizers of an objective function of the form,

$$\mathcal{J}(\mathbf{a};\lambda,\alpha,\mathbf{B},\mathbf{C}) = \lambda||\mathbf{B}^{-1/2}\mathbf{a}||_\alpha + (\mathbf{d}-\mathbf{HFa})^T\mathbf{C}^{-1}(\mathbf{d}-\mathbf{HFa}), \tag{9}$$

where λ is a scalar regularization parameter; $\alpha \in \{0,1,2\}$ is a scalar which determines the norm used in the regularization; $\mathbf{B}^{1/2}$ is the matrix square-root of $\mathbf{B}$, a positive-semidefinite weighting matrix (commonly identified with the background error covariance of $\mathbf{a}$ for the case $\alpha = 2$); $\mathbf{C}$ is a positive-definite weighting matrix, referred to as the observation error covariance matrix; $\mathbf{d}$ is a vector of observational data; and $\mathbf{H}$ is a linear observation operator which samples the fields of harmonic constants ($\zeta = \mathbf{Fa}$) to produce model-based estimates of the observations. Note that $\mathbf{H}$ may involve spatial interpolation (corresponding to observations of along-track harmonic constants), interpolation and harmonic synthesis in time (corresponding to observations of along-track SLA), interpolation, harmonic synthesis, and spatial gradients (corresponding to observations of ocean surface currents), or other linear operations (such as along-track differencing). The definitions of the norm $||\mathbf{a}||_\alpha$ are as follows: $||\mathbf{a}||_0$ is the number of non-zero elements of $\mathbf{a}$, $||\mathbf{a}||_1 = \sum_i |a_i|$, and

TABLE 2. Exact-repeat orbit satellite altimeter missions used. Abbreviations follow the usage in the Radar Altimeter Database System (Scharroo et al. 2013).

Satellite mission	Time period	Orbit cycles
(TOPEX/Jason reference orbit, Δt = 9.9156d)		
TXA	1992–2002	4–364
J1A	2002–2009	1–259
J2A	2008–2015	1–303
J3A	2016–2021	1–216
(TOPEX/Jason interleaved orbit, Δt = 9.9156d)		
TXB	2002–2005	369–480
J1B	2009–2012	262–374
J2B	2016–2017	305–327
(Geosat orbit, Δt = 17.0505d)		
G1A	2000–2008	37–223
(ERS/Envisat reference orbit, Δt = 35.0000d)		
E2A	1995–2003	1–83
N1B	2002–2010	10–94
SAA	2013–2017	1–34

$||\mathbf{a}||_2 = \mathbf{a}^T\mathbf{a}$, where superscript T indicates transpose and i ranges over the elements of $\mathbf{a}$, taken as a real-valued vector.

The general formulation of (9) is used here because it is capable of representing many of the previous approaches to estimating the baroclinic tides. One of the key distinctions among methods is the choice of model complexity, specifically, the number of component plane waves used in stepwise regression (Zhao et al. 2014), the order of the polynomial model for the amplitude modulation (Zaron 2019), and the bandwidth of allowable waves around the predicted dispersion relation (Dushaw 2015). In (9), the gamut of component waves is determined by the non-zero entries of $\mathbf{B}$. The model complexity (i.e., the number of waves with significant or non-zero coefficients) can be controlled by the scalar λ, a tradeoff parameter, and α, the type of norm used.

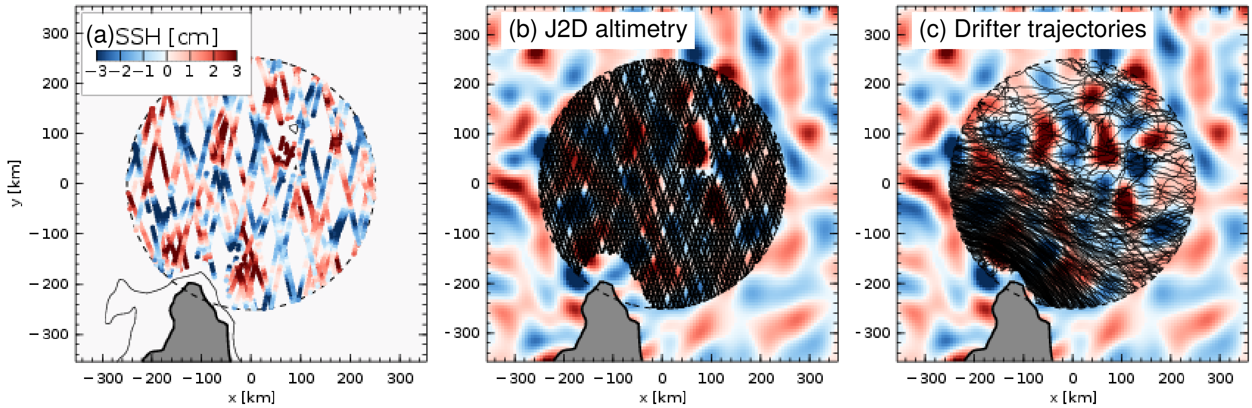


FIG. 2. Data locations within a representative data patch northeast of Madagascar (indicated with the red circle in Figure 1). (a) M_2 harmonic constants from ERM altimetry, (b) GM altimeter data from the J2D mission, which represents only 5% of GM data within this patch, and (c) drifter trajectories. The contour in panel (a) is the 500m isobath; the scale image in panels (b) and (c) is the optimal estimate of the M_2 tide using the same colorscale as panel (a); the dashed line is the boundary of the data patch, with radius $L_d = 250$km from the center of the tangent-plane.

3. Data

Data from three different sources are used to build and evaluate the family of estimators proposed in the previous section. The purpose of using different data sources is to conduct robust cross-validation studies, wherein the model estimated from one source is evaluated by comparison with another source.

a. SLA Harmonic Constants from Exact-Repeat Orbit Altimetry

The exact-repeat mission (ERM) altimetry data used here include those used in Zaron (2019), updated through 2020-12-31. The data were pre-processed and harmonic constants computed as described in Zaron (2019) with one change: prior to harmonic analysis, the DT-2021 version of the SSALTO/DUACS gridded mesoscale SLA is subtracted from the altimeter SLA, rather than the older version of the SSALTO/DUACS product (Zaron and Ray 2018) used previously.

Table 2 summarizes the ERM altimeter missions used, and Figure 2a illustrates these data within the D_m indicated in red in Figure 1. The data are typical of observations close to complex

TABLE 3. Long-repeat- and non-repeat-orbit satellite altimeter missions used. Abbreviations follow the Radar Altimeter Database System (Scharroo et al. 2013).

Satellite mission	Time period	Orbit cycles
Cryosat-2/phase A (C2A)	2010–2020	7–136
Jason-1/phase C (J1C)	2012–2013	382–425
Jason-2/phase C (J2C)	2017–2018	332–355
Jason-2/phase D (J2D)	2018–2019	356–383
Saral-AltiKa/phase B (SAB)	2016–2021	36–92

topography and exhibit amplitudes of several centimeters, correlated at nearby mission ground tracks, but without a clear large-scale spatial structure.

In order to reduce the effect of residual long-wavelength errors, the complex harmonic constants are differenced in the along-track direction. This approach is different from the along-track filtering employed in other works (e.g., Dushaw 2015; Zhao et al. 2016) since it is used within the least-squares data-fitting, rather than being done as an independent data processing step. The original harmonic constants (along-track non-differenced) are used in some of the cross-validation metrics, below, but along-track-differenced data are always used to build the wave models.

Confidence intervals for the harmonic constants are estimated using a Monte Carlo method to generate realizations of colored noise from the de-tided residuals (Matte et al. 2013). These confidence intervals are averaged along each ground track within the data analysis windows and used to identify and exclude outliers.

b. SLA Time Series from Long-Repeat- and Non-Repeat-Orbit Altimetry

The primary source of data for cross-validation are SLA measurements from long-repeat- and non-repeat-orbit altimeter missions, listed in Table 3. Together, these datasets will be referred to as geodetic mission (GM) altimetry.

In contrast to the ERM ground tracks, the GM ground tracks are closely spaced and are therefore useful for assessing models in the spatial gaps not sampled by other data. Figure 2b shows the locations of measurements from the J2D mission. The spatial density of the GM data is remarkable, considering that the J2D data shown is only about 5% of the total available. Quantitatively, the GM data are dominated by C2A, which accounts for about 50% of the total.

As with the ERM data, discussed above, the GM data are used in the form of both SLA and along-track SLA differences. We have experimented with the using the GM data in both forms to estimate the baroclinic tides, but the models obtained have lower prediction skill than those based on ERM data. Hence, the ERM data are used here for estimating the tides, while the GM data are withheld and used for cross-validation of the ERM-based estimates.

c. Velocity Time Series from Lagrangian Surface Drifters

Several authors have noted the presence of baroclinic tidal signals in velocity time series obtained from Lagrangian surface drifters (Elipot and Lumpkin 2008; Poulain and Centurioni 2015; Kodaira et al. 2016; Elipot et al. 2016; Zaron and Ray 2017; Zaron and Elipot 2021). The hourly drifter-derived surface-currents are used here to explore whether the baroclinic tides estimated from them are consistent with those estimated from altimetry. The Global Drifter Program (GDP) dataset, version 2.0, is used (Elipot et al. 2022b,a), which consists of approximately 165 million hourly estimates of position and velocity obtained from 17,234 individual drifters, collected between 1990 and 2021, although 86% of the data were collected after 2005. Data from both drogued and undrogued drifters are used since there appears to little shear near the ocean surface at the diurnal and semidiurnal tidal frequencies (Arbic et al. 2022).

Similar to the altimetry, it is useful to remove an estimate of the low-frequency signals prior to using the drifter-derived currents for tidal mapping. For this purpose, the original currents were filtered around the dominant M_2 and K_1 tidal frequencies with a bandpass filter of width 0.4 cycles-per-day (cpd) and 0.2 cpd. Without this filtering, significant mapping artifacts appeared in regions of strong currents with sparse data, such as within the equatorial current systems of the Pacific and Indian Oceans.

Figure 2c shows the drifter tracks in the D_m near Madagascar, and the tracks are indeed dense compared to the 50-to-120 km scales baroclinic waves over most of the patch. However, the data density is notably sparser in the northen half of the domain, compared to the southern half. Compared to the ERM and GM datasets described above, the GDP are heterogeneous over the globe.

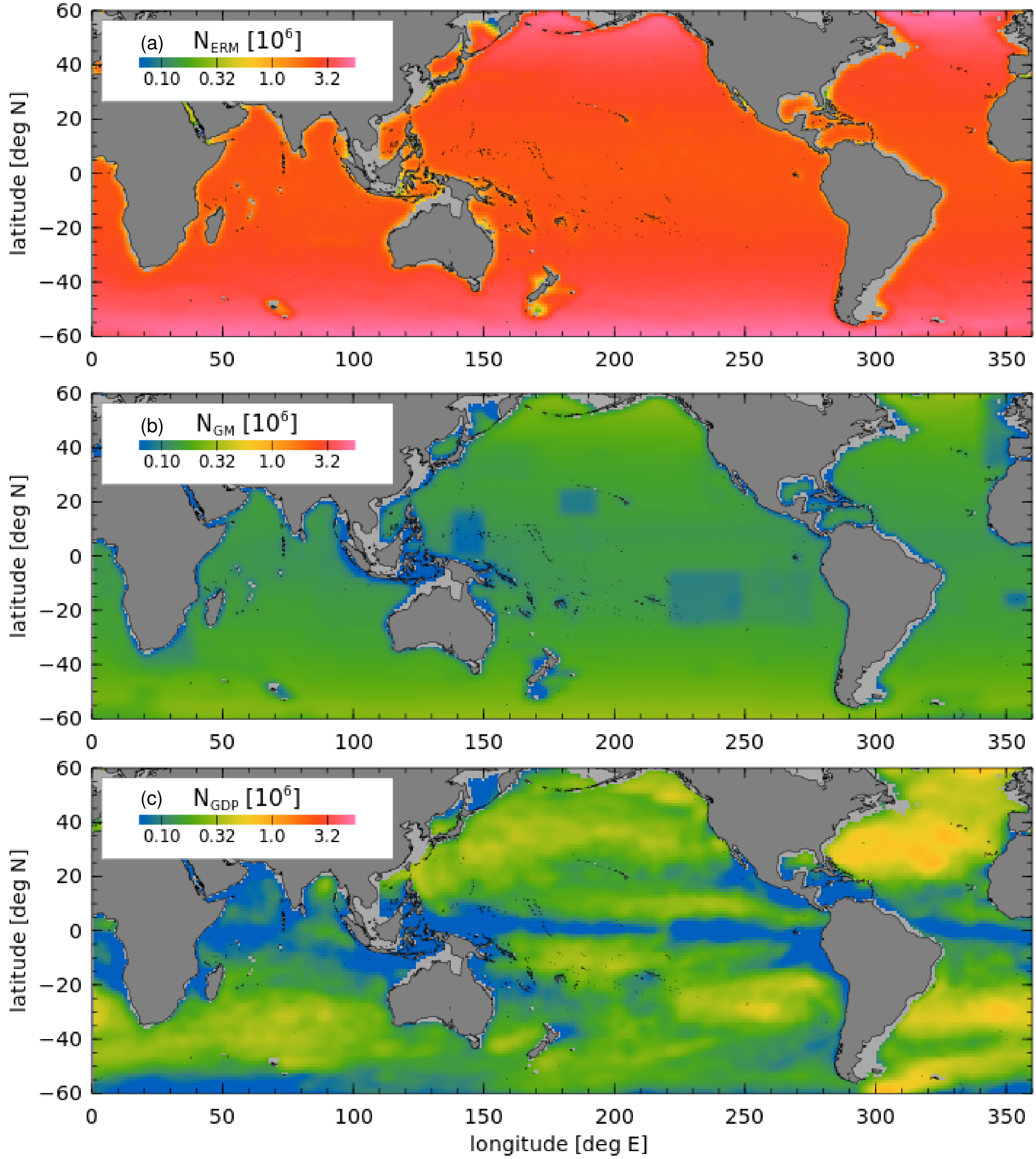


FIG. 3. Number of observations of different types. Note that the colorscale is non-uniform (log-scaled).

