An Assessment of Global Ocean Barotropic Tide Models Using Geodetic Mission Altimetry and Surface Drifters

Edward D. Zaron*

Portland State University, Portland, Oregon, USA

Shane Elipot

Rosenstiel School of Atmospheric and Marine Science, University of Miami, Miami, Florida, USA

**Corresponding author address:* Edward D. Zaron, Department of Civil and Environmental Engineering, Portland State University, P.O. Box 751, Portland, OR 97207

E-mail: ezaron@pdx.edu

ABSTRACT

The accuracy of three data-constrained barotropic ocean tide models is assessed by comparison with data from geodetic mission altimetry and ocean surface drifters, data sources chosen for their independence from the observational data used to develop the tide models. Because these data sources do not provide conventional time series at single locations suitable for harmonic analysis, model performance is evaluated using variance reduction statistics. The results distinguish between shallow and deep water evaluations of the GOT410, TPXO9A, and FES2014 models; however, a hallmark of the comparisons is strong geographic variability which is not well-summarized by global performance statistics. The models exhibit significant regionally-coherent differences in performance which should be considered when choosing a model for a particular application. Quantitatively, the differences in explained SSH variance between the models in shallow water are only 1-2% of the root-mean-square (RMS) tidal signal of about 50 cm, but the differences are larger at high latitudes, more than 10% of 30 cm RMS. Differences with respect to tidal currents variance are strongly influenced by small-scales in shallow water and are not well represented by global averages; therefore, maps of model differences are provided. In deep water, the performance of the models is practically indistinguishable from one another using the present data. The foregoing statements apply to the 8 dominant astronomical tides M_2, S_2, N_2, K_2, K_1, O_1, P_1, and Q_1. Variance reduction statistics for smaller tides are generally not accurate enough to differentiate the models' performance.

1. Introduction

Progress in the knowledge of tides, particularly in the deep ocean, may be largely attributed to measurements acquired by satellite altimeters beginning with the TOPEX/Poseidon mission in 1993 (LeProvost 2001; Ray and Egbert 2017). Global models of the barotropic ocean tides may be used to predict tidal sea levels with an accuracy of 1 cm or better in many parts of the world ocean (Ray and Byrne 2010; Stammer et al. 2014). This accuracy has been achieved, in large measure, by meticulous analysis of data from multiple altimeter missions, coastal tide gauges, and bottom pressure recorders. Because of the importance of accurate tidal predictions, nearly all available data are incorporated into the development of the models. While this approach has benefits, the relative lack of independent data makes it challenging to evaluate the latest models, to measure improvement, and to identify where and how they might be further refined.

The goal of this paper is to assess the most recent versions of independently-developed barotropic tide models available from different groups, and to update the assessment provided by Stammer et al. (2014) with more recent validation data. Emphasis is placed on using novel data, namely, (1) sea surface height (SSH) measurements collected by altimeter missions in long-repeat orbits not commonly used for tidal analysis, and (2) measurements of ocean surface velocity inferred from drifting buoys. Unfavorable aliasing of tidal frequencies (for altimetry) and Lagrangian sampling (for drifters) makes it problematic to compute accurate tidal harmonic constants from these data, which would provide the best information for comparing with tide models. Instead, a variance reduction statistic, the explained variance, is used to compare the models; however, interpreting this statistic requires attention to the influence of signals that are correlated with the tides of interest.

Stammer et al. (2014) compared three different types of tide models: (1) purely hydrodynamic models, which predict the tides by numerically integrating the equations of motion forced by the astronomical tidal potential with corrections for ocean self-attraction and solid earth loading, (2) purely empirical models, which predict tides using harmonic constants inferred from satellite altimeter data, and (3) data assimilative models, which systematically combine information used in the previous approaches. Herein we consider updated versions of three models, namely, the TPXO Atlas, version 9.2 (TPXO9A), the Finite Element Solution, version 2014 (FES2014), and the Goddard/Grenoble Ocean Tide model, version 4.10c (GOT410), developed using methods described in detail in Egbert and Erofeeva (2002), Lyard (1999), and Ray (1999), respectively. TPXO9A and FES2014 are models of the data assimilative type, (3) above, while GOT410 is an empirical model, (2) above. All three models are widely used to compute tidal predictions for a variety of applications in oceanography and geophysics. Among the other models considered in Stammer et al. (2014), note that the DTU model has been updated to DTU16, but since this used the older FES2012 model as a prior, it is not considered here. With the exception of the purely hydrodynamic models, the other models considered in Stammer et al. (2014) have not been updated and are not considered. The hydrodynamic model examined in Stammer et al. (2014), global HYCOM, has been subject to ongoing improvement and assessment elsewhere (Ngodock et al. 2016; Savage et al. 2017; Nelson et al. 2019).

The software used for computing tidal predictions from these models differs slightly in each case to accommodate the harmonic constants provided and the choice of minor constituents computed by inference. Because of these differences, comparisons of the tidal predictions are ambiguous to interpret, since differences may be due to errors in the tidal harmonic constants, differences in how atmospheric tides are treated, or differences in the inference methodology. Thus, previous assessments have generally emphasized comparisons of the models with observed harmonic constants, rather than tidal predictions, per se. While this approach usefully isolates the comparisons to individual tidal constituents, it limits the comparison data to those for which long time series are available for computing accurate harmonic constants.

As an alternative to comparisons of harmonic constants, the approach taken here uses variance reduction statistics to evaluate model performance. Thus, a partial tide prediction is computed for a single constituent, say, M_2, and this prediction is subtracted from the observations. The variance of the residual, and the difference in variance compared to the original, may then be compared among the models. An advantage of this approach is that it permits the use of more types of data for model intercomparison than could be used for comparisons of harmonic constants. This is particularly important because the largest differences among the models generally occur near the coastline, at spatial scales which cannot be resolved by tide gauges or reference mission altimetry data.

But variance reduction statistics are only meaningful to the extent that the sampling is sufficient to decorrelate signals at the tidal frequencies of interest from each other and from non-tidal signals. For example, when data from altimeters flying in non-repeating or long-repeat orbits are considered, satisfying this constraint requires a degree of spatial averaging which largely obviates the advantages of using these data. To overcome this difficulty, surface drifter data are investigated for comparisons of tidal velocity. One novelty of this approach is that the velocity is, in essense, a measure of the tidal SSH slope and friction in the models, so the comparisons may provide more validation of the model dynamics than the SSH alone. A large quantity of drifter data with one-hour sampling is available, which reduces the concern with long-period tidal aliases, and their correlations; however, the nature of the Lagrangian drifter measurements requires consideration of effects absent in Eulerian data.

Our path through this material is as follows. First, the main attributes of the tide models and their evolutions with respect to previous versions are reviewed. Following a brief overview of sampling issues relevant to explained variance statistics, the models are compared to altimeter data, emphasizing comparisons in shallow water, deep water, and high latitudes. After this, a large dataset of drifter-derived velocities is used to evaluate tidal current predictions, and, once again, sampling issues are reviewed. The paper finishes with a discussion of implications, and a summary of the main findings.

2. Overview of Tide Models

a. Goddard/Grenoble Ocean Tide Model, version 4.10c (GOT410)

The GOT410 model is the latest in the series of barotropic tide models developed at the Goddard Space Flight Center using the approach described in Schrama and Ray (1994). It was developed by analyzing satellite altimeter data to identify harmonic constants for the ten tidal frequencies listed in Table 1 (Ray 2013); although P_1 was determined by inference, mainly from K_1, because it was found to be more accurate than the direct estimate (Ray 2017 – Table 4). Version "c," used here, differs slightly from previous versions by accounting for tidal oscillations of the geocenter, which results in small changes to the attribution of the ocean and land components of the tides, mostly affecting K_1 and O_1 (Desai and Ray 2014).

The tide predictions obtained from GOT410 are computed from the harmonic constants available on its native grid at $(1/2)^\circ$ resolution. While this resolution is generally adequate for resolving the barotropic tides in the deep ocean, it is not intended to capture variability on continental shelves or near the coastlines.

b. TPXO Model, version 9.2-Atlas (TPXO9A)

TPXO9A is the most recent version of the global tide models in the TPXO series (Egbert et al. 1994; Egbert and Erofeeva 2002). It is constructed by assimilating altimeter data from multiple exact-repeat missions into a hydrodynamic model based on the Laplace Tidal Equations, modified to account for tidal dissipation, ocean self-attraction, and solid-earth loading. It is referred to as an "atlas" because it is obtained by combining a $(1/6)^{\circ}$-resolution global solution, TPXO9.1, with thirty separate $(1/30)^{\circ}$-resolution regional solutions for coastal areas, including the Arctic and Antarctic. All of the TPXO9A patches were obtained using consistent bathymetry and boundary conditions from the TPXO9.1 base solution, and they assimilate nearly all available exact-repeat mission altimetry and most available coastal tide gauge data. The resolution of the combined TPXO9A atlas is a uniform $(1/30)^{\circ}$.

The TPXO9A tidal estimates include several tides which are not part of the GOT410 model, namely, $2N_2$, MN_4, and MS_4 (Table 1). These tides are part of the FES2014 model, below, to which they may be compared. Although these tides are generally small, they do contribute to sea level predictions at the millimeter-, and sometimes larger-, scale, especially near the coast. Note that the TPXO9A model does not include the S_1 tide.

c. FES2014 Model

The FES2014 model is the latest version of the Finite Element Solution (FES) tide model (Lyard et al. 2006). Like the models in the TPXO series, the FES models assimilate essentially all available exact repeat mission altimeter data into a hydrodynamic model based on the Laplace Tidal Equations, and FES2014 also assimilates most available coastal and deep ocean data from tide gauges and bottom pressure recorders. Compared to FES2012, the hydrodynamic component of FES2014 includes refinements to the finite-element grid and it incorporates regional bathymetric data, both aimed at improving coastal tides. The native resolution of the FES2014 finite-elements ranges from roughly 5 km to 80 km; however, outputs are distributed on a uniform $(1/16)^\circ$ grid which is used here.

FES2014 contains estimates for more tidal frequencies than either the GOT410 or TPXO9A models, including several long-period tides, minor diurnal and semidiurnal tides, and nonlinear overtides (Table 1). The inclusion of these tides in the tide prediction software, and tides inferred from a smooth admittance function, leads to superior performance of FES2014 predictions when all the tides are summed. It is challenging to incorporate all of these tides in an assessment evaluation, though, since most sources of validation data are not capable of resolving all the components. For this reason, the present efforts are restricted to evaluating the solutions at the 8 primary, plus one overtide, frequencies, M_2, S_2, N_2, K_2, K_1, O_1, P_1, Q_1, and M_4.

3. Comparisons with Altimeter-derived Sea Level Data

The three models present very similar estimates of SSH harmonic constants. The co-tidal maps of the individual constituents conform closely to each other and to previous estimates (not shown). In order to evaluate the evolution of the models with respect to previous versions, the tide models are compared to independent data from satellite altimeters which were not used in the tide model development.

These missions, listed in Table 2, occupy orbits with long ground track repeat periods which were not optimized for tidal analysis. Rather than directly comparing harmonic constants, we instead look at the variance reduction as a measure of goodness-of-fit. Maps of residual variance reduction, shown later, are computed from all available orbit cycles. For convenience, data from the Jason-1/c, Jason-2/c, and Jason-2/d geodetic missions shall be referred to as Jason/GM data.

a. Preliminaries: variance reduction for assessment of tide models

The ability to identify harmonic constants from a time series depends on its length, which sets the minimum resolvable frequency separation, the Rayleigh bandwidth (Foreman 1977; Foreman et al. 2009). Although the amplitude and phase of a sinusoid are two independent parameters, if the length of a record is too short, any measurement error, non-tidal signal, or un-modeled tidal signal will project onto the resolved frequencies and lead to very poor estimates (Cherniawsky et al. 2001).

A simple example illustrates that the same considerations apply for the interpretation of residual variance. Suppose the sea level observations are represented by a linear combination of two tidal components and noise,

$$h(t) = a_1 \cos(\omega_1 t - \phi_1) + a_2 \cos(\omega_2 t - \phi_2) + \varepsilon(t), \tag{1}$$

where (a_1, a_2), (ϕ_1, ϕ_2), and (ω_1, ω_2), are the amplitudes, phases, and frequencies of the tides, and ε is the noise. The tide prediction to be validated, $\hat{h}$, is given in terms of the amplitude, $\hat{a}_1$, and phase, $\hat{\phi}_1$, of the first component,

$$\hat{h}(t) = \hat{a}_1 \cos(\omega_1 t - \hat{\phi}_1). \tag{2}$$

The variance reduction associated with this tidal prediction is,

$$r = \langle h^2 \rangle - \langle (h - \hat{h})^2 \rangle, \tag{3}$$

where $\langle x \rangle$ is the average of $x(t)$ at the measurement times. If the model is exactly correct, $\hat{a}_1 = a_1$ and $\hat{\phi}_1 = \phi_1$, then the variance reduction is,

$$r = \frac{1}{2} a_1^2 + a_1 a_2 \langle \cos((\omega_1 - \omega_2)t + \phi_1 - \phi_2) \rangle, \tag{4}$$

where it has been assumed that the noise has zero mean and is uncorrelated with the tidal signals, and the time series is long enough to approximate $\langle \cos^2(\omega_i t - \phi_i) \rangle = \frac{1}{2}$ and $\langle \cos[(\omega_1 + \omega_2)t - (\phi_1 +$

$\phi_2)]\rangle = 0$. The variance reduction associated with the partial tide, $\frac{1}{2}a_1^2$, is positive; however, if the sampling is not sufficient to decorrelate the tides, the second term can be of either sign and even dominate the first depending on the size of a_2. The conditions to minimize the second term are analogous to the Rayleigh criterion, $T > |\omega_1 - \omega_2|^{-1}$, for separating closely-spaced frequencies. Variance reduction statistics for a given tidal frequency are subject to interference from nearby tidal frequencies, similar to harmonic analysis.

Applying the above analysis to long-repeat orbit and drifting orbit altimeter missions is complicated because useful variance reduction statistics must be computed from a combination of temporal averages (over a range of dates or set of orbit cycles) and spatial averages (over a range of latitudes and longitudes or set of orbit passes). Zaron (2018) considered this issue for harmonic analysis of CryoSat-2 data, and the topic is reconsidered here from the perspective of variance reduction.

Figure 1 illustrates the sampling in time and longitude of the CryoSat-2 mission, which occupies a 368-day long-repeat orbit. Within one degree of longitude of a given observation (x-axis) subsequent observations occur at intervals of 2-days, 13.5-days, 15.5-days, 26.9-days, etc. (open circles). The measurements sample the tides at various phases and enable the separation of tidal frequencies by harmonic analysis or, equivalently, the sampling decorrelates the sinusoids and enable the determination of variance reduction statistics. The figure indicates the synodic periods of the M_2/S_2 frequency pair, indicating the length of record needed to separate the two frequencies (in red, indicated graphically with the dashed lines terminated in a filled circle). In this case, the sample approximately 2 days after the initial pass, about 0.9° to the east, decorrelates the M_2/S_2 pair after only 15 days. Unfortunately, the usefulness of the 15 day synodic period for data within $\pm 1°$ longitude bins is less than might be expected, since the high inclination CryoSat-2 orbit plane precesses slowly around the earth, it takes nearly a year to sample enough phases of S_2 to reliably decorrelate it from M_2 (cf., 324 day and 332 day synodic periods associated with the 15.5-day and 42.4-day subcycles, respectively). In other words, the subsequent samples with 2-day offsets at nearby longitudes do not efficiently sample the phases of S_2.

The impact of combining spatial averaging with the temporal sampling is illustrated in Figure 2 using data from the La Jolla, CA, tide gauge. First, the longest gap-free record from the station, 21-years from 1976 to 1997, is harmonically analyzed to determine harmonic constants. The M_2 tide is predicted using the known harmonic constant, and the variance reduction is computed for different length of record (LOR), denoted τ, according to the sampling of the CryoSat-2 mission (Figure 2a) within longitude bins of $\pm(1/4)^\circ$ (black), $\pm(1/2)^\circ$ (green), and $\pm 1^\circ$ (red). The theoretical maximum variance reduction is known from the given harmonic constant, $(A_{M2}^2/2)$, but explained variance oscillates, depending on the size of the longitude bins defining the subsampling. Even for $\pm 1^\circ$ bins, the variance reduction statistic is not accurate for $\tau < 10$yr.

The same comparison has been made using the Jason-1/c sampling pattern, which has a 407-day orbit repeat period (Figure 2b). The orbit is a small perturbation to the TOPEX/Poseidon reference mission orbit, and it has more favorable tidal sampling properties. In this case, the variance reduction statistic is stable for $\tau > 3$yr, for longitude bins of $\pm 0.5^\circ$ and larger.

The conclusion from these examples is that variance reduction statistics computed with long-repeat orbit altimeter data do not have any special value for diagnosing tide model differences near the coastlines, above that afforded by the exact-repeat reference missions. While variance reduction provides a direct metric of model performance, a criterion analogous to the Rayleigh criterion dictates that a substantial degree of spatial averaging is necessary to reduce the sampling errors. In spite of this, the data from the long-repeat missions are independent of the data assimilated into the tide models, so they provide a useful metric of performance, but not at the spatial resolution suggested by their close ground track spacing.

b. Results of altimeter comparisons

Table 3 lists the root-mean-square of the residual SSH after removing tide predictions provided by the GOT410, TXPO9A, and FES2014 tide models. The results are presented separately for observations in water shallower and deeper than 1000 m, as well as separately for the Arctic and Antarctic regions (poleward of $\pm 66°$) with CryoSat-2. To avoid the influence of differences in prediction software and the treatment of minor constituents, the prediction is computed as the sum of the eight major tidal constituents (M_2, S_2, N_2, K_2, K_1, O_1, P_1, and Q_1). The comparisons utilize the predictions for the geocentric tide, which involves a model for the load tide, i.e., the vertical tidal movement of the sea floor, and the same GOT4.10c load tide model is used in all cases. To insure the results are as comparable as possible with Stammer et al. (2014), corrections for equilibrium long-period tides (Cartwright and Taylor 1971; Cartwright and Edden 1973) are also applied to the altimetry prior to comparison with the tide models.

The results are consistent with an overall convergence of the results at the level of precision that can be obtained with residual variance comparisons, and an incremental improvement over similar results tabulated in Stammer et al. (2014). One noteworthy aspect of the new models is the now-identical level of performance in shallow and deep water with respect to the Jason/GM dataset, even though the overall SSH signal is much larger in shallow water. The standard error, σ_{sub}, has been estimated by computing the residual errors over independent subsamples of the full data records, using 100-day subrecords for Jason/GM and 1-year subrecords for CryoSat-2, as described in more detail in the Appendix. The standard error thus accounts for reduced degrees of freedom associated with the temporal and spatial correlations of SSH at periods shorter than about 100 days, caused by processes such as storm surges, variable boundary currents, and mesoscale eddies.

Although the differences between the residual variance among the models are smaller than the standard errors in Table 3, it would be incorrect to conclude that the model performances are identical or that the data are insufficient for distinguishing their relative performance. Because the same SSH data are used to evaluate each model, sample errors for the residual variances are correlated between the models, and σ_{sub} is not an appropriate measure of the sample error of residual variance *differences* between models.

Table 4 reports the statistics concerning the pairwise differences in residual variance. The standard error estimate for the variance difference, σ_r^2, is computed using the same sub-records as described above, but it is significantly smaller than σ_{sub}^2 because the correlated component of the sample variability is now absent. A second estimate for sample error, σ_d^2, attributed to the unknown and unresolved minor tidal constituents (analogous to the role of S_2 for the example shown in Figure 2), is found to be negligible compared to σ_r^2. The reader is referred to the Appendix for details about the computation of error estimates.

Considering the performance in shallow water, for example, Table 4 indicates that the residual variance of GOT410 is 1.4 cm^2 larger than that of TPXO9A, which suggests that TPXO9A performs better when measured with respect to Jason/GM data. And a larger difference of 9 cm^2 is evident when compared with CryoSat-2 data. Based on the results in the Table, the FES2014 model has the lowest residual variance when averaged over the domains indicated. The performance of the tide models in deep water, where all the differences are 0.25 cm^2 or less, corresponding to 0.5 cm root-mean-square, are likely to be equivalent for all practical purposes; however, the differences are unambiguously larger than the standard errors in all cases. In shallow water and polar regions the differences between the models may be of practical significance, but the performance varies greatly within these regions and one model or the other might be better depending on the region of application.

Maps of the differences in SLA residual variance with respect to the Jason/GM and CryoSat-2 data are shown in Figures 3 and 4, where the data have been binned at 2-degree resolution (corresponding to the $\pm1^\circ$ averaging indicated in Figure 1). FES2014 is used as the reference to which GOT410 and TPXO9A are compared, so negative values in the maps indicate that FES2014 performs better than the model with which it is compared. As demonstrated above, the variance reduction statistics invariably contain spatially-correlated sampling errors, dependent on the errors in the resolved tides (M_2, S_2, N_2, K_2, K_1, O_1, P_1, and Q_1) as well as the unresolved tides. Table 4 indicates very small differences in model performance in the deep ocean, and this is evident in the maps.

In the deep ocean, the differences between the FES2014 and GOT410 do not appear to be significant, as regions with the largest differences, e.g., near Antarctica, depend on whether Jason/GM or CryoSat-2 data are used for comparison (Figure 3 versus 4). Less ambiguous differences among the models are evident near the coastlines, consistent with the slight advantage of FES2014 noted in Table 4. The data are generally in agreement that FES2014 reduces sea level variance more than GOT410, as would be expected from the coarse grid of the latter. In a few locations, e.g., near New Caledonia (20°S,160°E), GOT410 and TPXO9A appear to have an advantage.

In conclusion, Jason/GM and CryoSat-2 data enable an assessment of tide model skill at more locations than could be obtained from comparison with other non-assimilated sea level data, e.g., without the gaps between tracks as would occur with reference-mission altimetry. The results provide a useful description of model performance and show that, on average, FES2014 provides better predictions of the 8 major tides in shallow water and at high-latitudes; although, there are important regional differences where either of the other models are more accurate. The utility of the altimetry is nonetheless limited by the spatial averaging required to reduce interference with correlated signals at nearby (aliased) tidal frequencies. The extent of spatial averaging needed to obtain stable results (2° bins here) is so large that it diminishes the ability to identify small spatial-scale errors near the coastline, and this, in part, motivates the use of drifter-derived current data for evaluating the models, below.

4. Comparisons with Surface Drifters

The Global Drifter Program (GDP) is an array of satellite-tracked drifting surface buoys from which are derived estimates of 15-m velocity along the buoys' trajectories (Lumpkin and Pazos 2007; Lumpkin et al. 2017). The complete historical GDP dataset is available as a drifter position and velocity product at 6-hour intervals (Lumpkin and Centurioni 2019). For this study a subset at hourly resolution is used (Elipot et al. 2016), version 1.02, which comprises 144 million tuples of location and velocity estimates from 15,329 individual drifter trajectories, totalling 80,765 contiguous segments. The hourly dataset contains drogued and undrogued drifters tracked by both the Argos positioning system and the Global Positioning System (GPS). Compared to the 6-hourly product, the hourly product is a dramatic improvement in the time resolution of the data, reducing data gaps and providing considerably better resolution of oceanic variability than was previously possible (e.g. Elipot and Lumpkin 2008; Elipot et al. 2010; Lumpkin and Elipot 2010; Elipot et al. 2016).

Previous studies have used GDP-derived velocities to map tidal currents. For example, Poulain and Centurioni (2015) used GDP data, interpolated to hourly time steps by kriging, to estimate tidal surface currents over the world oceans, using both drogued and undrogued drifter data. Their approach involved the estimation of tidal harmonic constants from 15-day segments of drifter trajectories, using least-squares harmonic analysis to identify the 4 largest tides, M_2, S_2, K_1, and O_1, coupled with inference of 4 smaller tides, P_1, Q_1, N_2, and K_2 (inference relations provided by the TPXO tide model). They mapped the harmonic constants at 2° resolution to resolve the barotropic tide and some coastal tides. Among the findings of Poulain and Centurioni (2015), it is noteworthy that M_2 tidal currents exceed 5 $\mathrm{cm\,s^{-1}}$ in nearly all shallow seas and coastal areas, and currents in excess of 40 $\mathrm{cm\,s^{-1}}$ are present in numerous areas near the continental shelves. Other studies with GDP data include Kodaira et al. (2016) and Zaron and Ray (2017), in which GDP-derived velocities were compared with models for the baroclinic tide using harmonic constants and variance reduction statistics, respectively.

a. Preliminaries: Lagrangian vs. Eulerian Estimates of Tidal Currents

Similar to the SSH data considered above, the GDP data may be used to either directly provide the harmonic constants of interest, or variance reduction statistics may be used. Unlike the altimetry, the temporal sampling of which aliases the tides to long periods, the GDP temporal sampling is hourly, so the usual Rayleigh criterion provides a guide to choosing which harmonic constants are identifiable from a given length of record. The issue which complicates the use of drifter data is related to the Lagrangian character of the observations. A typical drifter trajectory traverses a path along which both the tidal and non-tidal currents can vary strongly, even within a relatively small geographic region (Crawford et al. 1998).

Simulations of Lagrangian sampling through a field of tidal currents have been examined to illustrate these features of drifter data. Two GDP drifter trajectories are shown in Figure 5a, each consisting of approximately 72 days of hourly observations of drifter location and velocity. One drifter trajectory is confined to deep water (white line, depth $>$ 3000 m) while the other is on the continental shelf and slope (red line, depth $<$ 3000 m). The M_2, S_2, K_1, and O_1 components of the TPXO9A model have been used to predict tidal currents along each trajectory, and the power spectra of these predictions sampled along the drifter paths are shown in Figures 5b and 5c (thin red lines). The spectra are computed using Welch's method by averaging periodograms estimated from 15-day subrecords, so the frequency resolution for the deep and shallow drifters is identical.