268 *d. Summary*

269 Figure 3 maps the number of observations within the D_m domains for M_2, discussed in the Results
270 section, below. The ERM data comprise by far the most numerous and homogeneously-distributed

form of data; although, the ERM data fall exclusively along a relatively sparse set of ground tracks (e.g., there are about 6000 data locations shown in Figure 2a). The sustained observations along the TP/Jason reference orbit (29-years) and ERS/Envisat orbit (20-years) comprise the bulk of the ERM data. The GM dataset is, overall, homogeneous, with the exception of regions where the Cryosat-2 sampling mode has been changed over the years (e.g., near 20°N, 180°E); it comprises about 19-years of satellite data. In contrast, the GDP data are not evenly distributed, as a consequence of both the distribution of deployment locations and the Lagrangian character of the surface trajectories.

Preliminary experiments sought to combine all the data types within each D_m, but the results (not shown) were very sensitive to the relative weighting of the observations. In principle, one would expect the optimal weighting to be equal to the inverse of the error covariance of the observations; however, the error covariance appears to be dominated by the presence of non-tidal signals, rather than instrumental errors, and developing consistent estimates for the covariance proved to be too challenging. Instead, the preliminary experiments motivated the approach used below, where the data sources are used one-at-a-time, and the λ parameter is optimized through cross-validation with the other datasets.

4. Mapping Methodology, Part II: Implementation

a. Choosing α, λ, $\mathbf{B}$, and $\mathbf{C}$

What values of the parameters, α, λ, $\mathbf{B}$, and $\mathbf{C}$, lead to the best estimates of $\mathbf{a}$? This is a big design space which is only partially explored here. Optimal values of these parameters might depend on geographic location, D_m, the size of the data window, L_d, and the data used for mapping, $\mathbf{d}$. The criteria for judging whether $\mathbf{a}$ is the "best estimate" is also important. Among the design choices investigated in this section, ERM altimeter data comprise $\mathbf{d}$, and a comparison dataset of GDP data are used to assess the parameter choices.

Previous work has assumed either that the component waves are exactly described by the dispersion relation for mode-n linear internal waves (e.g., Zhao et al. 2014),

$$|\mathbf{k}_n|^2 = \frac{\omega_j^2 - f^2}{c_n^2}, \tag{10}$$

where f is the latitude-dependent Coriolis frequency, and c_n is the non-rotating eigenspeed of mode-n waves (Kelly 2016), or that the waves fall within a prescribed bandwidth around the dispersion relation (Dushaw 2015; Zaron 2019). There are several rationales for allowing wavenumbers which fall off the dispersion relation: (1) the waves are locally-defined with respect to particular values of f and c_n which vary spatially within the data windows, (2) there may be a directly-forced component of the tidal SLA which does not fall on the dispersion relation, and (3) time-variability of the propagation medium can lead to a distribution of waves around the mean dispersion relation. To allow for wavenumbers that fall off the dispersion relation, the nonzero elements of $\mathbf{B}$ are required to correspond to wavenumbers within a prescribed fractional bandwidth, ν, such that $(1-\nu)|\mathbf{k}_n| \leq |\mathbf{k}| \leq (1-\nu)^{-1}|\mathbf{k}_n|$, following Dushaw (2015). Recall, following equation (9), that for $\alpha = 2$ the matrix $\mathbf{B}$ may be regarded as the background error covariance of the generalized Fourier coefficient vector, $\mathbf{a}$; hence, more generally, the diagonal of $\mathbf{B}^{1/2}$ is taken as an a priori estimate of the magnitude of the elements of $\mathbf{a}$. Here, $\mathbf{B}$ is assumed to be diagonal, and the wavenumber components are a priori uncorrelated. Experiments were conducted (not shown) in which bandwidth parameters of $\nu \in \{0.45, 0.30, 0.23, 0.11\}$ were used to estimate modes $n = 1, \ldots, 3$. Through trial and error, it was found that the 0.23 and 0.11 bandwidths were best for modes 1 and 2, respectively, values used to generate the results reported, below. Experiments found that mode-3 could not be stably estimated from ERM altimetry.

The matrix $\mathbf{C}$ is taken as the identity, which is equivalent to assuming the data errors are uncorrelated and homogeneous within the domain D_m when equation (9) is regarded as a Bayesian estimator. As already mentioned, the assignment of different error levels to the different data types did not lead to notably better estimators, and, likewise, the weighting of data from different ERM missions did not improve the results either.

The choices for α and λ are related. For $\alpha = 0$, λ may be varied to yield sparse least-squares estimates of $\mathbf{a}$. This type of estimator is closely related to the stepwise least-squares algorithms of Zhao et al. (2014) and Zaron (2019). In contrast, the choice $\alpha = 2$ is the prototype of estimators used in variational data assimilation where the full gamut of possible waves are used, and λ controls the goodness-of-fit to the data and provides protection against overfitting (Dushaw 2015). The new approach used here, $\alpha = 1$, combines the sparsity-favoring properties of the $\alpha = 0$ estimator with control on overfitting similar to the $\alpha = 2$ estimator.

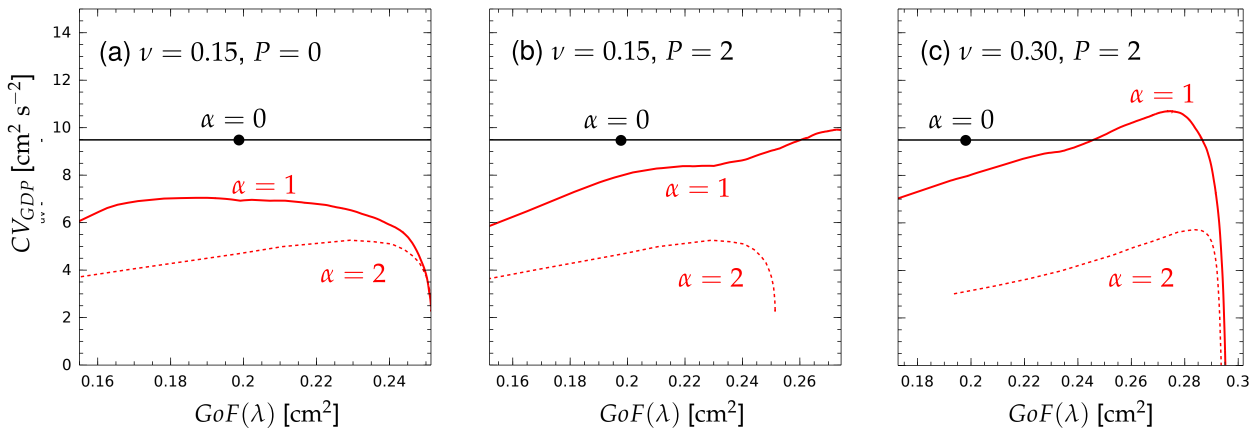


FIG. 4. An example of the performance of three different forms of the estimator (9) for $\alpha = 0, 1, 2$, for mapping the M_2 baroclinic tide from harmonically analyzed ERM altimetry near the Hawaiian Ridge. The influence of λ is represented by the goodness-of-fit metric on the x-axis, $GoF(\lambda) = (|\mathbf{d}_{ERM}|^2 - |\mathbf{d}_{ERM} - \mathbf{H}_{ERM}\mathbf{Fa}|^2)/N_{ERM}$. The performance of the estimators is shown by the cross-validation metric on the y-axis, $CV_{GDP} = (|\mathbf{d}_{GDP}|^2 - |\mathbf{d}_{GDP} - \mathbf{H}_{GDP}\mathbf{Fa}|^2)/N_{GDP}$. The line for the $\alpha = 0$ case depicts the performance of the HRET8.1 model (Zaron 2019); it is shown as a line for comparison of CV_{GDP} with the other estimators, but it represents a single estimate of $\mathbf{a}$ with GoF indicated by the solid dot. Performance for three cases: (a) a narrow-bandwidth B, $\nu = 0.15$, using the simplest plane-wave basis, $P = 0$, (b) a narrow-bandwidth, $\nu = 0.15$, but with a quadratically-modulated plane-wave basis, $P = 2$, (c) a wide-bandwidth, $\nu = 0.3$, and a quadratically-modulated plane-wave basis, $P = 2$.

Figure 4 illustrates the influence of different choices of α, λ, and ν on the performance of the estimator (9). The x-axis represents variations in λ as measured by a goodess-of-fit (GoF) metric, $(||\mathbf{d}_{ERM}||^2 - ||\mathbf{d}_{ERM} - \mathbf{H}_{ERM}\mathbf{Fa}||^2)/N_{ERM}$, which measures the explained variance of the estimate with respect to the data used, $\mathbf{d}_{ERM}$. In this example, $\mathbf{d}_{ERM}$ consists of the same ERM data as used in the HRET8.1 model (Zaron 2019). On the y-axis, the performance of the estimator is measured using cross-validation (CV_{GDP}) with respect to an independent dataset, $\mathbf{d}_{GDP}$, by the explained variance $(||\mathbf{d}_{GDP}||^2 - ||\mathbf{d}_{GDP} - \mathbf{H}_{GDP}\mathbf{Fa}||^2)/N_{GDP}$, where $\mathbf{d}_{GDP}$ represents GDP surface current observations, and $\mathbf{H}_{GDP}$ is the measurement operator corresponding to $\mathbf{d}_{GDP}$, an $N_{GDP} \times 1$ vector.

The $\alpha = 0$ estimator is taken from HRET8.1, which utilizes exactly 6 quadratically-modulated ($P = 2$) component waves; this estimate is the reference against which the others are evaluated, and it is identical in each panel (indicated with the dot labelled $\alpha = 0$ on the line of constant CV_{GDP}). The $\alpha = 1$ and $\alpha = 2$ estimates are parameterized by λ within each panel, for different choices of $\mathbf{B}$ (bandwidth) and P (polynomial modulation). The three different cases as are as follows: (a) A narrow bandwidth, $\nu = 0.15$, is used, and no polynomial amplitude modulation is used, $P = 0$. In this case, the $\alpha = 0$ estimate performs better than either of the $\alpha = 1$ or $\alpha = 2$ alternatives. The $\alpha = 1$ estimate consistently explains more variance than the $\alpha = 2$ estimate, the conventional L_2 estimator, and it is more efficient, in the sense that it achieves its maximum CV_{GDP} statistic for a smaller value of the GoF statistic. (b) A narrow bandwidth, $\nu = 0.15$, is used, and a quadratic amplitude modulation is used, $P = 2$. In this case, the $\alpha = 1$ estimate exceeds the performance of both the $\alpha = 0$ and $\alpha = 2$ estimates. (c) A wider bandwidth, $\nu = 0.30$, is used, and a quadratic amplitude modulation is used, $P = 2$. The added bandwidth allows for further improvements in the CV_{GDP} metric and conclusively shows that the $\alpha = 0$ and $\alpha = 2$ solutions are sub-optimal.

Reasons for the differences among the estimators are not completely understood. If (9) is interpreted in a Bayesian context, then the better performance of the $\alpha = 1$ estimator compared to the $\alpha = 2$ estimator may be related to the non-Gaussian distribution of the component wave coefficients. Whereas both the $\alpha = 0$ and $\alpha = 1$ estimators are inherently sparsity-selecting (Candès et al. 2006), only the $\alpha = 1$ estimator allows for controlling the goodness-of-fit (avoiding overfitting) among the non-zero elements of $\mathbf{a}$.

This study was partly motivated by the desire to implement a non-arbitrary stopping criterion in the estimator of Zaron (2019). Experiments using stopping criteria based on the F-test, Mallows C_p (Mallows 1973), and Aikake's Information Criterion (Akaike 1974) yielded sub-optimal results when evaluated using cross-validation. For a modern review of these, and other, approaches to stopping criteria and their sensitivity to unmodeled signals or non-Gaussian noise, the dissertation by Lysen (2009) is recommended. For these reasons, the estimators used below exclusively use a cross-validation metric (i.e., comparison against withheld or independent data) to determine the optimal value of the regularization parameter, λ.

b. Numerical methods and implementation details

The generalized Fourier coefficient vector, $\mathbf{a}$, consists of the coefficients of the generalized Fourier components on D_m. Many of these coefficients are zero, though, due to the a priori bandwidth selection, as defined by $\mathbf{B}$, which is taken as a diagonal matrix consisting either of 0's or 1's. Let the vector $\mathbf{b}$ denote the elements of $\mathbf{a}$ corresponding to the gamut of possible generalized wavenumbers. Let $\mathbf{b}$ be such that $\mathbf{a} = \mathbf{B}_b^{1/2}\mathbf{b}$, where $\mathbf{B}_b^{1/2}$ is the rectangular matrix containing only those of columns of $\mathbf{B}^{1/2}$ with non-zero elements. With this transformation, equation (9) fits the canonical form,

$$\mathcal{J}(\mathbf{b};\lambda,\alpha,\mathbf{B},\mathbf{C}) = \lambda||\mathbf{b}||_\alpha + (\mathbf{d}-\mathbf{A}\mathbf{b})^T\mathbf{C}^{-1}(\mathbf{d}-\mathbf{A}\mathbf{b}), \tag{11}$$

where $\mathbf{A} = \mathbf{HFB}_b^{1/2}$.

Different methods were used to compute the minimizer of (11), depending on the value of α. For the case, $\alpha = 2$, given λ, equation (11) is minimized using the conjugate gradient algorithm (Shanno 1985). Alternately, when the gamut of non-zero wavenumbers is small enough, equation (11) may be minimized using direct methods by solving,

$$(\lambda\mathbf{I} + \mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})\mathbf{b} = \mathbf{A}^T\mathbf{C}^{-1}\mathbf{d}, \tag{12}$$

which was used as a check on the conjugate gradient solver.