Several distinct processes contribute to the character of the tidal spectra. For reference, the idealized line spectra of Eulerian currents evaluated from the harmonic constants at the mid-point of each trajectory are shown (heavy black lines). These idealized spectra would be broadened by the effects of finite record length, windowing, and averaging of Welch's method (red shading shown for M_2 only); but, the tidal peaks would be unambiguous for Eulerian data. The broadened spectrum of the M_2 currents along the trajectory (gray filled region) shows the extent to which the nearby S_2 line would be obscured because of the Lagrangian sampling. The spectrum from the deep water trajectory (Figure 5b) is essentially the same as would be obtained from Eulerian data because the barotropic tidal currents are nearly homogenous in the deep water. In contrast, the spectrum from the shallow water trajectory (Figure 5c) is considerably broadened, with variance from the peaks overlapping.

The broadening of the Lagrangian spectra can be interpreted as a consequence of the convolution theorem. Using a Taylor series expansion, the current along the trajectory, $u(X(t),t)$, can be related to the Eulerian current at a point along the trajectory, X_o, via

$$u(X(t),t) = u(X_o,t) + \nabla u(X_o,t) \cdot (X(t) - X_o). \tag{5}$$

Therefore, the Fourier transforms of the Lagrangian, $u(X(t),t)$, and Eulerian, $u(X_o,t)$, currents differ by a term which is the convolution of the transform of the spatial gradient of the Eulerian current, $\nabla u(X_o,t)$, and the transform of the position vector, $X(t) - X_o$, the integral of the velocity. Another, equally descriptive, explanation for the spectral broadening is the spatial non-stationarity of the tidal current. Tidal currents on the continental shelf vary strongly along the Lagrangian trajectories, particularly near the shelf break and near islands.

This example illustrates the tradeoff between frequency resolution (record length) and spatial resolution when Lagrangian currents are used for identification of tidal currents. It suggests that it is better to analyze the GDP data within spatial bins near the continental shelves, treating it like Eulerian rather than Lagrangian data. The harmonic constants, or variance reduction statistics, can then be determined from time series formed from spatially-binned data. The unknown spatial variability within the bins would then contribute to sampling error.

To examine this possibility, Figure 6 illustrates characteristics of the M_2 tidal current ellipses on the continental shelf and in the deep ocean, within the same region shown in Figure 5. The GDP currents within each region have been subjected to conventional least-squares harmonic analysis by binning the data and treating them as unevenly-sampled time series ($N = 2662$ observations in the shallow region and $N = 1661$ in the deep region). The quantity of data are sufficient to stably identify the ellipse parameters of the M_2, S_2, K_1, and O_1 tides with a formal error of about 0.2 $\mathrm{cm\,s^{-1}}$ for the major axis (Cherniawsky et al. 2001). Note that the magnitude of the currents is very different on the shelf and in the deep water, 20 $\mathrm{cm\,s^{-1}}$ and 4 $\mathrm{cm\,s^{-1}}$.

Tidal velocity ellipses from the barotropic FES2014 and TPXO9A models are shown, as well as the baroclinic High Resolution Empirical Tide (HRET) model (Zaron 2019), where each ellipse corresponds to a single point along the GDP trajectories in either the shallow (panels b-d) or deep (panels e-g) water. In the shallow region, the FES2014 ellipses are more similar to the observed ellipse than TPXO9A ellipses; however, in the deep water, the barotropic current ellipses are much too small. Comparison of the observed current ellipse with HRET in deep water indicates that the observations are likely dominated by the baroclinic — rather than barotropic — currents. The spread in the model ellipses (in gray) is a consequence of the spatial variability within the regions, and it is much larger than the formal error estimate of the measured ellipse parameters (computed, but not shown). The validity of the formal error estimate depends on assumptions about the probability distribution and power spectrum of the residual currents, both of which are difficult to characterize from the unevenly sampled time series. The difficulty this creates is obvious: without adequate error estimates for the observed harmonic constants, it is impossible to assess the significance of model deviations from the observations.

From the above discussion, we conjecture that variance reduction statistics should be useful for assessing the quality of the barotropic tidal current predictions on the continental shelves, where barotropic currents may be expected to be larger than the baroclinic currents. In contrast, in the deep ocean, where baroclinic tidal currents can be much larger than the barotropic currents, the variance reduction statistics will be less useful, because the barotropic predictions (and their errors) are correlated with the larger baroclinic currents. To discern differences in variance reduction associated with the barotropic tide in the deep ocean, it would be necessary to average the statistics over an area larger than the wavelength of the baroclinic tide.

b. Results of Comparisons with Surface Drifters

The above conjecture about the interference of baroclinic tidal currents is confirmed by maps of variance reduction statistics for the M_2 and K_1 tidal currents from the FES2014 model (Figure 7). Overall the variance reduction is positive-valued (red), particularly in areas with strong barotropic tidal currents, such as in the Bering Sea, off the U.S. East coast, in the Yellow Sea, and many other areas, which means that the FES2014 model does predict some of the velocity variance captured by the drifters. However, near sites of strong internal tide generation there are oscillations in the variance reduction of $\pm 10\ (\mathrm{cm\,s^{-1}})^2$. For M_2 these occur between the Aleutian Islands and the Hawaiian Ridge, near the Mariana Arc, and in the Tasman Sea; for K_1 the most prominent oscillations are near Luzon Strait. In this and subsequent figures, bins in which the density of observations is less than 20 per $(50\ \mathrm{km})^2$ are indicated with light gray.

In addition to the oscillations caused by the baroclinic tide, there are also spatial modulations in the noise level of the variance reduction. This is presumably related to the strength of the tidal currents and the number of GDP observations within each spatial bin. Figure 8 shows a map of the spatial density of GDP observations, expressed as the number of observations per $(50\ \text{km})^2$ bin. Elevated noise in the variance reduction statistics along the equator (Figure 7) is associated with the lower density of GDP observations.

The spatial density of GDP data is sufficiently large that it is useful for evaluating the tide models throughout much of the global oceans (Figure 8). Because the number of observations increases with the size of the region considered, there is a relationship between the spatial scale and the frequency resolution of the tides. In other words, to distinguish the model performance at two frequencies, say, M_2 and S_2, it is necessary to evaluate the residual variance from a sufficiently large number of hourly observations to decorrelate the tides at these frequencies. The Rayleigh criterion suggests that 350 hourly measurements (15 days) are sufficient to resolve (O_1, K_1, M_2, S_2), and 4382 hourly measurements (183 days) are sufficient to separate (Q_1, O_1, P_1, K_1, N_2, M_2, S_2, K_2). When the number density of Figure 8 is integrated over regions of different sizes (not shown), the 8 main constituents can be distinguished in mid-gyre within patches of $(200\ \text{km})^2$ or larger, but, near the coast, larger patches must be used.

When the variance reduction is averaged over 2° bins, the baroclinic tides are averaged out, and the sample fluctations of the variance reduction statistics are improved within the bins (Figure 9). The small variance reduction associated with the barotropic M_2 tide over the Mid-Atlantic Ridge is stably estimated (note the much narrower colorscale used in Figure 9 as compared with Figure 7). There is an unusual strip of elevated variance reduction at the inertial latitude of the K_1 tide, particularly noticeable near 30°S in the Indian Ocean. Overall, though, the largest variance reductions exceeding 10's of $(\mathrm{cm\,s^{-1}})^2$ are associated with the continental shelves and regions of strong barotropic tides, consistent with observations of Poulain and Centurioni (2015).

A global comparison of the GOT410, TPXO9A, and FES2014 models using GDP currents is summarized in Tables 5 and 6, using predictions for the sum of the 8 major tides. Two statistical measures of the models are shown: the root-mean-square (RMS) and the median absolute residual after removing the predicted tidal currents (*u*- and *v*-component residuals are combined to form a residual speed for both RMS and median statistics). Note that only three of the single-model residuals differ by more than the estimated standard error from the *Signal* in Table 5, all in shallow water (depth less than 1000 m). FES2014 exhibits a slight reduction of both RMS and median statistics; while GOT410 exhibits a RMS residual larger than the *Signal*, indicating that it is not useful for predicting tidal currents in shallow water. The latter result is expected, though, since the coarse-resolution GOT410 model was not intended for use in shallow water or near the coastlines (personal communication, Richard D. Ray 2020).

As explained in the Appendix, the sample errors are dominated by non-tidal variability, so the differences in residual variance between the models can be estimated more precisely than the residual variance of any single model. The differences in model performance, summarized in Table 6, are much larger than the sampling errors, but the large difference in the mean-square versus median-square statistics highlights the limited utility of these summary statistics. Although the models can be ranked in performance from GOT410, to TPXO9A, to FES2014 based on the global statistics in Table 6, the largest differences between the models occur at small scales and the best model for any specific application depends on location. Furthermore, the GDP sampling is poor in the Arctic Ocean, in the Hudson Bay and around much of Antarctica, and so the GDP provide no insight into the model differences in these regions where the altimeter data indicated performance differences.

Maps of differences in residual variance are provided for the comparison of FES2014 and TPXO9A in Figure 10. Large-amplitude differences between the models are resolved at small spatial scales where differences in excess of 25 $(\mathrm{cm\,s^{-2}})^2$ are evident along the northeast coast of South America, in the western East China Sea, and in other areas (Figure 10a). Consistent with Table 6, overall FES2014 has a smaller residual variance than TPXO9A in the coastal areas (blue shading), but numerous sites are present where the opposite is true (red shading). Smaller differences in residual variance can be resolved when the data are compared in larger bins (Figure 10b) and show that TPXO9A has smaller residual variance over much of the Mid-Atlantic Ridge, at a number of spots in the Western Pacific, e.g., near 10°N–160°E, and other regions, even though the summary statistics indicate smaller average residual for FES2014.

To illustrate the differences between the models at smaller scales, the map in Figure 11 shows an enlargement of Figure 10a, where a logarithmic colorscale is used. Blue-green indicates negative values, which correspond to smaller FES2014 residuals; while pink-red indicates positive values, which correspond to smaller TPXO9A residuals. As noted previously, the FES2014 residual is smaller over much of the East and South China Seas, the Gulf of Thailand, and the shelf adjoining the Andaman Sea and Malacca Strait. Figure 11 also illustrates geographical patterns related to the internal tide, barely visible around Luzon Strait (20°N–122°E), as well as large-amplitude anomalies near the coast. The range of current variance shown, $\pm 10^4$ $(\mathrm{cm\,s^{-1}})^2$, corresponds to a root-mean-square speed difference of 1 $\mathrm{m\,s^{-1}}$.

The red pixels at 22°S–150°E, on the northeast coast of Queensland, Australia, exhibit an apparent degradation of FES2014 relative to TPXO9A. The site is nearby McEwin Islet, where the tidal range exceeds 8 m (Australian Bureau of Meteorology 2020), near a funnel shaped estuary. This site is shoreward of a portion of the Great Barrier Reef, where the FES2014 model performance is unambiguously superior to the TPXO9A (blue pixels at 21°S–150°E, Figure 11), and it exemplifies the large-amplitude, but rare, differences between the models that differentiates the mean-square and median statistics in Table 6.

The individual drifter trajectories (Figure 12a) are sparse, partly because the swift coastal currents move the drifters quickly through the region. Examination of the residual currents, after de-tiding with FES2014 or TPXO9A, reveals that the performance differences in this area are caused by just a few trajectories in their closest approach to the coast, where the tidal ellipses predicted by the two models differ greatly (Figures 12b and c). The highlighted trajectory passes close to Leicester Island (22.25°S–150.5°E) where FES2014 predicts M_2 currents of nearly 200 $\mathrm{cm\,s^{-1}}$ (Figure 12b), but the tidal currents inferred by harmonic analysis of the trajectory shown are only about 50 $\mathrm{cm\,s^{-1}}$, which are comparable to the TPXO9A currents (Figure 12c). In this case the FES2014-predicted currents, while plausible, are much larger than either the observed or TPXO9A currents.

The comparison of the tide models along individual trajectories highlights the incredible level of detail afforded by the drifter data, but it also exposes the limitation of global statistical summaries to usefully describe differences in the models' performance. We had originally conjectured that the data from this buoy were spurious, but, upon examination, it does not appear to have malfunctioned or become grounded. Thus, large differences between the predicted and observed currents exist very close to the coastline, and one may wish to consider comparisons along individual trajectories when evaluating the tide models for use at particular sites. In any case, the large differences provide guidance as to where further model refinement may be warranted.

5. Discussion

Previous comparisons of barotropic ocean tide models using long-repeat/geodetic mission altimetry found small differences in performance in shallow areas (Stammer et al. 2014) which are also found here, with better precision, using the longer records now available. In deep water the differences between the models are so small that they are hardly of importance to most applications, but they may point to the need to consider the different groups' load tide models in the future, instead of uniformly employing the GOT4.10 load tide as was done here. The high-latitude data from CryoSat-2 indicate that TPXO9A performs significantly worse than either GOT410 or FES2014 poleward of 66°, but it would be useful to revisit the high-latitude comparisons of Stammer et al. (2014) using the most recent GRACE, GRACE-FO, and IceSat-2 data.

The comparisons of GDP data with tidal currents from the models leads to results similar to the altimeter-based comparisons. The GDP comparisons highlight the importance of strong, but small-scale, tidal currents, which are represented differently in summary statistics based on either mean versus median statistics.