For the case $\alpha = 1$, a generalization of the least-angle regression algorithm is used (Efron et al. 2004). In its original form, this elegant algorithm produces a sequence of scalar regularization parameters $\lambda = \lambda_k$ and vectors $b = \mathbf{b}_k$ which minimize $\mathcal{J}$, where $k = 1,\ldots,N_b$, with N_b being the dimension of the vector $\mathbf{b}$. At each step, the rank of the linear system to be solved is k and corresponds to the k-dimensional subspace of $\mathbf{A}$ which is most correlated with the residual from the previous step, $\mathbf{r}_{k-1} = \mathbf{d} - \mathbf{A}\mathbf{b}_{k-1}$. In this way, the algorithm proceeds to build the minimizer of $\mathcal{J}$ which has the sparsity property that $||\mathbf{b}||_0 = k$ and which would terminate after $k = N_b$ steps at the least-squares solution of $\mathbf{A}\mathbf{b} = \mathbf{d}$ (when $\lambda = 0$). The generalization of the algorithm used here is one which expands the subspace at each step with pairs of variables corresponding to the real and imaginary parts of the generalized Fourier coefficients (Yuan and Lin 2006). The algorithm is computationally and memory efficient since it is based on forming rank-2 updates of the QR-decomposition of $\mathbf{A}$. In practice it is not run to completion at step $k = N_b$; instead, the algorithm is terminated when the optimal agreement with the cross-validation dataset is obtained.

The $\alpha = 0$ case for minimizing (11) proceeds similar to the $\alpha = 1$ case, except that the least-squares solution of $\mathbf{A}\mathbf{b}_k = \mathbf{d}$ is computed at each step, where the nonzero entries of $\mathbf{b}_k$ comprise the same set of variables as used in the $\alpha = 1$ solver (described above). By construction, the solution sequence achieves a smaller residual (better goodness-of-fit) compared to the corresponding $\alpha = 1$ case; however, in practice, the cross-validation metric is worse than the $\alpha = 1$ case (cf., Figure 4).

TABLE 4. Estimators. Seven estimators for each of the tides are computed, and these are labelled according to the dataset employed for computing generalized Fourier coefficients, and the dataset employed for cross-validation (optimization of λ). The notation Eab is used, where a indicates the data used to build the model, and b indicates the data used for cross-validation (1=ERM, 2=GM, 3=GDP). The estimate E^* is the linear combination of the E12 and E32 estimates which optimizes the explained variance with respect to the GM data (Appendix A). (Note that the E21 and E23 estimates are not shown, but are available from the authors.)

Estimate	$\mathbf{d}$	$\mathbf{d}_{CV}$	description
E12	ERM	GM	fit to ERM, optimize w.r.t. GM
E13	ERM	GDP	fit to ERM, optimize w.r.t. GDP
E21	GM	ERM	fit to GM, optimize w.r.t. ERM
E23	GM	GDP	fit to GM, optimize w.r.t. GDP
E31	GDP	ERM	fit to GDP, optimize w.r.t. ERM
E32	GDP	GM	fit to GDP, optimize w.r.t. GM
E*	ERM, GDP	GM	optimal linear combination of E13 and E31 w.r.t. GM

5. Results

The above-described methodology has been applied to estimate the baroclinic SLA and surface currents associated with the tides listed in Table 1. In every case the $\alpha = 1$ estimator with $P = 2$ (quadratic) amplitude modulation is used. A data window (disk radius) of $L_d = 250$ km is used for the M_2 tide, while a $L_d = 500$ km data window is used for the other tides, justified as follows: (1) for the diurnal tides, K_1 and O_1, because of their longer wavelengths, (2) for S_2, because of the sparse ground track sampling by ERM altimetry, and (3) for N_2, MA_2, and MB_2, because of their small signal-to-noise ratio in the data.

The three data sources described in Section 3 are used in pairs to create different estimates, as enumerated in Table 4. Rather than fitting a wave model to all the data sources simultaneously, each data source is used, in turn, and the λ parameter (which controls the goodness-of-fit to the data) is selected to optimize the estimate with respect to another of the datasets. For example, E12 involves building the wave model from ERM data, but using the GM data to select the optimal λ. In other words, the E12 estimate minimizes the prediction error with respect to the GM data, while fitting the ERM data. In addition to the pairwise fit-and-predict pairs, an estimate E^* is computed which is the optimal linear combination of the E12 and E32 solutions with respect to the GM data (see Appendix A).

Two aspects of this work are novel. First, the estimator defined by equation (9) involves the absolute value (the first term on the right-hand-side; $\alpha = 1$, an L_1-norm) as well as the sum-of-squares (the second term on the right-hand-side; an L_2-norm), so it is a mixed L_1/L_2-norm estimator. And, second, the use of different data types for building and cross-validating the models is another novel aspect of this work. While the GM data have been used previously for evaluating and comparing models (Zaron and Ray 2017; Carrere et al. 2021), it is not clear if these are the best data for this task, considering how non-tidal signals and various sources of noise overlap with the tidal alias frequencies (e.g., see Zaron 2018 for a detailed analysis of aliasing for the Cryosat-2 mission). Experiments were conducted to compare SLA versus along-track SLA differences for cross-validation, which revealed minor differences between the estimates (examples shown, below). Ultimately, along-track SLA differences were favored for cross-validation under the presumption that the errors are less correlated in the along-track direction than in the original SLA data.

a. M_2 tide

Figure 5 compares the M_2 baroclinic sea level anomaly estimate from Zaron (2019), HRET8.1, with the new estimates based on ERM data and GDP data, E12 and E32, respectively. Perhaps the most noteworthy aspect of the estimates is the remarkable visual similitude of the E12 and E32 estimates, which are based on completely independent data, and the similitude of these estimates with HRET8.1. Careful study of the panels does reveal subtle differences, though. For example, the HRET8.1 solution was masked to zero in much of the Southern Ocean and in the western boundary currents, while the E12 and E32 estimates smoothly go to zero in these regions. Also, the background noise in the E12 and E32 estimates appears smaller than in the HRET8.1 solution. While this is most visually evident in the southern Indian Ocean, explained variance metrics, shown below, confirm that the noise is in fact smaller throughout most of the global oceans. The same internal wave sources and large-scale beams are apparent in all three estimates. One difference among the estimates is the low-amplitude of the E32 solution in the Banda Sea (5°S,130°E) and

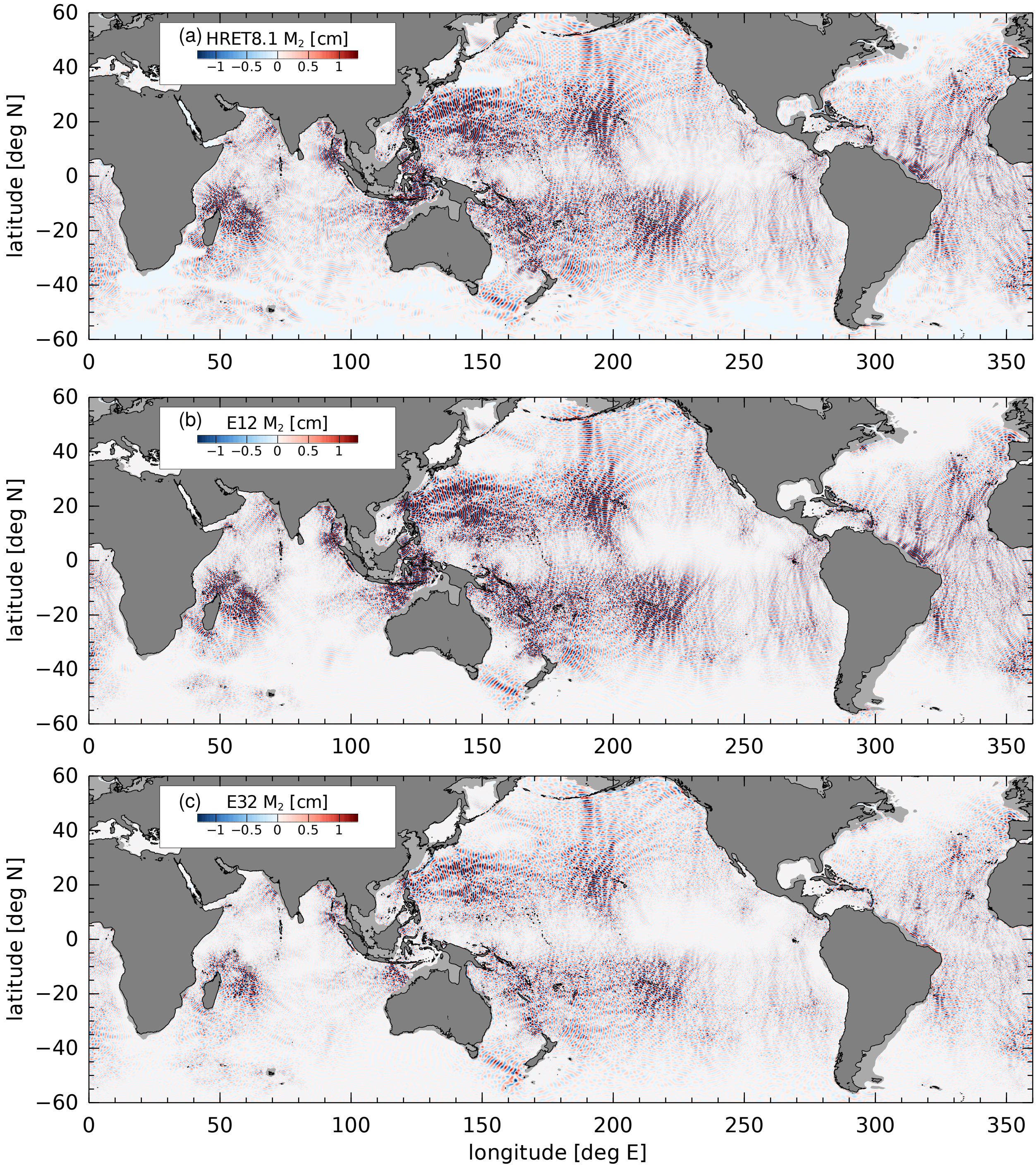


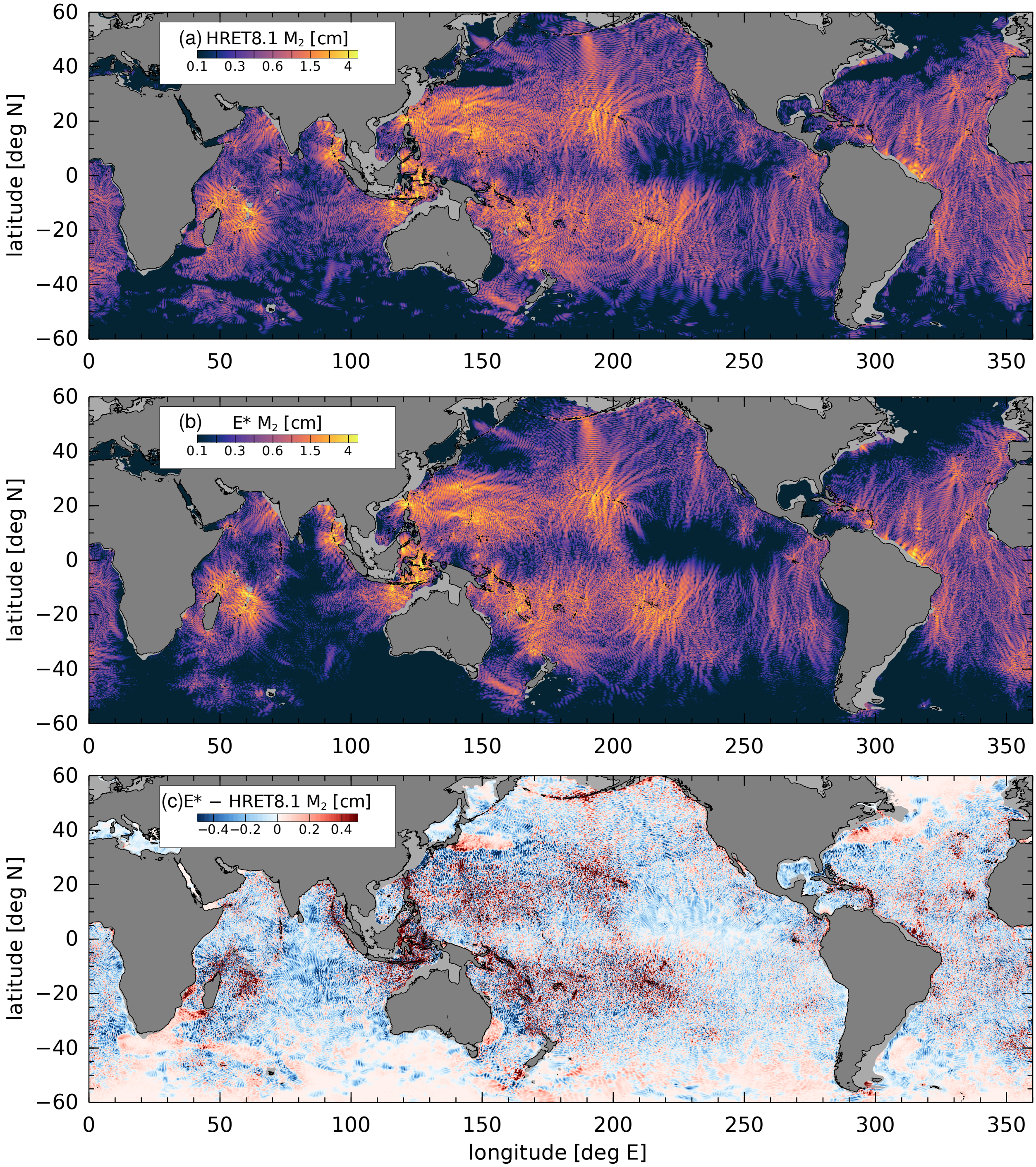
FIG. 5. Three estimates of the M_2 baroclinic sea level anomaly (real part, in-phase with the astronomical tidal potential): (a) HRET8.1, based on ERM data; (b) E12, based on ERM data using the estimator described in Section 2 with λ optimized by comparison to GM along-track sea surface slope data; (c) E32, based on GDP data, but otherwise as in (b).

and generally throughout the equatorial oceans. In addition, there are differences in clarity of the beam-like features in the South Atlantic, for example.

In order to develop the best possible estimate using both the ERM and GDP data, another estimate is formed, denoted E*, which is the optimal linear combination of E12 and E32 as cross-validated by the GM data. This estimate is computed as described in Appendix A. The generalized Fourier coefficients for the E12 and E32 estimates are computed using the independent ERM and GDP data; however, the optimal goodness-of-fit parameter λ, is determined by GM data in both cases. Thus, the E* estimate is not adding any new data to the estimate, it simply forms a weighted average which better agrees with the GM data used for validation.

The visual diffence between E* and the other solutions is subtle. Figure 6 illustrates the amplitude of the complex baroclinic sea level, η. It is apparent from comparison of panels (a) and (b) that the E* amplitude is smaller than the HRET8.1 amplitude in many regions where the amplitude of η is itself small (e.g., in the mid-Indian Ocean, and in the Eastern Equatorial Pacific). The difference in amplitudes, panel (c), illustrates the same point, but also shows that the E* estimate is larger than HRET8.1 in many of the generation sites. The E* estimate thus appears to be less noisy than HRET8.1, and more detailed near the generation sites.

Detailed comparisons of the E12 and E32 estimates, from which E* is constructed, show many small-scale differences, in spite of the large-scale similarity. Figure 7 illustrates the real part of E12 and E32 in the region around the Azores, a significant generation site in the North Atlantic, and a region with a high density of GDP data ($N_{GDP} \approx 4 \times 10^5$ within each D_m). The pattern of wave radiation, and the location of maxima and minima are reproduced in both estimates. However, the difference field, denoted δM$_2$, contains both small-scale noise and features which are coherent across multiple data windows, D_m. The difference field is a measure of the combined effects of random measurement error, mapping error, and systematic error possibly reflecting changes in the tides between the time periods over which the measurements were obtained (1993-2022 for the ERM data, and predominantly 2008-2020 for the GDP data). The difference between the E12 and E32 fields is assumed to be a measure of the uncertainty in E*. For this region, the root-mean-square amplitude differences are about 20% of the amplitude, and phase differences are about 15°. Phase differences are, of course, larger in regions with smaller amplitude signals. When the difference estimates are evaluated globally, differences in relative amplitude are directly related to



464 FIG. 6. Amplitude of the complex-valued baroclinic M_2 SLA for (a) the HRET8.1 estimate, and (b) the new E^* estimate, which is the optimal linear combination of the E12 and E32 estimates. The difference between the amplitudes, (b) minus (a), is shown in panel (c).

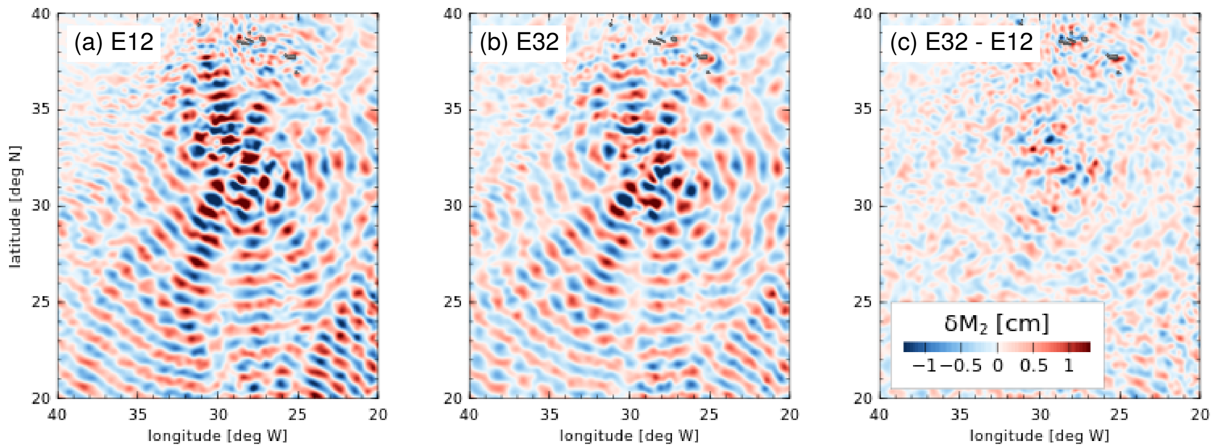


FIG. 7. The real part of the M_2 SLA for (a) E12, (b) E32, and (c) their difference, E32 minus E21, for a region around the Azores in the North Atlantic. The same colorscale is used in each panel, which is identical to that used in Figure 5 .

the number-density of GDP measurements (see Figure 3c) and increase to about 35% where N_{GDP} drops to 10^5 per D_m.

For practical purposes, one useful metric of the difference between the E* and HRET8.1 estimates is the prediction error, as measured by the explained variance with respect to different datasets. The explained variance (more-precisely, the reduction in mean-square residual from the model prediction; however, the mean values are negligable here and are expected to be zero), is defined as,

$$e_x = (|\mathbf{d}_x|^2 - |\mathbf{d}_x - \mathbf{H}_x\mathbf{F}\mathbf{a}|^2)/N_x, \tag{13}$$

where the x subscript indicates the data source, ERM, GM, or GDP. Figure 8 shows differences in the quantity, $\Delta e_x = e_x(\mathrm{E}^*) - e_x(\mathrm{HRET8.1})$, for the GM mission SLA (denoted Δe_{GM}), GM mission along-track SLA differences (denoted $\Delta e'_{GM}$), and GDP velocity data (denoted Δe_{GDP}). In most regions, particularly near the main internal tide generation sites, the E* estimate explains more variance than the older HRET8.1 estimate. Note that the magnitude of the change in explained variance is small for the GM data, and not likely to be of practical significance in most parts of the world Ocean; however, the difference is more substantial for the GDP dataset. Generally speaking, the small-magnitude and small-scale changes in baroclinic SLA (cf., Figure 5 and Figure 6c) alter

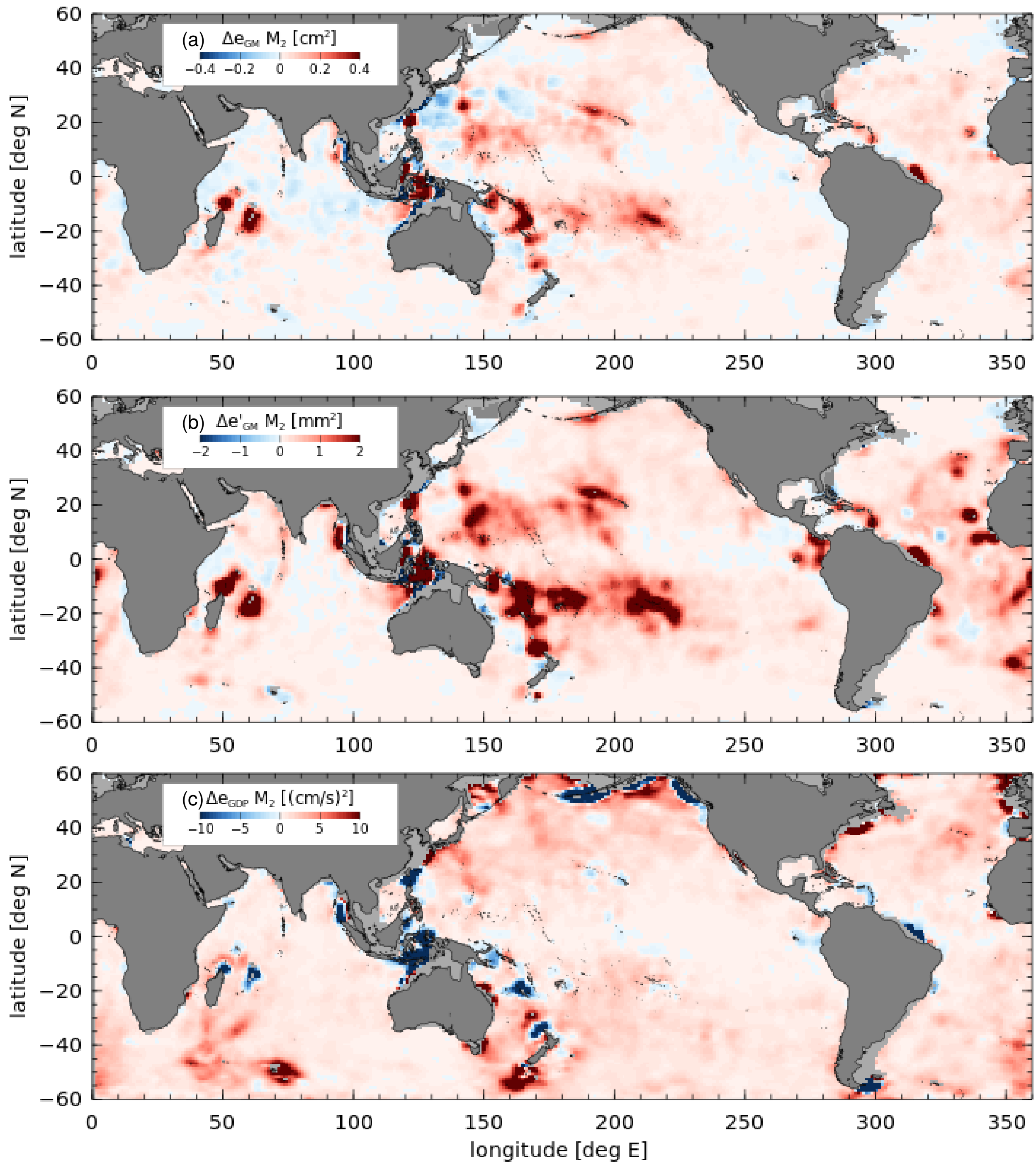


FIG. 8. Differences of explained variance, E^* minus HRET8.1. Positive values indicate that the E^* estimate explains more variance (e.g., it has smaller prediction error) than HRET8.1, based on the following datasets: (a) GM altimeter SLA, (b) GM altimeter along-track SLA differences, and (c) GDP surface velocity data.

the estimated sea surface slope in a manner which brings it in to better agreement with the observed velocity field.

b. MA_2, and MB_2 tides

The seasonal modulations of the M_2 tide, denoted with the Darwin symbols MA_2 and MB_2, were estimated in Zaron (2019) and included in HRET8.1. Figure 9 illustrates the amplitudes of these tides in the HRET8.1 and E^* estimates. The new estimates of the seasonal modulates are much smaller than the HRET8.1 estimates. Simultaneously, the E^* estimate explains more variance than the HRET8.1 estimate. Thus, it appears that the HRET8.1 estimates for these quantities were largely spurious.

There are two plausible reasons for the differences in the new estimates, compared to the old. First, MA_2 and MB_2 are dominated by the non-tidal environmental processes (Ray 2022), rather than the variability of the astronomical tidal potential. The E^* estimate is based on optimizing λ by comparison with GM data which were collected only within the latter half of the ERM record on which E12 is based, and the non-tidal processes responsible for MA_2 and MB_2 might decorrelate over these time periods. A similar argument applies to the GDP data used to form the E32 estimate. Another explanation for the difference involves the use of the $\alpha = 1$ estimator in the new approach; it may simply provide better protection against over-fitting, and result in less-noisy estimates. In any case, it is clear that the MA_2 and MB_2 estimates provided in HRET8.1 are not accurate.

c. S_2 and N_2 tides

The S_2 and N_2 tides represent the second and third largest astronomical tides in the semidiurnal group, with astronomical potential about 1/2 and 1/5 as large as M_2, respectively. The data available to estimate these two tides differ, though, because there are fewer (non-sun-synchronous orbit) satellite missions capable of observing S_2. Due to the sparse ERM sampling of S_2 and the small amplitude of N_2, both tides are estimated using a data window twice as large as used for M_2 (Table 1).

A comparison of the HRET8.1 and E^* estimates of S_2 identifies most of the same large-scale generation sites in both estimates (Figure 10). The newer estimate achieves better prediction error in mid-ocean regions associated with the main generation sites, but notable deficits occur