To contextualize the differences between the models, it is useful to compare the variance differences to the overall tidal signals. For example, the SSH variance differences in shallow water range from about 1 to 16 cm^2 (Table 4), while the tidal signal variance is about 2.7×10^3 cm^2 (computed from the difference of the *Signal* squared minus the residual squared in Table 3). Thus, the differences in SSH residuals in shallow water amount to less than 0.5% even at their largest. The high-latitude differences are larger, e.g., about 15 cm^2 out of 1035 cm^2, or about 1.5%. When these ratios are interpreted as RMS ratios, the latter is somewhat greater than 3 cm out of 30 cm, a relatively large percentage compared to the accuracy in deep water at lower latitudes. The comparison of tidal currents is less clear-cut because the median statistics are so much smaller than the RMS statistics (Table 5). Comparing the total signal, 30.2 $\mathrm{cm\,s^{-1}}$, with the best-case residual, 28.4 $\mathrm{cm\,s^{-1}}$, suggests that a smaller fraction of shallow current variance can be attributed to barotropic tides, $((30.2)^2 - (28.4)^2)/(30.2)^2$, about 10%. Thus, even the small differences in residual variance represent large fractions of the tidal variance.

The comparisons of the tide models has focussed on the 8 main tides (Q_1, O_1, P_1, K_1, N_2, M_2, S_2, K_2) because these tides are common to all models considered in this study, but the GDP data can also be used to evaluate predictions for other tides included in the individual models. For example, the variance reduction for M_4 derived from the FES2014 model is shown in Figure 13. In this case, many coastal areas exhibit negative values of the variance reduction (blue-green hues), indicating that tidal current predictions for this overtide are not yet usefully accurate. Comparisons with M_4 predicted by TPXO9A are similarly ambiguous (not shown). While the models differ in detail, neither is capable of making reliable current predictions for M_4 or the other minor tides. Readers interested in graphical comparisons of each of the tides predicted by both FES2014 and TPXO9A are referred to the Supplementary Data which accompanies this manuscript.

It is interesting to consider the dynamical significance of tidal current predictions. For the data assimilative tide models, the validation of currents is a stronger test of the models than the validation of sea surface height predictions. The latter are strongly constrained by the altimetry which they assimilate; however, the currents depend on the gradient of the surface height (pressure) as well as other terms in the force-balance, such as the bottom frictional stress, the baroclinic pressure gradient, and the Lagrangian acceleration. The influence of the baroclinic pressure gradients is manifest in deep water near known sources of the internal tide (Figure 7). The predominant balance is nonetheless between the Eulerian acceleration and the horizontal pressure gradient, so improvements in the prediction of tidal currents ought to correspond to improvements in the pressure gradient, or the sea surface slope. The variance in these quantities is approximately related by, $s^2 = (\omega/g)^2 \Delta u^2$, where s^2 is the variance of the surface slope, ω is the tidal frequency, g is the acceleration of gravity, and Δu^2 is the current variance. For the M_2 tide, the $10^2 (\mathrm{cm\,s^{-1}})^2$ variance reduction typical of the continental shelf (Figure 11) corresponds to 1 micro-radian in the RMS tidal slope. The tidal predictions provided by FES2014 in the blue areas, and provided by TPXO9A in the red areas, ought to enable better coastal tide corrections of sea surface slope, which is used to map short-wavelength undulations of the geoid (Sandwell et al. 2014).

6. Summary

This paper has sought to compare the latest versions of data-constrained models of the barotropic ocean tides, namely, FES2014, TPXO9A, and GOT410, which are widely used in tidal predictions of sea surface height, boundary conditions of regional ocean forecast models, and other applications. The quality of the models, as measured either by their agreement with harmonic constants inferred from in situ data or the skill of their tide predictions, has converged to a level in the deep ocean that makes it difficult to distinguish them. However, in the coastal regions and on continental shelves, there are unambiguous differences in model skill. It is a challenge to quantify these differences, though, because the models utilize nearly all available precise sea level measurements in their development.

Comparisons between the models are consistent whether altimeter or drifter data are used. The finer grid of FES2014 and, presumably, the high-resolution bathymetry of this model (some of which is based on non-public data) lead to better variance reduction statistics compared to TPXO9A and GOT410 near the coasts and in shallow water. At high latitudes the TPXO9A model does not explain as much variance as either GOT410 or FES2014, but this conclusion is based solely on CryoSat-2 altimetry. The GDP data can provide ground truth for tidal currents very close to shore, but the dynamic range and variability of tidal currents is so large that median statistics may be more useful than conventional averaging. Alternately, the analysis of individual drifter records may be necessary to interpret the large differences between the models close to shore.

In the deep ocean, the accuracy of the three models is essentially indistinguishable for the 8 largest tides (M_2, S_2, N_2, K_2, K_1, O_1, P_1, and Q_1). For the smaller astronomical tides, and nonlinear overtides, the variance reduction statistics are not broadly useful for distinguishing the model performances. The sampling error and frequency resolution are insufficient to identify the differences in variance reduction associated with these small-amplitude tides.

Acknowledgments. This research was supported by NSF awards #1850961 (Zaron) and #1851166 (Elipot), and NASA award #NNX16AH88G (Zaron). This research would not have been possible without the many people who support the collection and dissemination of the publicly-available altimeter and drifter data used here, whose efforts are acknowledged. The authors also acknowledge and appreciate the tide model developers, both for providing the models and for their help using them to predict tidal elevations and currents. The careful reading and commentary provided by two anonymous reviewers contributed to the quality of this paper, and their efforts are greatly appreciated.

Data Availability Statement. The data analyzed in this study are freely available from public websites or from the creators of the respective tide models. La Jolla tide gauge data are available from the NOAA Tides and Currents website, `https://tidesandcurrents.noaa.gov`, station number 9410230. The satellite altimeter data used here were extracted from the Radar Altimeter Database System, RADS (Naeije et al. 2002), which were analyzed using its *radsstat* and *rads2grid* programs. The FES2014 model was produced by Noveltis, Legos and CLS and is distributed by Aviso+, with support from CNES (`https://www.aviso.altimetry.fr/`). TPXO9A is available from its creators, Lana Erofeeva and Gary Egbert, at `https://www.tpxo.net/global/tpxo9-atlas`; version 2 of the atlas was used here, released April 1, 2020. GOT410 is available from Richard Ray, `richard.d.ray@nasa.gov`. The hourly surface drifter data used in this study are available through the Data Assembly Center of the Global Drifter Program at `https://www.aoml.noaa.gov/phod/gdp/hourly_data.php`.

APPENDIX

Error Estimates for Sample Variance

Estimates of residual and explained variance are used throughout the text to compare tide models. This appendix reviews some basic statistics which are useful for describing the sources of error in the variance estimates. To establish the notation, represent a scalar time series, h (either SSH or a velocity component), as the sum of a signal, x, attributed to the resolved tides of interest; a signal, y, attributed to unresolved tides; and broadband noise, ε, consisting of measurement noise plus non-tidal signals, $h = x + y + \varepsilon$. For the sake of convenience, assume that each component of h has zero mean. Let x_i denote the predicted tide provided by the i-th tide model, then the residual variance is defined as $\langle (h - x_i)^2 \rangle$, and the explained variance is simply $\langle h^2 \rangle - \langle (h - x_i)^2 \rangle$, where the angle brackets denote the mean over the sample of interest, which may involve both temporal and spatial averaging.

Because the true tide, x, is unknown, it is not possible to compute the tide model error, $\langle (x - x_i)^2 \rangle$, and this is the reason why comparisons of the residual variance, $\langle (h - x_i)^2 \rangle$, are used to assess the relative performance of the models. Because the tide models are very similar, the x_i estimates are correlated with each other, and the sample errors of $\langle (h - x_i)^2 \rangle$ are not independent. Consequently, the sample errors of variance *differences*, $\Delta_{ij} = \langle (h - x_i)^2 \rangle - \langle (h - x_j)^2 \rangle$, may be much less than the sample errors of either model, $\langle (h - x_n)^2 \rangle$ for $n = i$ or $n = j$. Sample error here refers to errors which arise from non-tidal signals and random errors, ε, as well as errors which arise from unresolved tides, y. In principle both of these errors will tend to zero as the sample size is increased, but the temporal and spatial structure of the errors must be accounted for when estimating the sampling error.

The variance of h is $\langle h^2 \rangle = \langle x^2 \rangle + \langle y^2 \rangle + \langle \varepsilon^2 \rangle + 2(\langle xy \rangle + \langle x\varepsilon \rangle + \langle y\varepsilon \rangle)$. The residual variance is

$$\langle (h - x_i)^2 \rangle = \langle (x - x_i)^2 \rangle + \langle y^2 \rangle + \langle \varepsilon^2 \rangle + 2(\langle (x - x_i)y \rangle + \langle (x - x_i)\varepsilon \rangle + \langle y\varepsilon \rangle). \qquad \text{(A1)}$$

For very accurate tide models, the sample error associated with $\langle y^2 \rangle$ and $\langle \varepsilon^2 \rangle$ may be comparable to the quantity of interest, $\langle (x - x_i)^2 \rangle$, which motivates the use of the explained variance (or, the variance reduction associated with the tide prediction),

$$\langle h^2 \rangle - \langle (h - x_i)^2 \rangle = \langle x^2 \rangle - \langle (x - x_i)^2 \rangle + 2\langle x_i(y + \varepsilon) \rangle, \qquad \text{(A2)}$$

which is unaffected by sample errors of $\langle y^2 \rangle$ and $\langle \varepsilon^2 \rangle$. For model comparisons, it is useful to write out the variance difference, mentioned above, as

$$\Delta_{ij} = \langle (x - x_i)^2 \rangle - \langle (x - x_j)^2 \rangle + 2\langle (x_i - x_j)(y + \varepsilon) \rangle. \qquad \text{(A3)}$$

Different strategies are used to estimate the sample errors associated with the residual variance (A1) and residual variance differences (A3). The standard error of the quantity $\langle (h - x_i)^2 \rangle$ (A1), denoted σ_{sub}^2, is estimated from N_{sub} independent subsamples within 10-day windows for the Jason/GM (one orbit cycle) or 1-year windows for CryoSat-2. The standard error, σ_{sub} in Table 3, is the average of the standard deviation estimates, divided by the square root of the number of subsamples $N_{sub}^{1/2}$. The error estimate thus accounts for the reduced degrees of freedom associated with correlated sea-level variability within each data window (periods less than 10 days or 1 year for Jason/GM and CryoSat-2, respectively). It is provided as a reference to understand the expected error of residual sea level estimates. Note that this sampling error is primarily caused by non-tidal sea level variability; uncorrelated instrument error is effectively averaged out by taking global and regional averages of the given 1 Hz altimeter data. Also, note that a different error estimate is computed for each of the three models, and model pairs; however, only the values for either TPXO9A or TPXO9A-FES2014 are provided in the tables, since these are the values salient to distinguishing performances differences between TPXO9A and FES2014.

The term $2\langle(x_i - x_j)(y+\varepsilon)\rangle$ in the variance difference (A3) is associated with a random component, $2\langle(x_i - x_j)\varepsilon\rangle$, and a deterministic component, $2\langle(x_i - x_j)y\rangle$, where x_i and x_j are given but ε and y are unknown. The random component, denoted σ_r^2, is estimated by computing Δ_{ij} on independent data windows, as described above. The deterministic component, denoted σ_d^2, is harder to estimate because it depends on y, which may be correlated among independent data subsamples. To make an estimate of σ_d we compute $2\langle(x_i - x_j)y\rangle$ over the full data record using tidal predictions for several of the minor tides from the FES14 model in place of y (larger representatives of the minor tides are used, S_1, J_1, ν_2, and M_4). The root-sum-of-squares of these quantities is then taken as σ_d^2.

References

Australian Bureau of Meteorology, 2020: McEwin Queensland Tide Predictions. Accessed 2020-02-04, http://www.bom.gov.au/tides.

Cartwright, D. E., and A. C. Edden, 1973: Corrected tables of tidal harmonics. *Geophys. J. Roy. Astron. Soc.*, **33**, 253–264.

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

Cherniawsky, J. Y., M. G. G. Foreman, W. R. Crawford, and R. F. Henry, 2001: Ocean tides from TOPEX/Poseidon sea level data. *J. Atm. and Ocean. Tech.*, **18**, 649–664.

Crawford, W. R., J. Y. Cherniawsky, and M. G. Foreman, 1998: Rotary velocity spectra from short drifter tracks. *J. Atmos. Oceanic Technol.*, **15**, 731–740.

Desai, S. D., and R. D. Ray, 2014: Consideration of tidal geocenter variations for satellite altimeter observations of ocean tides. *Geophys. Res. Lett.*, **41**, 2454–2459.

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

Egbert, G. D., and S. Y. Erofeeva, 2002: Efficient inverse modeling of barotropic ocean tides. *J. Atm. and Ocean. Tech.*, **19**, 183–204.

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.*, **121 (5)**, 2937–2966.

Elipot, S., R. Lumpkin, and G. Prieto, 2010: Modification of inertial oscillations by the mesoscale eddy field. *J. Geophys. Res.*, **115**, C09 010.

Foreman, M. G., 1977: Manual for tidal heights analysis and prediction. Pacific Marine Science Report 77–10, Institute of Ocean Sciences, Patricia Bay, 66 pp.

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.

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.

LeProvost, C., 2001: Ocean tides. *Satellite Altimetry and Earth Sciences*, L.-L. Fu, and A. Cazenave, Eds., International Geophysics Series, Vol. 69, Academic Press, San Francisco, 267–303.