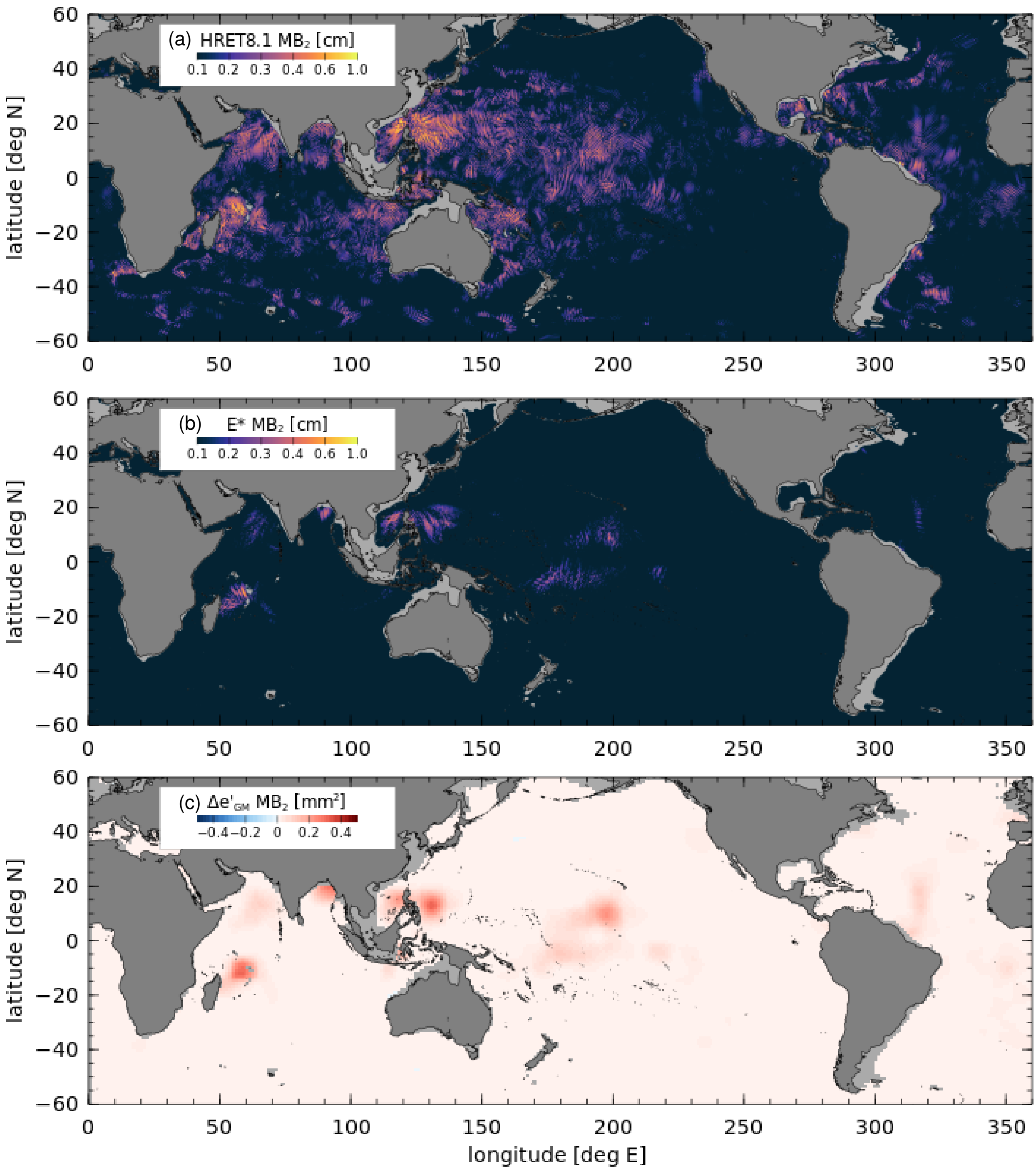
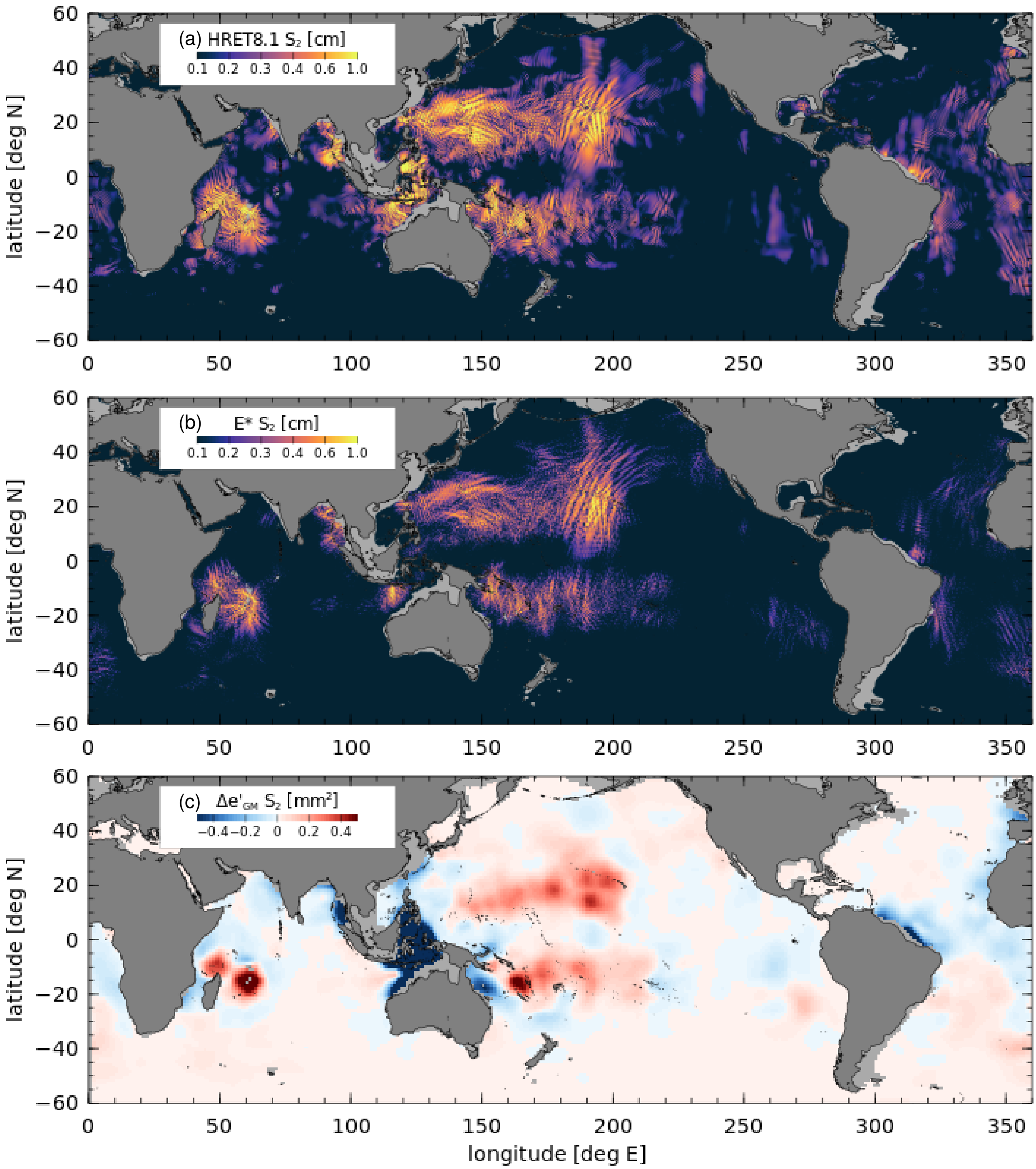


FIG. 9. Amplitude of the MB_2 tide, an annual modulate of M_2: (a) MB_2 from HRET8.1, and (b) MB_2 from E^*. (c) Difference of explained variance, E^* minus HRET8.1. The new estimates of MB_2 and MA_2 (not shown) are much smaller than in HRET8.1, and the explained variance is larger, which suggests that the HRET8.1 estimates for these frequencies was dominated by error.



538 FIG. 10. S_2 tide comparison: (a) HRET8.1 amplitude, (b) E^* amplitude, and (c) $\Delta e'_{GM}$, the difference in
539 explained variance compared to along-track GM SLA differences.

in regions such as the Indonesian Seas, near the Amazon plume, and near the Andamans on the eastern boundary of the Indian Ocean. It is hypothesized that the $\alpha = 0$-like approach of HRET8.1 is better in these regions, which are all close to land boundaries, where the inhomogeneities in the observations and larger non-tidal signals (due to lower accuracy of the mesoscale corrections near land) obscure the distinction between "active" and "non-active" degrees of freedom in $\alpha = 1$ algorithm. In other word, attempts to balance the active modes (a priori assumed to be signal) and in-active modes (a priori assumed to be a mix of signals and noise) are not useful when the in-active modes are dominated by noise. Although more systematic parameter tuning could lead to more accurate estimates of S_2 in these regions, this has not been attempted.

The N_2 tide was not estimated in HRET8.1, so there are no alternatives available for comparison of the E^* estimate (Figure 11). Overall, the maps of the N_2 tide are similar to M_2, as expected from the identical spatial pattern of forcing and nearby frequencies of the constituents, but the N_2 fields are notably sparser, consistent with their smaller signal-to-noise ratio and the use of the $\alpha = 1$ estimator. Incorporation of N_2 in predictions of baroclinic sea level should be expected to explain a few millimeters of root-mean-square sea level near the main generation sites.

d. K_1 and O_1 tides

The diurnal K_1 and O_1 tides were estimated previously in Zaron (2019), and it is interesting to see the extent to which that estimate can be improved with the new estimator and the addition of GDP data. Compared to HRET8.1, the new estimates are remarkably similar. The amplitudes of the new estimates are slightly larger and exhibit more detail than in HRET8.1, as shown in the Eastern Pacific and Atlantic for K_1 in Figures 12a and 12b. Waves emanating from Luzon and Halmahera Sea (Figure 12c) and southward from the Lombok Strait (Figure 12d) are resolved better than previously (not shown). Although the K_1 waves are smaller amplitude than M_2, it is evident that they propagate long distances within their waveguide between ±30° latitude. Similar features are found in the new O_1 solution.

The increased detail of the new diurnal tide estimates is associated with better explained variance statistics, shown for K_1 in the far western Pacific and Indonesian Seas (Figure 13). The extent of improvement for K_1 depends more on the specific data source used for comparison than M_2, though (cf., Figure 8). Considerable area now shows essentially the same or even slightly worse

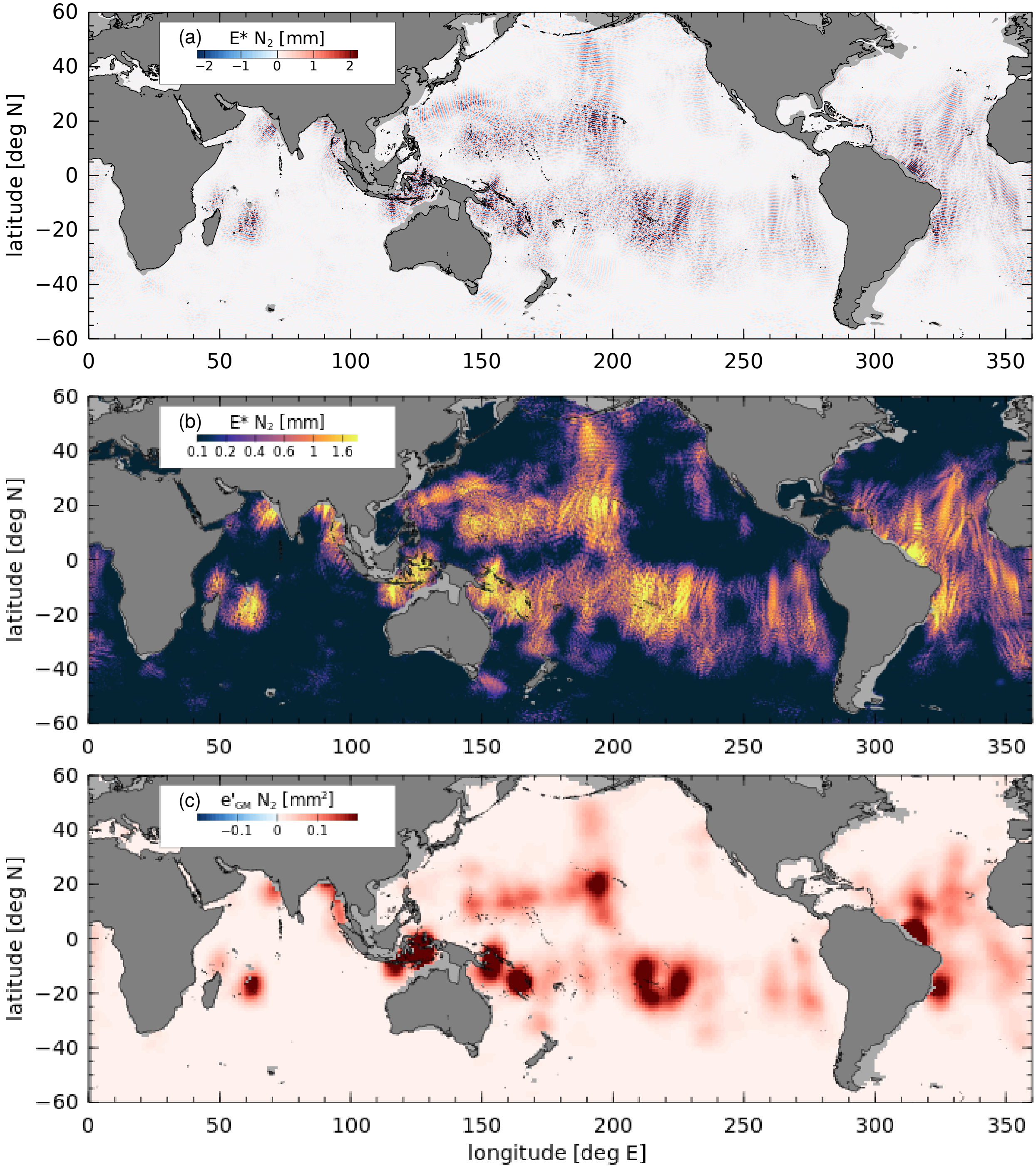


FIG. 11. N_2 tide metrics for the E* estimate: (a) the real part of SLA, (b) amplitude, and (c) explained variance with respect to along-track GM altimetry SLA differences. Note that the colorscale is scaled by a factor of 1/5 compared to plots for M_2 in Figures 5-8, corresponding to the magnitude of the astronomical potential.

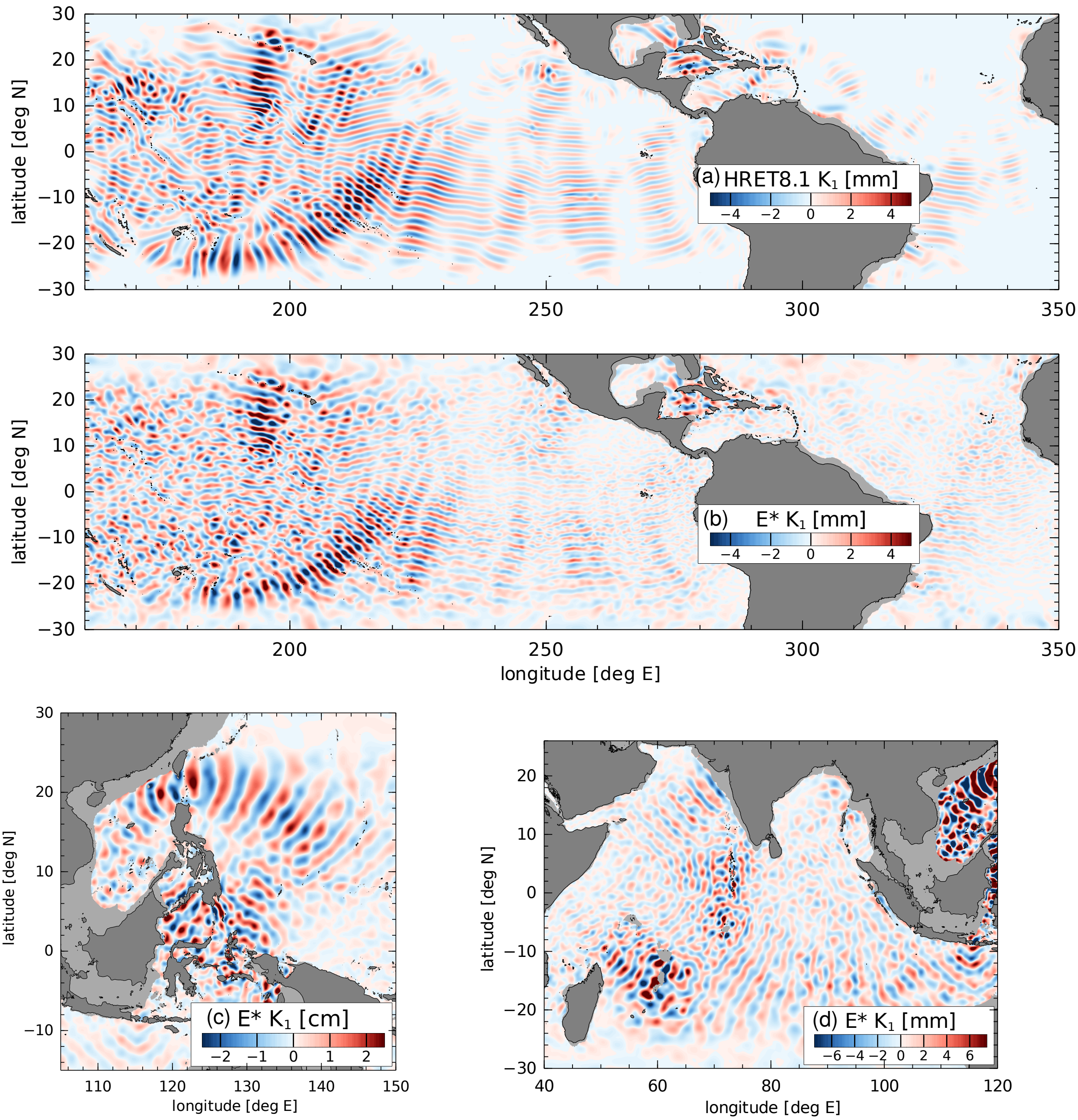


FIG. 12. K_1 baroclinic sea level anomaly (real part, in-phase with the astronomical tidal potential): (a) HRET8.1, based on ERM data; (b-d) E^*, based on combined ERM and GDP data using the estimator described in Section 2 with λ optimized by comparison to GM along-track sea surface slope. Note the different colorscales used in panels (b-d).

explained variance for the comparison with GM SLA data (Figure 13a); although, the comparison with GM along-track SLA difference is more uniformly favorable, which is expected since these

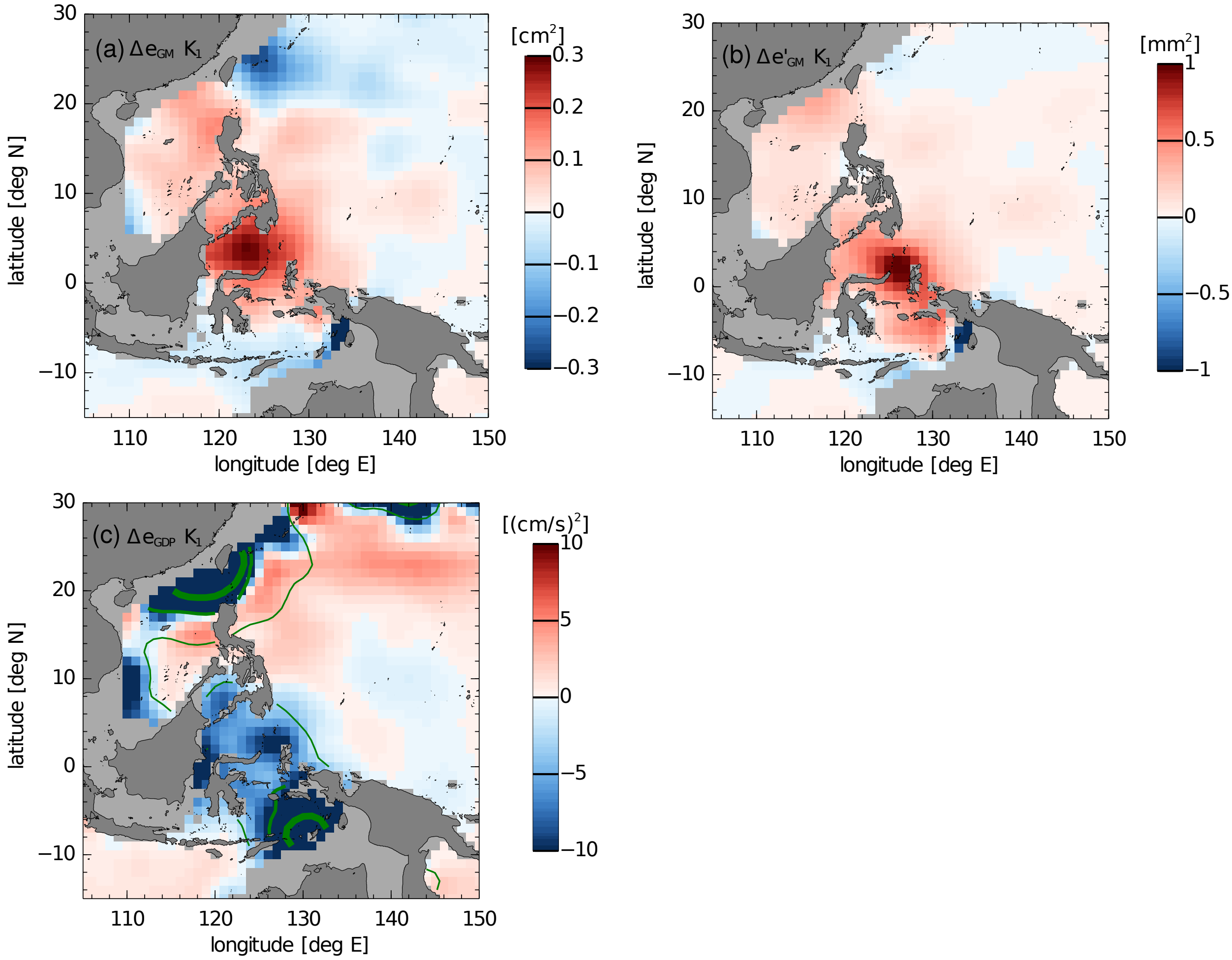


FIG. 13. Differences of explained variance, E* minus HRET8.1, within the same domain pictured in Figure 12c. Positive values indicate that the E* estimates explains more variance than HRET8.1. (a) GM altimeter SLA, (b) GM altimeter along-track SLA differences, and (c) GDP surface velocity data. The green contours in panel (c) indicate the estimated standard error of Δe_{GDP} at three levels: 1 $cm^2\,s^{-2}$ (thin), 3 $cm^2\,s^{-2}$ (medium), and 9 $cm^2\,s^{-2}$ (thick).

data are used to optimize the λ regularization parameter. The comparisons with GDP-derived currents shows the limitations of these data, caused in part by non-uniform spatial density and contamination by near-inertial currents near the K_1 inertial latitude, ±30°. The explained variance metric, Δe_{GDP} in Figure 13c, is overlaid with contours of its expected errors as estimated by a modification of bootstrap resampling which accounts for correlations in the data errors (data are resampled in 100-hr segments). Regions which explain less variance than HRET8.1 (shaded with blue) generally coincide with regions where the standard error is larger than 3 $cm^2\,s^{-2}$.

One interesting feature of K_1 is the apparent decay of the waves within about 3° of the inertial latitude (apparent near 190°E in Figure 12b and between 60°E and 80°E in Figure 12d). The structure of diurnal waves near their inertial latitude depends strongly on the direction of propagation, with an Airy-function structure for the meridional component (Hendershott 1981; Dushaw and Worcester 1998). But we should also expect the waves to be particularly sensitive to modulations of relative vorticity which will change the effective value of the inertial frequency in the propagation environment (Kunze 1985). Thus, the apparent decay of the waves near ±30° could be due to wave structures which cannot be represented by the present kinematic wave model, or due to time-dependent modulations which reduce the phase-locked signal. In either case, the interesting features of the diurnal waves deserve further investigation.

The O_1 tide is largely similar to the above description of K_1. It contains no striking new qualitative features, and is therefore not shown.

e. Summary of Results

Table 5 summarizes the main differences in the new E^* estimates compared to HRET8.1. It is noteworthy that the best-resolved M_2 tide is essentially the same amplitude in both estimates, but the new model explains nearly a factor of 5 more GDP velocity variance than HRET8.1. The use of the new $\alpha = 1$ estimator seems to reduce the noise level and problems with the MA_2 and MB_2 estimates from HRET8.1. In every case except S_2, the measures of explained variance are either the same or improved. There is no reason to think the dynamics of S_2 are fundamentally different from those of M_2 or N_2, so the larger difference between E^* and HRET8.1 for this frequency likely reflect its larger uncertainty, a consequence of the relatively poor sampling by altimetry.

6. Discussion

The new estimates of baroclinic tides, above, are useful for the prediction and description of the baroclinic waves. An illustrative example is shown in Figure 14a, which maps the surface kinetic energy associated with the baroclinic M_2 tide. The surface kinetic energy (Figure 14a) is computed from η using surface velocities computed with equation (8); it provides an estimate which may be useful in planning future ocean surface current measurement satellite mission concepts, for example (Rodríguez et al. 2019; Du et al. 2021). Figures 14b and 14c compare the rotary kinetic energy

TABLE 5. Summary statistics for E* compared to HRET8.1. $\langle|\eta|^2\rangle^{1/2}$ is the area-weighted root-mean-sqaure of the estimated tidal SLA amplitude. Other columns indicate the square root of the area-weighted mean of the explained variance with respect to ERM data (e_{ERM}), GM data (e_{GM}), along-track-differenced GM data (e'_{GM}), and GDP data (e_{GDP}). HRET8.1 did not include an estimate of N_2, which is indicated with an "-".

Tide	Model	$\langle\|\eta\|^2\rangle^{1/2}$ [cm]	e_{ERM} [cm]	e_{GM} [cm]	e'_{GM} [cm]	e_{GDP} [cm s^{-1}]
O_1	E*	0.24	0.18	0.18	0.02	0.67
	HRET8.1	0.22	0.17	0.18	0.02	0.36
K_1	E*	0.32	0.24	0.26	0.03	0.72
	HRET8.1	0.27	0.23	0.24	0.03	0.45
N_2	E*	0.05	0.06	0.05	0.02	0.37
	HRET8.1	-	-	-	-	-
MA_2	E*	0.03	0.02	0.03	0.01	0.35
	HRET8.1	0.11	0.00	0.00	0.00	0.00
M_2	E*	0.60	0.41	0.51	0.14	1.13
	HRET8.1	0.61	0.40	0.48	0.12	0.52
MB_2	E*	0.03	0.03	0.04	0.01	0.38
	HRET8.1	0.11	0.00	0.00	0.00	0.00
S_2	E*	0.15	0.16	0.14	0.04	0.62
	HRET8.1	0.19	0.15	0.15	0.04	0.13

of the observed Lagrangian currents (Elipot et al. 2016) with the predicted tidal currents. The examples show that the predicted currents exhibit considerable spreading in the frequency domain solely due to the Lagrangian character of the observations, independent of any non-phase-locked tidal variability. It is also interesting to note that the predicted tides in the semidiurnal band are a factor of 2 to 3 weaker than the observed currents. This difference in energy level likely reflects the presence of non-phase-locked tidal variability in the observations. In contrast, the kinetic energy near -1 cycle-per-day (cpd) consists of a mixture of tidal and inertial energy, and it is impossible to assess any bias in the predicted diurnal tidal currents.

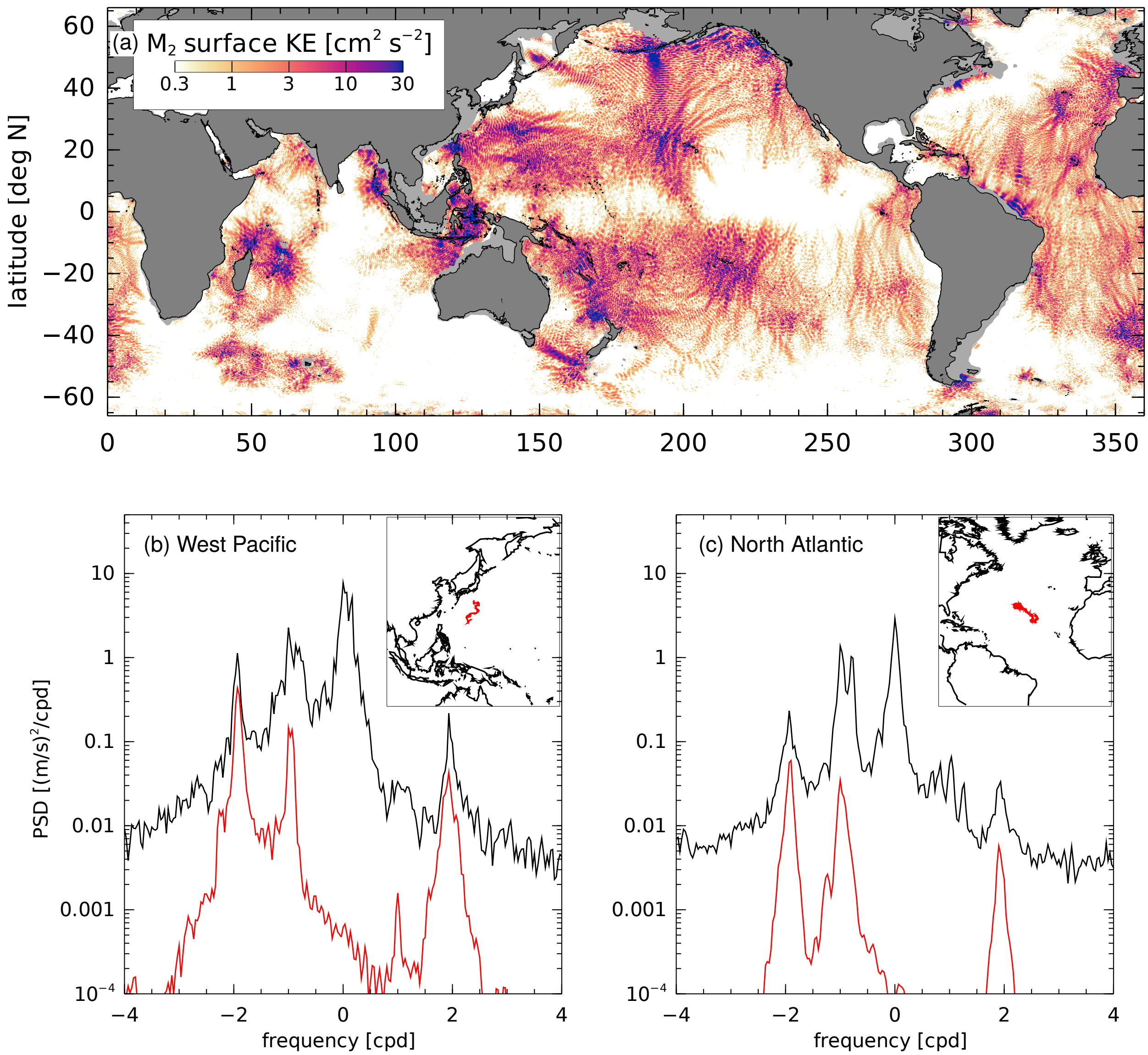


FIG. 14. (a) Kinetic energy of the baroclinic M_2 tide from the E^* estimate at the ocean surface. Note that the latitude range shown is expanded slightly compared to Figure 5 in order to show some small features at high latitude, such in the Labrador Sea. The rotary kinetic energy spectra of the observed Lagrangian currents (black) and predicted tidal currents (red) are shown for representative drifter trajectories in (b) the West Pacific and (c) the North Atlantic. Due to the drifters' proximity to 30° latitude, the kinetic energy near -1-cycle-per-day (cpd) consists of a mixture of tidal and inertial energy.