Lumpkin, R., and L. Centurioni, 2019: Global Drifter Program quality-controlled 6-hour interpolated data from ocean surface drifting buoys., accessed 2020-04-13.

Lumpkin, R., and S. Elipot, 2010: Surface drifter pair spreading in the North Atlantic. *J. Geophys. Res.*, **115**, C12 017.

Lumpkin, R., T. Özgökmen, and L. Centurioni, 2017: Advances in the Application of Surface Drifters. *Ann. Rev. Mar. Sci.*, **9 (1)**, 59–81.

Lumpkin, R., and M. Pazos, 2007: Measuring surface currents with SVP drifters: the instrument, its data, and some recent results. *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*, A. Griffa, A. D. Kirwan, A. Mariano, T. Ozgokmen, and T. Rossby, Eds., Cambridge University Press, 39–67.

Lyard, F., 1999: Data assimilation in a wave equation: a variational representer approach for the Grenoble tidal model. *J. Comput. Phys.*, **149**, 1–31.

Lyard, F., F. Lefevre, T. Letellier, and O. Francis, 2006: Modelling the global ocean tides: modern insights from FES2004. *Oc. Dyn.*, **56**, 394–415.

Naeije, M., E. Doornbos, L. Mathers, R. Scharroo, E. Schrama, and P. Visser, 2002: Radar Altimeter Database System: exploitation and extension (RADSxx). Tech. Rep. NUSP-2 report 02-06, NUSP-2 project 6.3/IS-66, Delft Institute for Earth-Oriented Space Research (DEOS), Delft, Netherlands. ISBN 90-5623-077-8.

Nelson, A., B. Arbic, E. D. Zaron, A. Savage, J. Richman, M. Buijsman, and J. Shriver, 2019: Towards realistic baroclinic semidiurnal tide nonstationarity in hydrodynamic models. *J. Geophys. Res.*, **124**, 6632–6642.

Ngodock, H. E., I. Souopgui, A. J. Wallcraft, J. G. Richman, J. F. Shriver, and B. K. Arbic, 2016: On improving the accuracy of the M_2 barotropic tides embedded in a high-resolution global ocean circulation model. *Ocean Mod.*, **97**, 19–26.

Poulain, P.-M., and L. Centurioni, 2015: Direct measurements of World Ocean tidal currents with surface drifters. *J. Geophys. Res.*, **120**, 1–18.

Ray, R. D., 1999: A global ocean tide model from Topex/Poseidon altimetry: GOT99.2. Tech. Rep. 209478, NASA Tech. Memo, Goddard Space Flight Center, Greenbelt, Maryland, 58 pp. URL https://ntrs.nasa.gov/search.jsp?R=19990089548.

Ray, R. D., 2013: Precise comparisons of bottom-pressure and altimetric ocean tides. *J. Geophys. Res.*, **118 (9)**, 4570–4584.

Ray, R. D., 2017: On tidal inference in the diurnal band. *J. Atm. and Ocean. Tech.*, **34**, 437–446.

Ray, R. D., and D. A. Byrne, 2010: Bottom pressure tides along a line in the southeast Atlantic Ocean and comparisons with satellite altimetry. *Ocean Dyn.*, **60**, 1167–1176.

Ray, R. D., and G. D. Egbert, 2017: Chapter 13: Tides and satellite altimetry. *Satellite Altimetry over Oceans and Land Surfaces*, D. Stammer, and A. Cazenave, Eds., CRC Press, 427–458.

Sandwell, D. T., R. D. Muller, W. H. Smith, E. Garcia, and R. Francis, 2014: New global marine gravity model from CryoSat-2 and Jason-1 reveals buried tectonic structure. *Science*, **346 (6205)**, 65–67.

Savage, A. C., and Coauthors, 2017: Spectral decomposition of internal gravity wave sea surface height in global models. *J. Geophys. Res.*, **122 (10)**, 7803–7821.

Schrama, E., and R. D. Ray, 1994: A preliminary tidal analysis of TOPEX/POSEIDON altimetry. *J. Geophys. Res.*, **99 (C12)**, 24 799–24 808.

Stammer, D., and Coauthors, 2014: Accuracy assessment of global barotropic ocean tide models. *Rev. Geophys.*, **52 (3)**, 243–282.

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., 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.

LIST OF TABLES

TABLE 1. Barotropic tide models considered. The Darwin symbols for the tides indicate the frequencies which are provided by all the models (in regular font; except for S_1 which is not included in TPXO9A); those provided by TPXO9A and FES2014 in parentheses; and those only provided by FES2014 in small font and parentheses.

Model	Tides	Resolution	References
GOT410	Q_1, O_1, P_1, K_1, S_1, N_2, M_2, S_2, K_2, M_4	$(1/2)^\circ$	Ray (2013)
TPXO9A	Q_1, O_1, P_1, K_1, N_2, M_2, S_2, K_2, M_4 ($2N_2$, MN_4, MS_4)	$(1/30)^\circ$	Egbert and Erofeeva (2002)
FES2014	Q_1, O_1, P_1, K_1, S_1, N_2, M_2, S_2, K_2, M_4 ($2N_2$, MN_4, MS_4) (S_a, M_m, M_f, MS_f, M_{tm}, MS_{qm}, J_1, ε_2, μ_2,ν_2, MKS_2, λ_2, L_2, T_2, R_2, M_3, N_4, S_4, M_6, M_8)	$(1/16)^\circ$	Lyard et al. (2006)

TABLE 2. Satellite missions used for SSH comparisons. Mission phase and orbit cycle definitions follow the usage in the Radar Altimeter Database System (Naeije et al. 2002).

Mission/phase	Orbit cycles	Dates	Duration [days]
Jason-1/c	382–425[a]	2012-05-07 – 2013-06-21	410
Jason-2/c	332–355	2017-07-11 – 2018-07-18	372
Jason-2/d	356–383	2018-07-25 – 2019-10-01	433
CryoSat-2	007–125[b]	2010-09-25 – 2019-12-02	3428

[a] Cycles 382–409 were used in Stammer et al. (2014).

[b] Cycles 1–40 were used in Stammer et al. (2014).

TABLE 3. Comparison of root-mean-square residual SSH (units of cm) after applying the barotropic tide correction from the three models, measured in different regions: shallow water refers to depths less than 1000 m, deep water refers to depths greater than 1000 m, Arctic refers to latitudes northward of 66°N, and Antarctic refers to latitudes southward of 66°S. The row labelled *Signal* is the SSH standard deviation before the ocean tide correction is applied. To avoid the influence of differences in prediction software and their treatment of minor constituents, the correction is computed as the sum of the eight major tidal constituents (M_2, S_2, N_2, K_2, K_1, O_1, P_1, and Q_1). To make the results as comparable as possible with Stammer et al. (2014), corrections for equilibrium long-period tides (from Cartwright and Edden 1973) and the GOT4.10c load tide (used for all models) are also applied. The standard error estimate of the residual, σ_{sub}, is inferred from the variability of the SSH residual variance over disjoint 100-day sub-records for Jason/GM and 1-year sub-records for CryoSat-2.

	Shallow		Deep		CryoSat-2	
Model	Jason/GM	CryoSat-2	Jason/GM	CryoSat-2	Arctic	Antarctic
GOT410	10.3	12.	10.2	10.	11.	7.7
TPXO9A	10.2	12.	10.2	10.	12.	8.5
FES2014	10.2	12.	10.2	10.	11.	7.7
Signal	54.4	54.	31.4	32.	34.	25.9
Std. error, σ_{sub}	0.8	2.	0.4	2.	2.	1.5

TABLE 4. Comparison of variance differences (units of cm^2) among the tide models in different regions (as in Table 3). The three models are compared pairwise, as indicated in the first column; for example, GOT410 - TPXO9A indicates the residual variance of GOT410 minus the residual variance of TPXO9A (i.e., a positive value indicates that the second model in the pair has smaller residual). Two estimates of the standard error are provided. The first estimate, σ_r^2, is inferred from the variability averaged over sub-records as in Table 3. Note that $\sigma_r^2 \ll \sigma_{sub}^2$ because the residual SSH is correlated between the models. The second estimate of the standard error, σ_d^2, is related to the correlation of the un-resolved tides with the resolved tides over the data records. Larger values indicate the degree to which the resolved (major) tides cannot be distinguished from unresolved (minor) tides due to the finite-record lengths, as explained in the Appendix. The uncertainty of the variance differences is dominated by non-tidal SSH signals rather than the un-resolved tidal signals, i.e., $\sigma_r^2 > \sigma_d^2$.

	Shallow		Deep		CryoSat-2	
Models compared	Jason/GM	CryoSat-2	Jason/GM	CryoSat-2	Arctic	Antarctic
GOT410 - TPXO9A	1.4	9.0	-0.14	-0.17	-14.9	-12.5
GOT410 - FES2014	2.7	15.7	0.03	0.09	0.6	0.3
TPXO9A - FES2014	1.3	6.7	0.17	0.25	15.5	12.8
Std. error, σ_r^2	0.1	0.7	0.01	0.05	2.0	1.3
Std. error, σ_d^2	0.06	0.03	0.002	0.001	0.1	0.3

TABLE 5. Comparisons of root-mean-square and median residual speeds (units of $\mathrm{cm\,s^{-1}}$) after applying the barotropic tide correction to GDP measured currents in shallow water (depth less than 1000 m) and deep water (depth greater than 1000 m). The row labelled *Signal* corresponds to the signal before the ocean tide correction is applied. As in Tables 3 and 4 the tidal current is computed as the sum of the eight major tidal constituents (M_2, S_2, N_2, K_2, K_1, O_1, P_1, and Q_1). The standard error estimate, σ_{sub}, is inferred from the variability of the residual variance when the GDP data is divided into 8 approximately-equal-length subrecords.

	Shallow		Deep	
Model	RMS	median	RMS	median
GOT410	39.8	25.4	29.6	25.2
TPXO9A	30.3	24.7	29.6	25.2
FES2014	28.4	24.3	29.6	25.2
Signal	30.2	26.0	29.7	25.3
Std. error, σ_{sub}	0.5	1.5	0.3	1.3

TABLE 6. Comparison of differences in residual current statistics (units of $\mathrm{cm}^2\mathrm{s}^{-2}$) among the tide models in shallow and deep water, for the same 8 constituents used in Table 5. The column labelled "mean" is the mean difference in the residual current squared between the models indicated. The column labelled "median" is the median difference in the residual current squared between the models indicated. As in Table 4, positive values indicate that the second model in the pair has a smaller residual. The standard error, σ_r^2, is inferred from the variability averaged over 8 approximately-equal-length subrecords of the GDP data.

	Shallow		Deep	
Models compared	mean	median	mean	median
GOT410 - TPXO9A	670	0.36	2.5	0.000
GOT410 - FES2014	780	0.56	3.0	0.000
TPXO9A - FES2014	110	0.21	0.5	0.000
Std. error, σ_r^2	20	0.05	0.1	0.002

LIST OF FIGURES

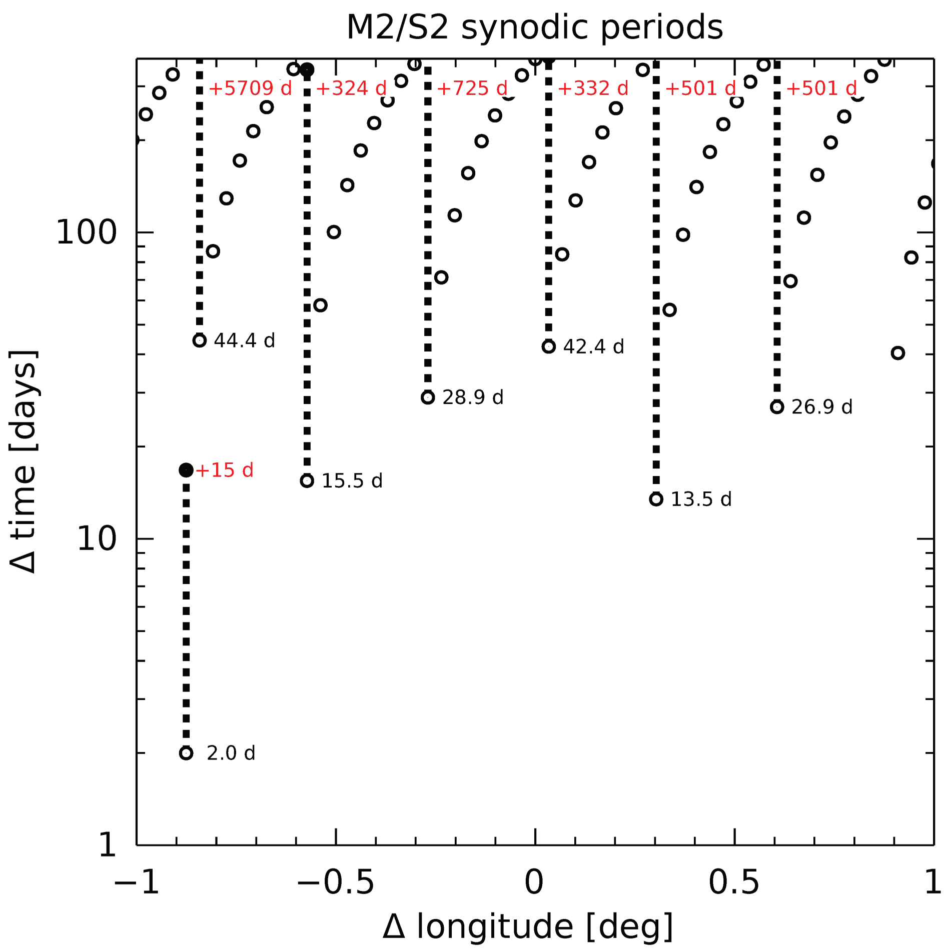

FIG. 1. The sampling of CryoSat-2 in longitude (x-axis) and time (y-axis) illustrates how the spatial binning of CryoSat-2 observations samples tidal variability at different tidal alias periods in relation to the M_2/S_2 synodic period (red). At the initial time Cryosat-2 samples the ocean at Δ longitude=0, and subsequent samples are indicated by the open circles. The number next to each open circle is the time increment from the initial sample. Thus, 2 days after the initial sample, the CryoSat-2 orbit samples about 0.9° to the west; at 13.5 days it samples 0.3° to the east; at 15.5 days it samples 0.6° to the west; etc. Associated with each of these sample intervals is a set of tidal alias periods and synodic periods, the latter shown here in red for the M_2/S_2 pair. The synodic period is the time needed for one of the frequencies in the pair to cycle through one period more than the other frequency in the pair. It is analogous to the Rayleigh period in that it is the shortest record length from which the two frequencies can be stably resolved using conventional harmonic analysis. The dashed lines terminating in the filled circles indicate the synodic period added to the sample interval to indicate the nominal length of record needed to distinguish the M_2 and S_2 waves.