The new tidal estimates may also stimulate new approaches to oceanographic data analysis. For example, Figure 15 illustrates the probability distribution function of the squared baroclinic M_2 amplitude in two forms. The first, Figure 15a, quantifies the spatial inhomogeneity of the beams

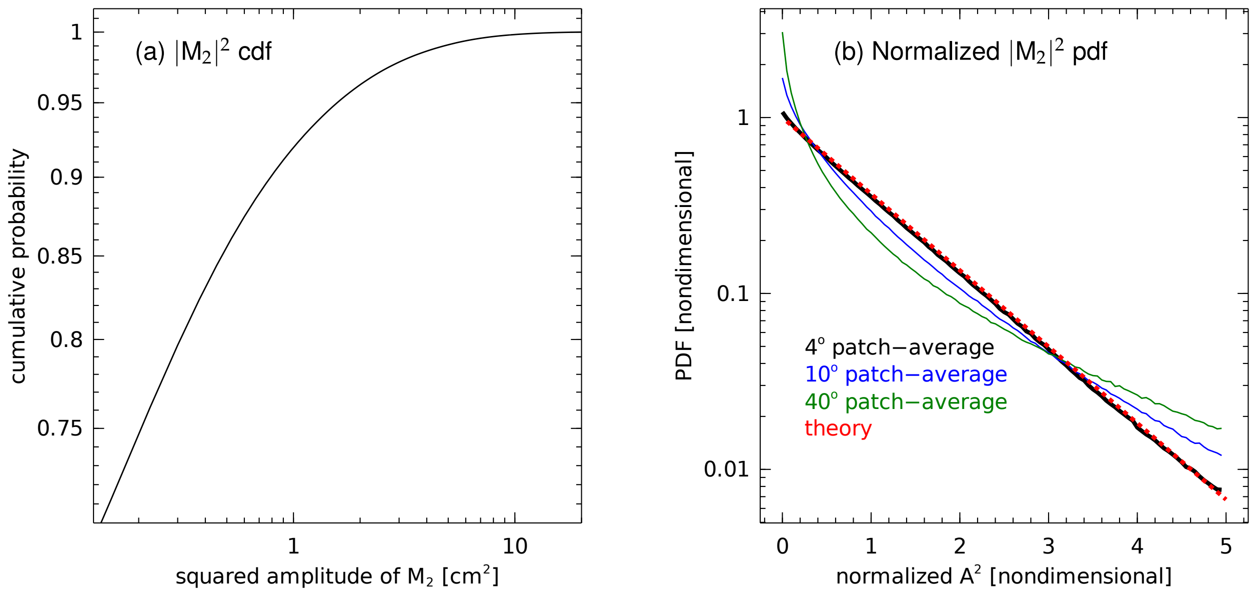


FIG. 15. The probability density function (pdf) of squared baroclinic M_2 amplitude. The pdf is computed by sampling the E* gridpoints, weighted by the cosine of latitude, so it represents the density of oceanic area (from 66°S to 66°N, and depth greater than 250 m) with given squared amplitude. (a) The cumulative probability density (cdf) quantifies the spatial heterogeneity of the internal tides, showing, for example, that baroclinic tides with amplitude exceeding 1 cm amplitude only occupy about 7.5% of the oceanic area. (b) For scales less than 400 km, the pdf of the normalized amplitude, $A^2 = |M_2|^2/\langle|M_2|^2\rangle$ (black line), agrees with the χ^2_2-distribution (red dashed line), consistent with strong multi-wave interference. Over larger scales the pdf is influenced by spatial heterogeneity of the wave field (blue and green lines).

of tidal energy; over 90% of the area of the deep ocean is associated with an M_2 amplitude of 1 cm or less. The second, Figure 15b, shows the probability density function and compares it with the exponential distribution, a model for squared wave amplitude in the presence of strong interference; it shows that within patches of the ocean in which the the wave field may be considered spatially homogeneous, its statistics are well-characterized by the exponential distribution.

It is important to recall that the E* estimate describes the phase-locked component of the baroclinic tides. The instantaneous tide, i.e., the sum of the phase-locked and modulated components, could be significantly different (Zaron 2022). It is also useful to note that the presence of diffraction patterns in the baroclinic tides is not necesssarily an indication of the phase-stability of the instantaneous tides, as has been asserted (Dushaw et al. 2011). A time-mean diffraction pattern could arise solely because of time-dependent modulations. For example, if two separate point sources turned on and off alternately, the mean wave field would exhibit an interference pattern even though none exists at any instant. Whether such a mechanism could explain any of the observed features of the wave field is speculative, but it is an interesting contrast to the interference of steady wave sources (Rainville et al. 2010).

The small amplitudes of the MA_2 and MB_2 tides found here is hypothesized to be related to non-phase-locked variability of the processes which modulate the M_2 tides. A better approach to capturing the baroclinic tidal variability at annual periods might use non-harmonic or year-by-year estimates, as implemented recently by Zhao (2022).

There are additional questions related to the form of the estimator used here. The mixed L_1/L_2-norm estimator is well known from the compressive-sampling literature (Candès et al. 2006), but it is not clear why it performs better than an L_2 estimator in the present application. There are other arbitrary choices, too, such as the size of the data-window and the wavenumber bandwidth. Given the heterogeneous nature of the fields to be mapped, the non-tidal noise, and the character of the observing arrays, it is unlikely that a single set of parameters are optimal for all of the local domains, D_m.

One particularly vexing issue concerns the appropriate estimators near the coastline or topographic features where barotropic tidal corrections may contain errors and where the barotropic and baroclinic dynamics may be coupled. This is a serious limitation of the kinematic wave approach, since the separation of barotropic and baroclinic sea level anomalies depends on the utility of the dispersion relation which, for all estimates made so far, has been derived in the limit of flat-bottom topography. It is not clear how the existing approaches to mapping or assimilation based on dynamical models (e.g., Egbert and Erofeeva 2014) could be modified to exceed the accuracy of the kinematic models, though. The mixed L_1/L_2 approach could be adapted for use in assimilative models to allow better control of model complexity; although, this would not address the approximations inherent in models based on a truncated modal decomposition (Lahaye and Llewellyn Smith 2020).

The presentation of results in Section 5, above, used explained variance to assess the tidal estimates; however, an alternative approach to assessment could be provided by estimates of standard errors. Users of HRET concerned with tidal prediction are generally interested in the accuracy of the prediction relative to the instantaneous tide, so the standard error of only the

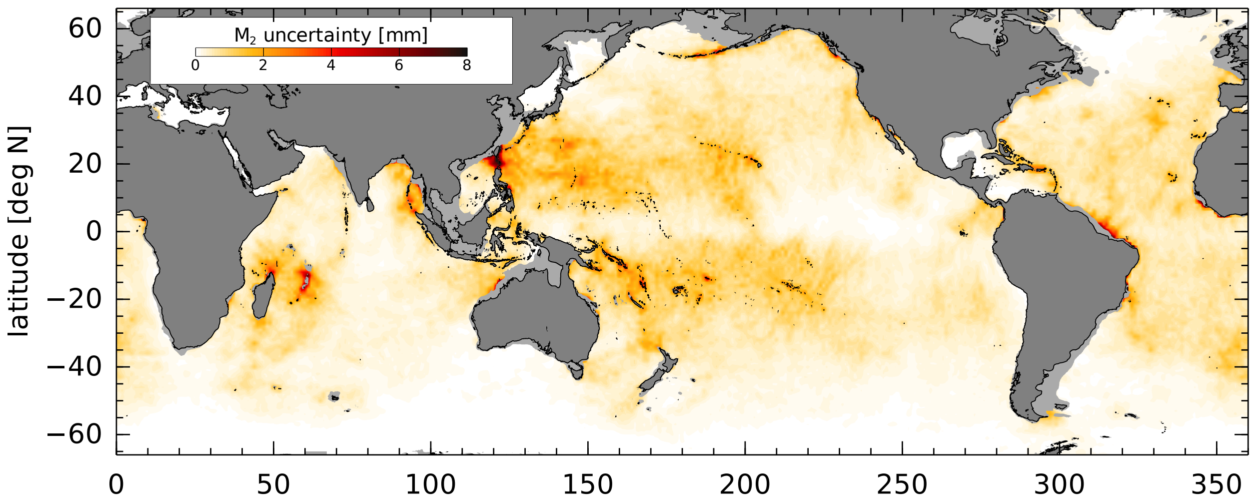


FIG. 16. An error estimate for the M_2 harmonic constants based on the assumption that the E12 and E32 estimates are affected by the same (non-instrumental) noise sources, namely, non-phase-locked tides and model structural error.

phase-locked tide may be of little practical use. Nonetheless, standard errors for the phase-locked tide could help assess differences among models, and a brief discussion of these errors follows.

Errors in the tidal estimates provided here arise from a combination of systematic and random errors in the observational data, model structural error (uncertainty in the form of the functions used to describe the data, equation 2), and noise due to non-phase-locked tides. If it is assumed that the latter two factors are dominant and affect the altimeter- and GDP-derived estimates equally, then the errors in these estimates should depend only on the number of data used for each estimate. In other words, assume that the E12 and E32 estimates are independent but affected by noise with same variance, σ^2. The errors of the estimates should be, $\sigma^2_{E12} = \sigma^2/N_{E12}$ and $\sigma^2_{E32} = \sigma^2/N_{E32}$, respectively, where N_e is the number of data minus the number of parameters estimated in estimate $e \in \{E12, E32\}$. With these assumptions, the difference field $\Delta = E12 - E32$ may be used to estimate σ^2,

$$\sigma^2 = \langle \Delta^2 \rangle \left(\frac{1}{N_{E12}} + \frac{1}{N_{E32}} \right), \tag{14}$$

where the angle-brackets denote spatial averaging over 1-degree cells. This error estimate for M_2 is shown in Figure 16.

7. Conclusions

The estimates for the baroclinic tides derived herein will be useful for the prediction of baroclinic tidal variability in the open ocean. The new methdology for estimating and mapping the wave fields is a definite improvement compared to prior methods. It is hoped that these estimators may be useful for analysis of temporal and spatial subsets of data, to identify non-phase-locked tidal variability, or to estimate the dispersion-relation parameters from the mapped fields.

Considering the small quantitative improvement of the present estimates compared to the older HRET8.1 estimates for mapping or predicting SLA, further efforts to map the time-mean phase-locked baroclinic tides with kinematic waves appear to be of questionable value. It seems that the greatest gains in baroclinic tide prediction will result from mapping or predicting the "instantaneous" tides, for which new methods need to be developed. The approaches in Egbert and Erofeeva (2021), Le Guillou et al. (2021), Ubelmann et al. (2022), and Zhao (2022) appear promising.

Nonetheless, the tidal estimates presented here are useful for predicting the baroclinic tidal SLA and ocean surface currents. The harmonic constants on a regular $(1/20)^\circ$ grid between the latitudes of $\pm 66^\circ$, and associated tidal prediction software, are publicly available[1] for the M_2, S_2, N_2, K_1, and O_1 tides. The netcdf-formatted files contain the E12 (altimeter-only) and E31 (drifter-only) tidal estimates, in addition to the optimal E^* estimate, and harmonic constants are provided for the vector surface currents as well as the SLA. Sub-surface tidal currents and other baroclinic tidal fields may be predicted using the generalized Fourier coefficient representation of each tide; these coefficients comprise roughly 60 GB of data for each constituent and are available from the authors upon request.

Acknowledgments. Computations were performed using the Portland Institute for Computational Science cluster computer, *Coeus*, acquired with support from NSF award #DMS1624776 and ARO award #W911NF-16-1-0307. Zaron was supported by NSF award #1850961 and NASA award #80NSSC21K0346, and Elipot was supported by NSF award #1851166. We appreciate the efforts of funding agencies, scientists, and technicians who have created the datasets used here, which are publically-disseminated as described in the Data Availability Statement, below.

[1]See the files at `https://ingria.ceoas.oregonstate.edu/fossil/SMCE/dir?ci=tip&name=HRET14`. The following are direct links to download the netcdf files: **HRET14 SSH file** **HRET14 surface velocity file**. [Note that the direct links will be removed and a permanent archival DOI or URL will be provided for the final draft].

Data Availability Statement

All of the data used in this research are publically available. The along-track altimetry data were extracted from the Radar Altimeter Database System (RADS), http://rads.tudelft.nl/rads/rads.shtml, which is an effort of the Department of Earth Observation and Space Systems at TU Delft and the NOAA Laboratory for Satellite Altimetry. The surface drifter data were obtained from the NOAA Global Drifter Program web page, https://www.aoml.noaa.gov/phod/gdp/hourly_data.php. The DT-2021 version of the SSALTO/DUACS multi-satellite gridded sea level anomaly maps were obtained from the Copernicus Marine Environmental Monitoring Service, http://marine.copernicus.eu. The HRET8.1 tidal estimates are available at the lead author's website, https://ingria.ceoas.oregonstate.edu.