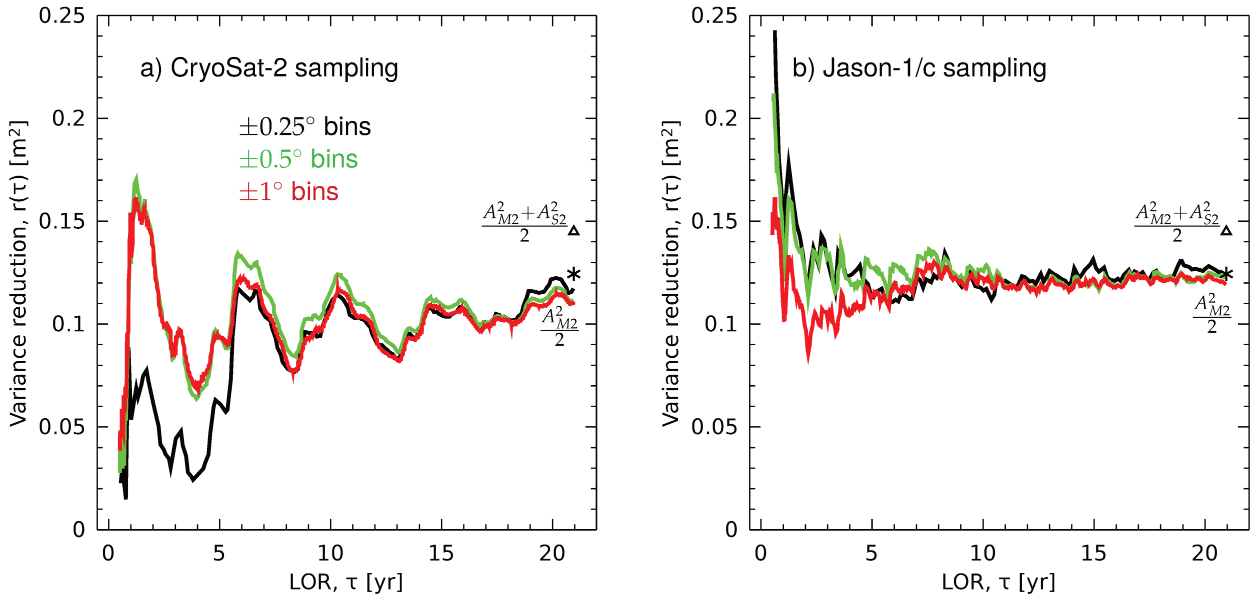

FIG. 2. Variance reduction statistics as a function of length of record (LOR, τ) and spatial binning (black, green, and red lines) using harmonic constants from a coastal tide gauge (La Jolla, CA). Variance of the M_2 signal (one-half the amplitude squared) is indicated with the asterisk at $\tau = 21$yr; and the variance of the M_2 plus S_2 signal is indicated with the triangle. (a) The variance reduction computed via CryoSat-2 sampling oscillates and converges slowly, even when collected within $\pm1^\circ$ bins. (b) In contrast, the variance reduction computed via Jason-1/c sampling is much more stable for records longer than about 3yr, when data are collected within longitude bins of 0.5° and larger.

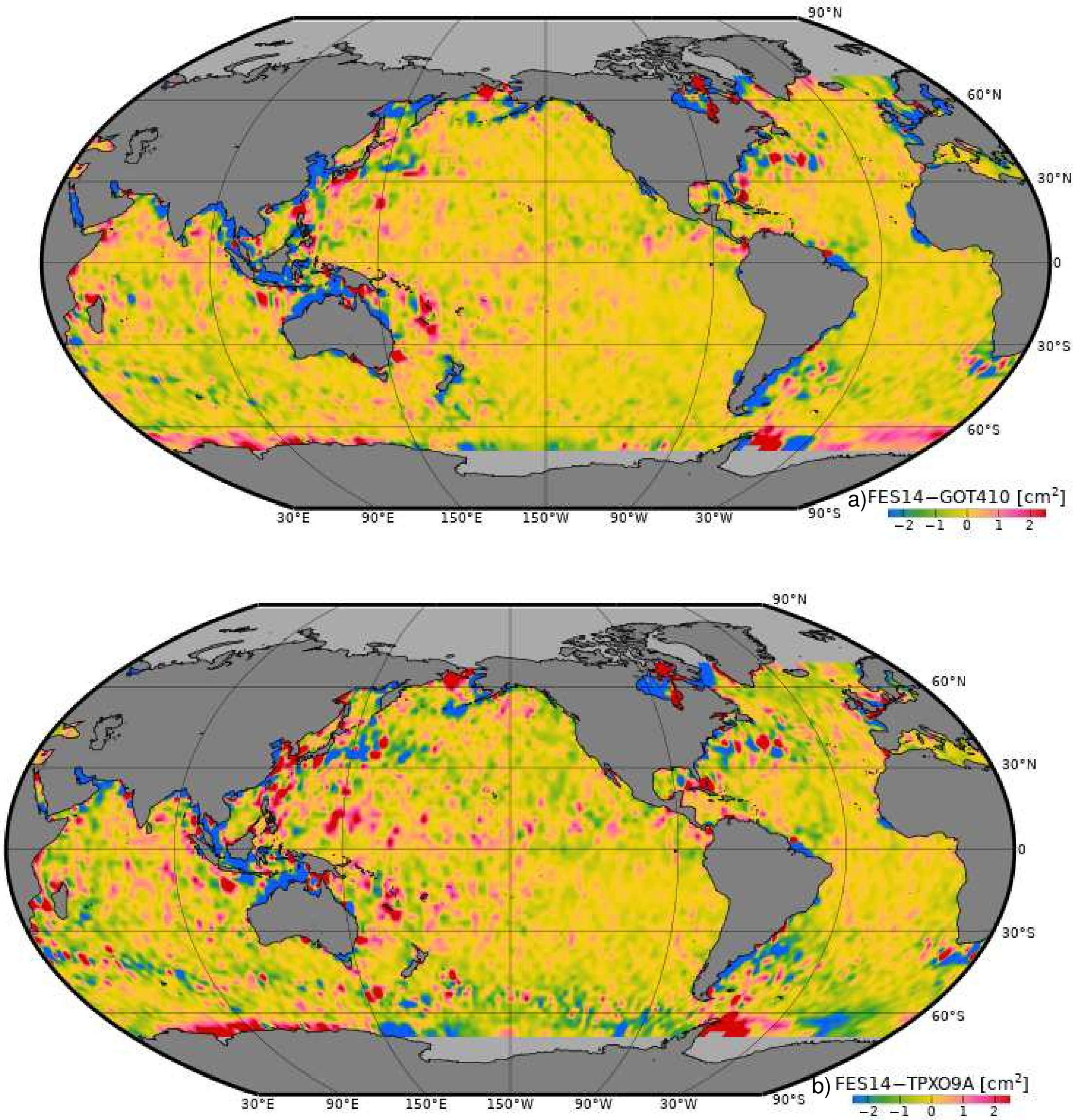

FIG. 3. Jason/GM residual variance maps. (a) Residual variance, FES2014 tide prediction vs. GOT410 tide prediction. (b) Residual variance, FES2014 vs. TPXO9A. Green-blue hues indicate where the residual variance (variance reduction) of FES2014 is smaller (larger) than the model to which it is compared.

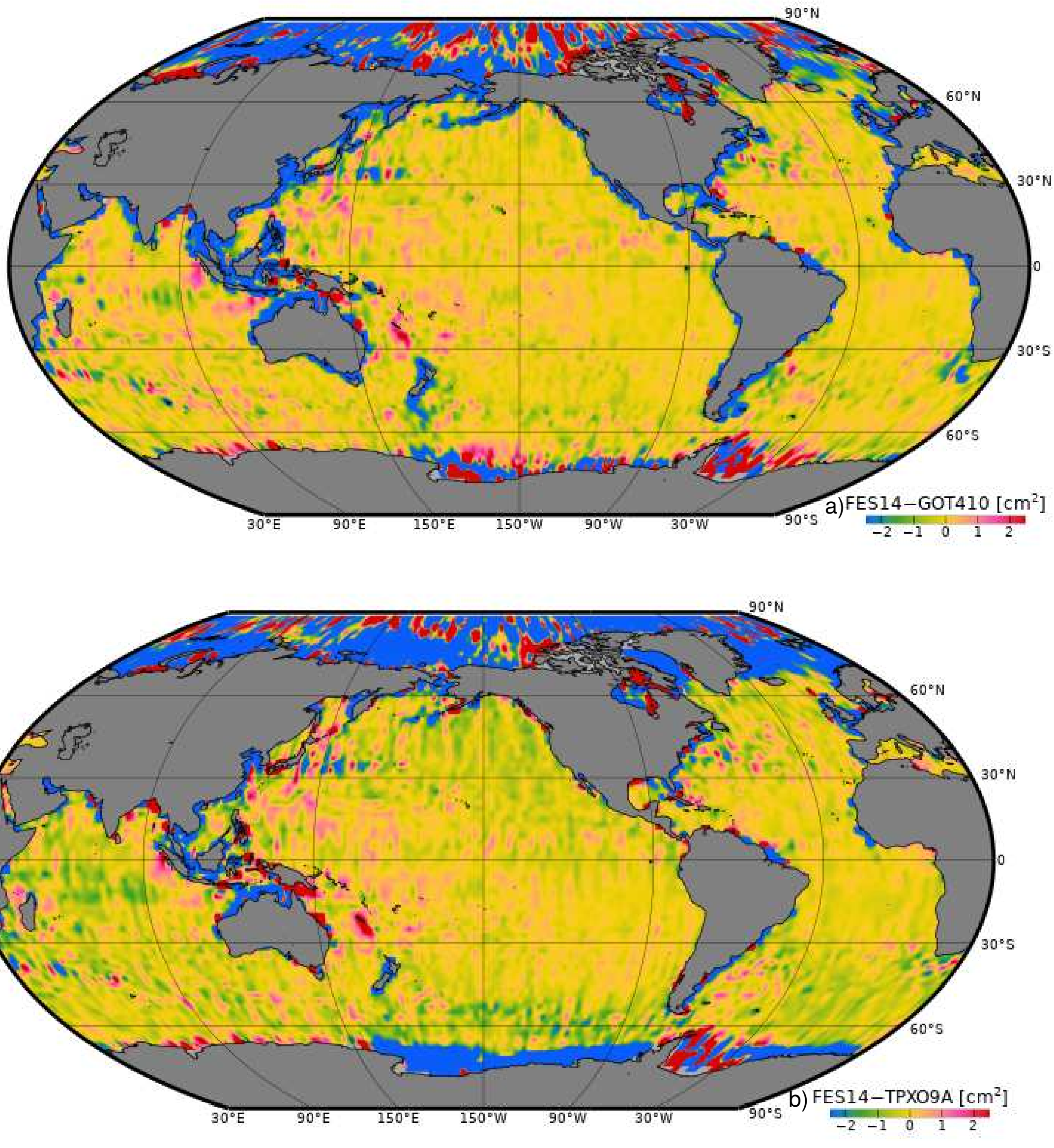

FIG. 4. CryoSat-2 residual variance maps. As in Figure 3.

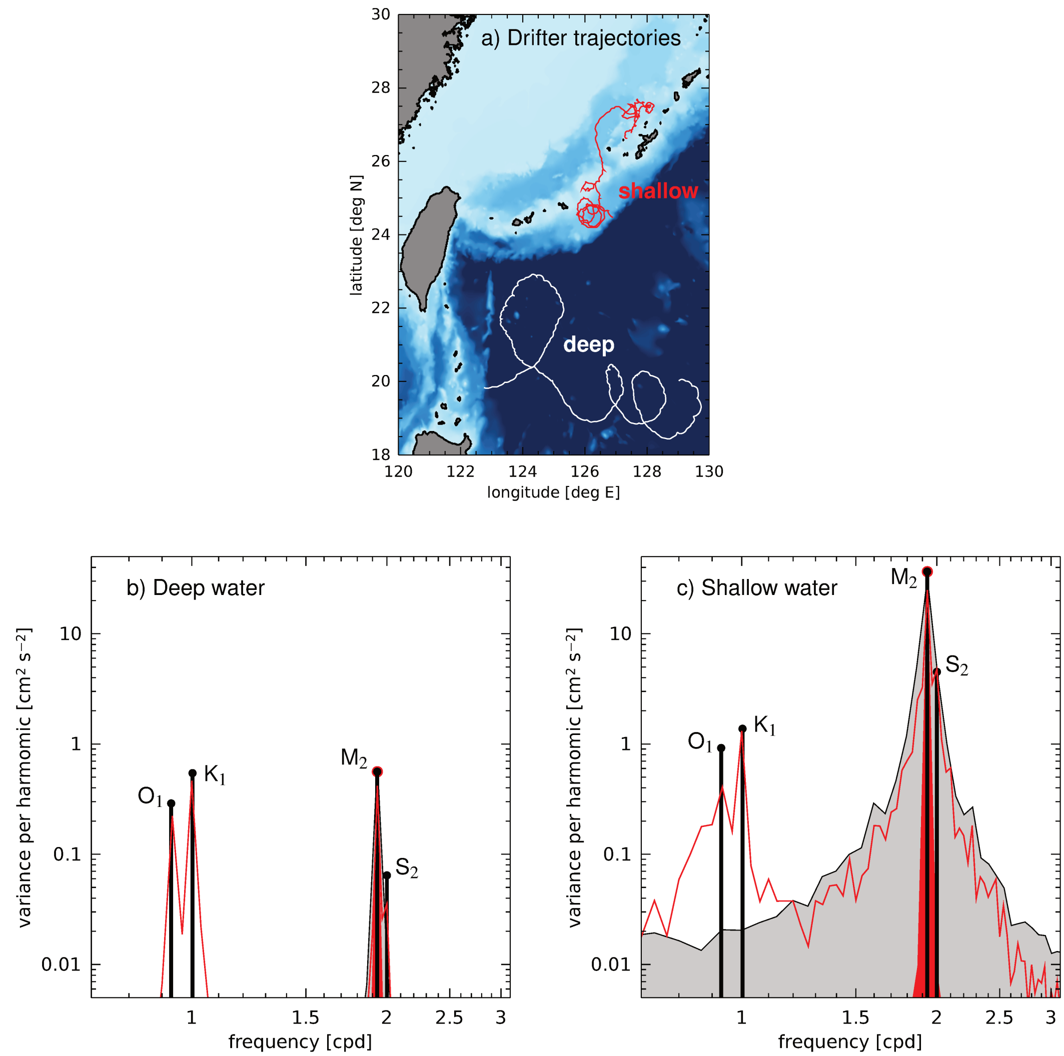

FIG. 5. Tidal kinetic energy spectra sampled along drifter paths. a) Gap-free drifter trajectories in deep (>3000 m) and shallow (<3000 m) water (GDP buoys #64765560, 2017-09-16 to 2017-11-27, and #63894840, 2017-10-31 to 2018-01-14, respectively). Tidal velocity predictions for the M_2, S_2, K_1, and O_1 tides from TPXO9A have been sampled along the trajectories to illustrate the character of tidal signals in Lagrangian spectra in (b) and (c). The sum of the spectra for the two velocity components has been multiplied by the Fourier bandwidth, giving the y-axis units of velocity variance ($\mathrm{cm}^2\mathrm{s}^{-2}$). The spectrum of the tidal velocity (red line) is narrow in (b) deep water, and broad in (c) shallow water. For reference, idealized line spectra (thick black lines) show the tide model kinetic energy at the mid-point of each trajectory. The spectrum of the M_2 velocity is shown for Eulerian sampling at this same location (shaded red, centered on M_2; barely visible in (b)) and for Lagrangian sampling (shaded gray; indistinguishable in (b)). The line broadening in (c) is a consequence of spatial variability of the predicted tidal currents along the trajectory.

FIG. 6. M_2 tidal current ellipses computed from harmonic analysis of spatially binned drifter data (black lines in panels b-g) are compared with model current ellipses sampled along trajectories within the bins (gray lines in panels b-g). a) Drifter data were sampled on the shallow continental shelf (red, < 200 m depth, $N = 2662$ samples) and in the deep ocean (white, > 4000 m depth, $N = 1661$). The current ellipses from the barotropic models (in gray; plotted from the models' harmonic constants every 10th-point along the trajectories) for b) FES2014 and c) TPXO9A are similar in magnitude to the observed current ellipse (in black); the d) HRET current ellipse is too small to be visible at this scale. In the deep water, the e) FES2014 and f) TPXO9A current ellipses are much smaller than the observed, which is more similar to the g) HRET ellipse.

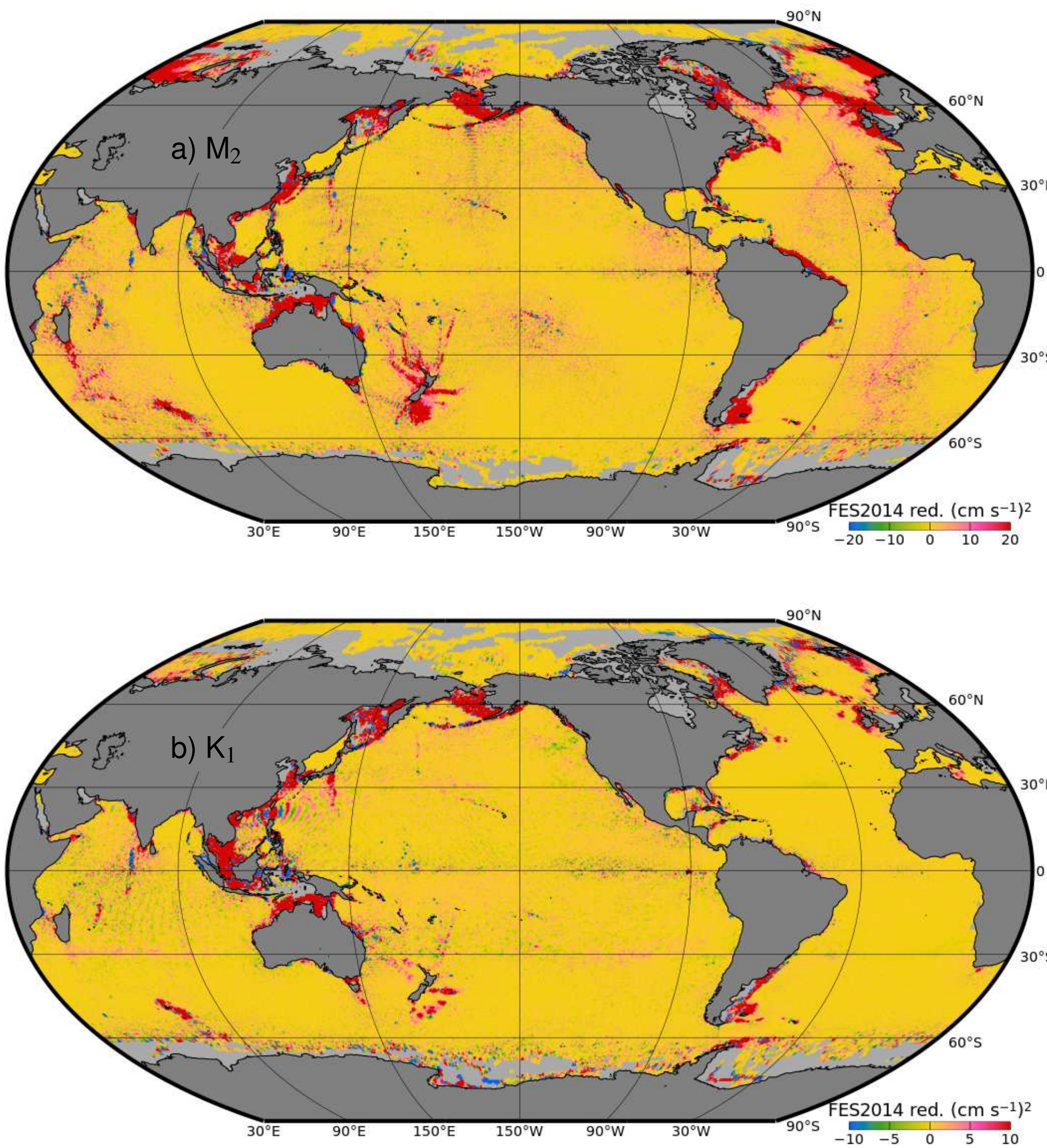

FIG. 7. Velocity variance reduction for the (a) M_2 and (b) K_1 tidal currents predicted by FES2014, averaged within 0.5° bins. The variance reduction is computed as the difference between the variance of the original GDP currents and the GDP currents after removing the predicted barotropic tidal current with FES2014. Baroclinic currents are correlated with barotropic currents and cause spatial oscillations in these high-resolution maps.

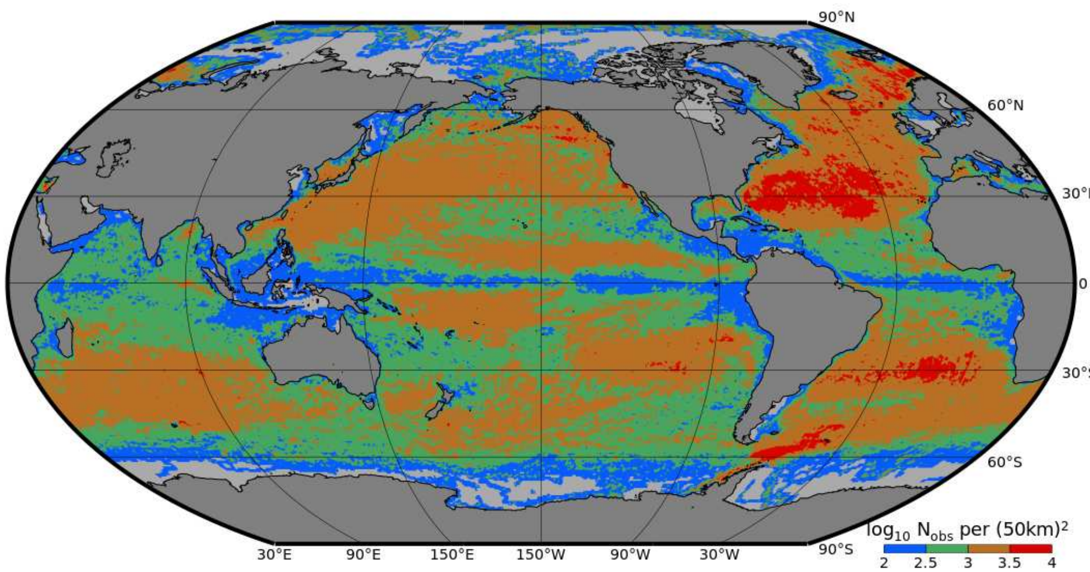

FIG. 8. The spatial density of GDP hourly currents observations, expressed as a number density per $(50\text{ km})^2$ spatial bin. Because the observations are generally made with hourly time resolution, roughly 350 observations are sufficient to discriminate the M_2, S_2, K_1, and O_1 tides, while 4382 observations are needed to discriminate the M_2, S_2, N_2, K_2, K_1, O_1, P_1, and Q_1 tides. Thus, within red-shaded regions there are enough observations to unambiguously identify all 8 major tides within, approximately, $(50\text{ km})^2$ patches. Within the green-brown-red shaded regions there are enough observations to identify the 4 major tides in the same size patches.