The authors have submitted the data files containing the tidal SLA and surface current harmonic constants to the NASA PO.DAAC and Zenodo for data archiving; it is anticipated that a DOI for these datasets will be included in the published version of this manuscript. The reviewers may access the data files and related software at the following url: https://ingria.ceoas.oregonstate.edu/fossil/SMCE/dir?ci=tip.

APPENDIX A

Definition of the E* estimator

As stated in the text, the E* estimate is an optimal linear combination of the E12 and E32 estimators. Let $\mathbf{a}^*$, $\mathbf{a}^{(12)}$, and $\mathbf{a}^{(32)}$ denote the generalized Fourier coefficients for these estimates, respectively, with $\mathbf{a}^* = a\mathbf{a}^{(12)} + (1-a)\mathbf{a}^{(32)}$ for scalar, a, to be deteremined. Within each data patch, a is chosen to maximize the explained variance with respect to the along-track-differenced GM data, the quantity,

$$J(a) = \mathbf{d}_{GM}^T \mathbf{d}_{GM} - (\mathbf{d}_{GM} - \mathbf{H}_{GM}\mathbf{F}\mathbf{a}^*)^T(\mathbf{d}_{GM} - \mathbf{H}_{GM}\mathbf{F}\mathbf{a}^*). \tag{A1}$$

The coefficient a is obtained by solving conditions where the quadratic function $J(a)$ is maximum,

$$\frac{1}{2}\frac{dJ(a)}{da} = [\mathbf{H}_{GM}\mathbf{F}(\mathbf{a}^{(12)} - \mathbf{a}^{(32)})]^T[\mathbf{d}_{GM} - \mathbf{H}_{GM}\mathbf{F}(a\mathbf{a}^{(12)} + (1-a)\mathbf{a}^{(32)})] = 0. \tag{A2}$$

Using the notation $\mathbf{d}_{GM}^{(12)} = \mathbf{H}_{GM}\mathbf{F}\mathbf{a}^{(12)}$ for $\mathbf{a}^{(12)}$ sampled at the GM data sites, and likewise for $\mathbf{d}_{GM}^{(32)}$, the optimal value of a is derived as follows:

$$0 = (\mathbf{d}_{GM}^{(12)} - \mathbf{d}_{GM}^{(32)})^T(\mathbf{d}_{GM} - a\mathbf{d}_{GM}^{(12)} - (1-a)\mathbf{d}_{GM}^{(32)}) \tag{A3}$$

$$0 = (\mathbf{d}_{GM}^{(12)} - \mathbf{d}_{GM}^{(32)})^T(\mathbf{d}_{GM} - \mathbf{d}_{GM}^{(32)} - a(\mathbf{d}_{GM}^{(12)} - \mathbf{d}_{GM}^{(32)})) \tag{A4}$$

$$0 = (\Delta\mathbf{d}_{GM})^T(\mathbf{d}_{GM} - \mathbf{d}_{GM}^{(32)} - a(\Delta\mathbf{d}_{GM})) \tag{A5}$$

$$0 = (\Delta\mathbf{d}_{GM})^T(\mathbf{d}_{GM} - \mathbf{d}_{GM}^{(32)}) - a(\Delta\mathbf{d}_{GM})^T(\Delta\mathbf{d}_{GM}) \tag{A6}$$

$$a = \frac{(\Delta\mathbf{d}_{GM})^T(\mathbf{d}_{GM} - \mathbf{d}_{GM}^{(32)})}{(\Delta\mathbf{d}_{GM})^T(\Delta\mathbf{d}_{GM})}. \tag{A7}$$

The E* estimate is computed independently on each domain, D_m, and these solutions are blended together, following Section 2b, to obtain the maps of global fields which are shown in the main text.

References

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Trans. on Automat. Cont.*, **19**, 716–723.

Arbic, B. K., and Coauthors, 2022: Near-surface ocean kinetic energy distributions from drifter observations and numerical models. *J. Geophys. Res.*, **127 (10)**, e2022JC018 551, https://doi.org/10.1029/2022JC018551.

Candès, E., J. Romberg, and T. Tao, 2006: Stable signal recovery from incomplete and inaccurate measurements. *Comm. on Pure and Applied Math.*, **59 (8)**, 1207–1223.

Carrere, L., and Coauthors, 2021: Accuracy assessment of global internal tide models using satellite altimetry. *Ocean Sci.*, **17**, 147–180.

Dong, W., O. Buhler, and K. S. Smith, 2020: Frequency diffusion of waves by unsteady flows. *J. Fluid Mech.*, **905**, R3.

Du, Y., and Coauthors, 2021: Ocean surface current multiscale observation mission (oscom): Simultaneous measurement of ocean surface current, vector wind, and temperature. *Prog. Oceanogr.*, **193**, 102 531.

Dushaw, B., 2015: An empirical model for mode-1 internal tides derived from satellite altimetry: Computing accurate tidal predictions at arbitrary points over the world oceans. Tech. Rep. APL-UW TM 1-15, University of Washington Applied Physics Laboratory, 114 pp.

Dushaw, B. D., and P. F. Worcester, 1998: Resonant diurnal internal tides in the North Atlantic. *Geophys. Res. Lett.*, **25 (12)**, 2189–2192.

Dushaw, B. D., P. F. Worcester, and M. A. Dzieciuch, 2011: On the predictability of mode-1 internal tides. *Deep Sea Res.*, **58**, 677–698.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, 2004: Least angle regression. *The Annals of Statistics*, **32 (2)**, 407–499.

Egbert, G. D., and S. Y. Erofeeva, 2014: Mapping M2 internal tides using a data-assimilative reduced gravity mode. American Geophysical Union, AGU Fall Meeting 2014AGUFMOS43F.

Egbert, G. D., and S. Y. Erofeeva, 2021: An approach to empirical mapping of incoherent internal tides with altimetry data. *Geophys. Res. Lett.*, **48**, e2021GL095 863, https://doi.org/10.1029/2021GL095863.

Elipot, S., and R. Lumpkin, 2008: Spectral description of oceanic near-surface variability. *Geophys. Res. Lett.*, **35**, L05 606.

Elipot, S., R. Lumpkin, R. C. Perez, J. M. Lilly, J. J. Early, and A. M. Sykulski, 2016: A global surface drifter data set at hourly resolution. *J. Geophys. Res.*, **65 (1)**, 29–50, https://doi.org/10.1002/2016JC011716.

Elipot, S., A. Sykulski, R. Lumpkin, L. Centurioni, and M. Pazos, 2022a: A dataset of hourly sea surface temperature from drifting buoys. *Sci. Data*, **9**, 567, https://doi.org/10.1038/s41597-022-01670-2.

Elipot, S., A. Sykulski, R. Lumpkin, L. Centurioni, and M. Pazos, 2022b: Hourly location, current velocity, and temperature collected from global drifter program drifters world-wide, https://doi.org/10.25921/x46c-3620, dataset, accessed 5/30/2023.

Foreman, M. G., J. Y. Cherniawsky, and V. A. Ballantyne, 2009: Versatile harmonic tidal analysis: Improvements and applications. *J. Atm. and Ocean. Tech.*, **26**, 806–817.

Hendershott, M. C., 1981: Long waves and ocean tides. *Evolution of Physical Oceanography*, B. A. Warren, and C. Wunsch, Eds., MIT Press, 293–341.

Kafiabad, H., M. Savva, and J. Vanneste, 2019: Diffusion of inertia-gravity waves by geostrophic turbulence. *J. Fluid Mech.*, **869**, R7.

Kelly, S. M., 2016: The vertical mode decomposition of tides in the presence of a free surface and arbitrary topography. *J. Phys. Oceanogr.*, **46**, 3777–3788.

Kelly, S. M., A. F. Waterhouse, and A. C. Savage, 2021: Global dynamics of the stationary M_2 mode-1 internal tide. *Geophys. Res. Lett.*, **48**, e2020GL091 692.

Kodaira, T., K. R. Thompson, and N. B. Bernier, 2016: Prediction of M_2 tidal surface currents by a global baroclinic ocean model and evaluation using observed drifter trajectories. *J. Geophys. Res.*, **121 (8)**, 6159–6183.

Kunze, E., 1985: Near-inertial wave propagation in geostrophic shear. *J. Phys. Oceanogr.*, **15**, 544–565.

Lahaye, N., and S. G. Llewellyn Smith, 2020: Modal analysis of internal wave propagation and scattering over large-amplitude topography. *J. Phys. Oceanogr.*, **50 (2)**, 305–321.

Le Guillou, F., and Coauthors, 2021: Joint estimation of balanced motions and internal tides from future wide-swath altimetry. *J. Adv. Model. Earth Sys.*, **13**, e2021MS002 613.

Lysen, S., 2009: Permuted inclusion criterion: a variable selection technique, URL https://repository.upenn.edu/edissertations/28, publicly Accessible Penn Dissertations, accessed 5/30/2023.

Mallows, C. L., 1973: Some comments on c_p. *Technometrika*, **15**, 661–675.

Matte, P., D. A. Jay, and E. D. Zaron, 2013: Adaptation of classical tidal harmonic analysis to non-stationary tides, with application to river tides. *J. Atm. and Ocean. Tech.*, **30**, 569–589.

Poulain, P.-M., and L. Centurioni, 2015: Direct measurements of world ocean tidal currents with surface drifters. *J. Geophys. Res.*, **120**, 6986–7003.

Rainville, L., T. S. Johnston, G. S. Carter, M. A. Merrifield, R. Pinkel, P. F. Worcester, and B. D. Dushaw, 2010: Interference pattern and propagation of the M_2 internal tide south of the Hawaiian Ridge. *J. Phys. Oceanogr.*, **40**, 311–325.

Ray, R. D., 2022: Technical note: On seasonal variability of the M2 tide. *Ocean Sci.*, **18 (4)**, 1073–1079.

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

Ray, R. D., and E. D. Zaron, 2016: M_2 internal tides and their observed wavenumber spectra from satellite altimetry. *J. Phys. Oceanogr.*, **46**, 3–22.

Rodríguez, E., M. Bourassa, D. Chelton, J. T. Farrar, D. Long, D. Perkovic-Martin, and R. Samelson, 2019: The winds and currents mission concept. *Front. Mar. Sci.*, **6**, https://doi.org/10.3389/fmars.2019.00438.

Savva, M., and J. Vanneste, 2018: Scattering of internal tides by barotropic quasigeostrophic flows. *J. Fluid Mech.*, **856**, 504–530.

Scharroo, R., E. W. Leuliette, J. L. Lillibridge, D. Byrne, M. C. Naeije, and G. T. Mitchum, 2013: RADS: Consistent multi-mission products. *Proc. of the Symposium on 20 Years of Progress in Radar Altimetry, Eur. Space Agency Spec. Publ., ESA SP-710*, Venice, 20-28 September 2012, 4, URL http://rads.tudelft.nl/rads/rads.shtml.

Shanno, D., 1985: Globally convergent conjugate-gradient algorithms. *Math. Program.*, **33**, 61–67.

Shriver, J. F., J. G. Richman, and B. K. Arbic, 2014: How stationary are the internal tides in a high-resolution global ocean circulation model? *J. Geophys. Res.*, **119**, 2769–2787.

Ubelmann, C., L. Carrere, C. Durand, G. Dibarboure, Y. Faugère, M. Ballarotta, F. Briol, and F. Lyard, 2022: Simultaneous estimation of ocean mesoscale and coherent internal tide sea surface height signatures from the global altimetry record. *Ocean Science*, **18 (2)**, 469–481, https://doi.org/10.5194/os-18-469-2022.

Wendland, H., 1995: Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.*, **4**, 389–396.

Yuan, M., and Y. Lin, 2006: Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B*, **68 (1)**, 49–67.

Zaron, E. D., 2018: Ocean and ice shelf tides from CryoSat-2 altimetry. *J. Phys. Oceanogr.*, **48**, 975–993.

Zaron, E. D., 2019: Baroclinic tidal sea level from exact-repeat mission altimetry. *J. Phys. Oceanogr.*, **49 (1)**, 193–210.

Zaron, E. D., 2022: Baroclinic tidal cusps from satellite altimetry. *J. Phys. Oceanogr.*, **52 (12)**, 3123–3137.

Zaron, E. D., and S. Elipot, 2021: An assessment of global ocean barotropic tide models using geodetic mission altimetry and surface drifters. *J. Phys. Oceanogr.*, **51 (1)**, 63–82, https://doi.org/10.1175/JPO-D-20-0089.1.

Zaron, E. D., and R. D. Ray, 2017: Using an altimeter-derived internal tide model to remove tides from in situ data. *Geophys. Res. Lett.*, **44**, 4241–4245.

Zaron, E. D., and R. D. Ray, 2018: Aliased tidal variability in mesoscale sea level anomaly maps. *J. Atm. and Ocean. Tech.*, **35 (12)**, 2421–2435.

Zhao, Z., 2022: Development of the yearly mode-1 M_2 internal tide model in 2019. *J. Atm. and Ocean. Tech.*, **39 (4)**, 463–478.

Zhao, Z., 2023: Satellite evidence for strengthened m2 internal tides in the past 30 years. *Geophys. Res. Lett.*, **50**, e2023GL105 764.

Zhao, Z., M. H. Alford, and J. B. Girton, 2012: Mapping low-mode internal tides from multisatellite altimetry. *Oceanography*, **25 (2)**, 42–51.

Zhao, Z., M. H. Alford, J. B. Girton, L. Rainville, and H. L. Simmons, 2016: Global observations of open-ocean mode-1 M_2 internal tides. *J. Phys. Oceanogr.*, **46**, 1657–1684.

Zhao, Z., M. H. Alford, R. C. Lien, M. C. Gregg, and G. S. Carter, 2014: Internal tides and mixing in a submarine canyon with time-varying stratification. *J. Phys. Oceanogr.*, **42 (12)**, 2121–2142.