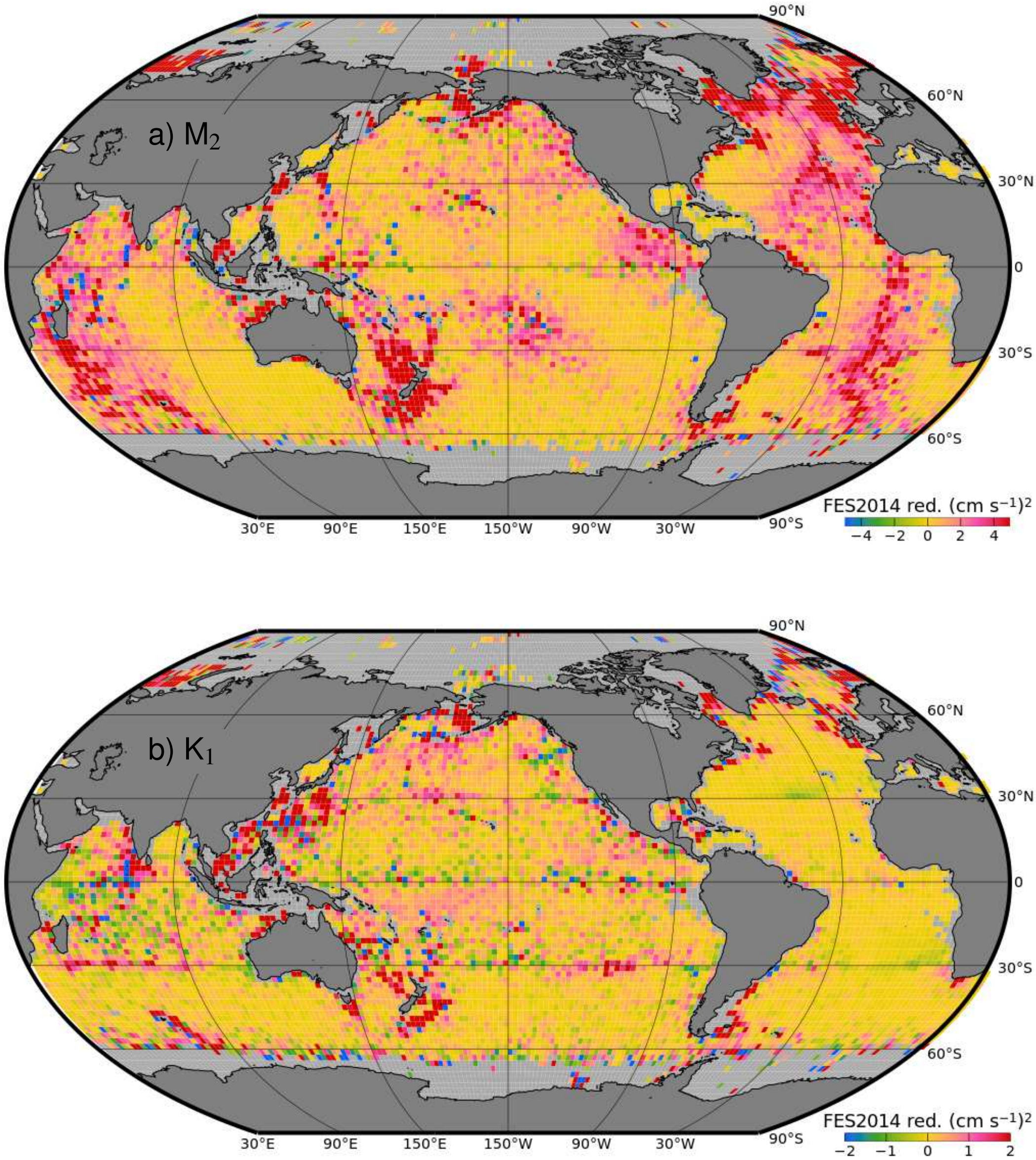

FIG. 9. Drifter current variance reduction for the (a) M_2 and (b) K_1 tidal currents predicted by FES2014, averaged within 2° bins. Note the much smaller range of colorscale used in this plot compared with Figure 7. Averaging within larger spatial bins reduces the oscillations due to baroclinic tides as well as reducing the sampling error.

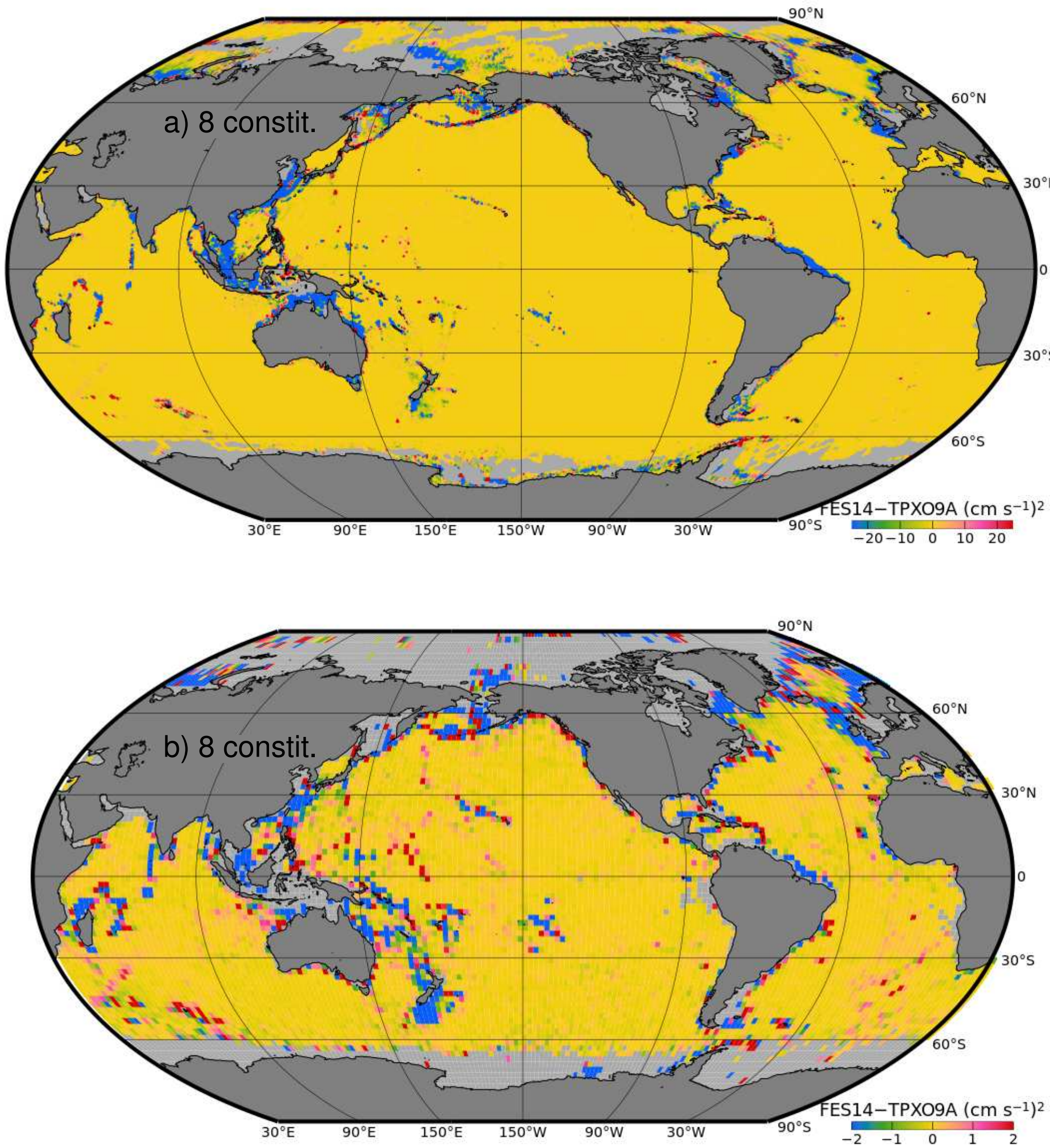

FIG. 10. Difference in drifter velocity residual variance, FES2014 minus TPXO9A, for tidal current predictions summed over the (M_2, S_2, N_2, K_2, K_1, O_1, P_1, Q_1) tides for data within a) 0.5° bins and b) 2° bins. Note the different colorscales used in the panels. Green-blue hues correspond to better FES2014 predictions, while pink-red hues correspond to better TPXO9A predictions.

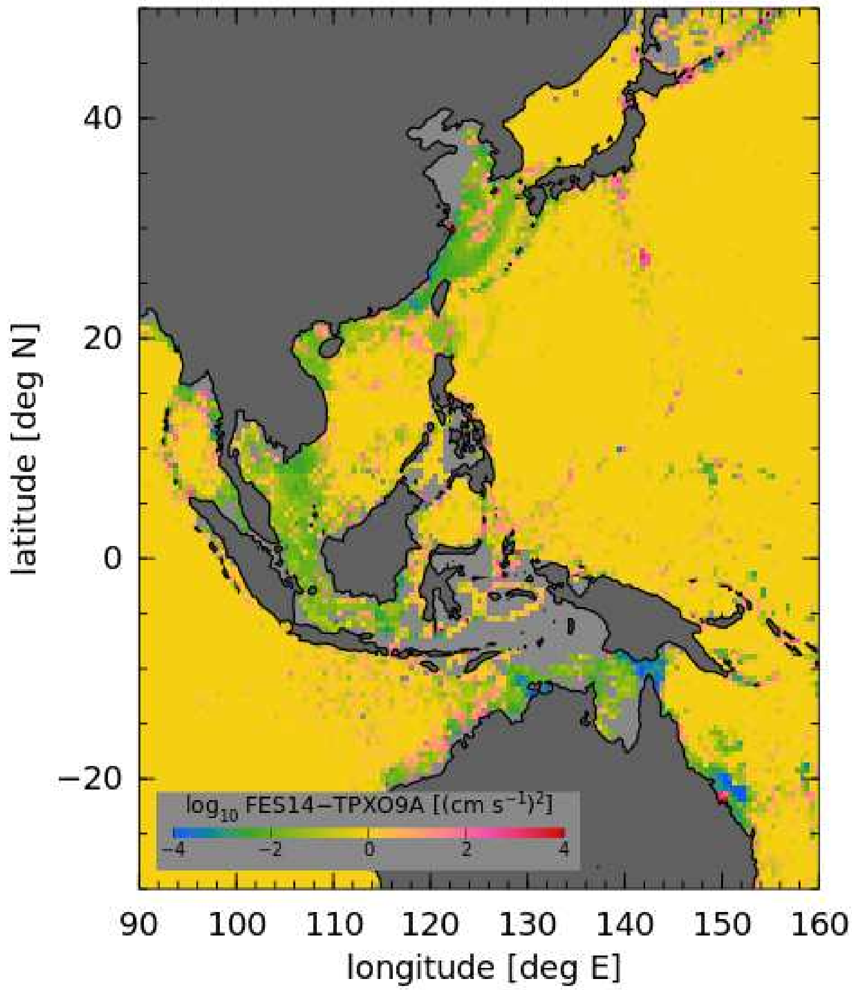

FIG. 11. Log-scaled difference in drifter velocity residual variance, FES2014 vs. TPXO9A (enlarged and different colorscale from Figure 10a). The logarithm of the absolute value of the FES2014 minus TPXO9A residual variance is shown, multiplied by the sign of the difference, so that, for example, -2 corresponds to -10^2 $(\mathrm{cm\,s^{-1}})^2$ difference of residuals (absolute values smaller than 1 $(\mathrm{cm\,s^{-1}})^2$ are truncated to zero). Green-blue hues correspond to better FES2014 predictions, while pink-red hues correspond to better TPXO9A predictions. The log-scaling highlights the sometimes large differences between the FES2014 and TPXO9A tidal currents on continental shelves and in shallow seas. Some apparent anomalies occur near the coastline, for example, near 22°S–150°E, which is examined in detail in Figure 12.

FIG. 12. Drifter trajectories and M_2 tides near 22°S–150°E. a) Drifter trajectories (light red) are sparse in the vicinity of McEwin Islet (indicated) where the observed tidal range exceeds 8 m. Currents along the highlighted trajectory (bright red) are largely responsible for the red pixels near 22°S–150°E in Figure 11, where TPXO9A agrees better with GDP-inferred currents than FES2014. The M_2 tidal ellipse computed from the highlighted drifter trajectory is shown (solid black ellipse) on top of the current ellipses sampled along the trajectory in the (b) FES2014 and (c) TPXO9A models, respectively (gray ellipses). The TPXO9A currents clearly agree with observations better than FES2014 at this site, even though currents along the nearby coast are better described by FES2014 (cf., Figure 11).

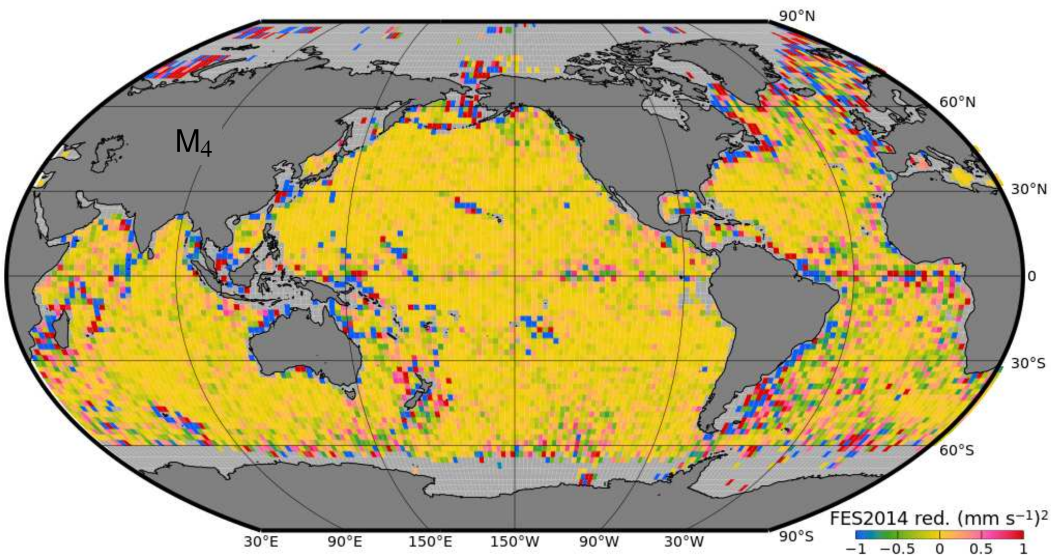

FIG. 13. FES2014 drifter velocity variance reduction for M_4 predictions averaged within 2° bins.