Title

Harmonic Analysis of Non-Phase-Locked Tides with Red Noise Using the red_tide Package

Journal

Journal of Atmospheric and Oceanic Technology, 39(7)

ISSN

0739-0572

Authors

Kachelein, Luke
Cornuelle, Bruce D
Gille, Sarah T
et al.

Publication Date

2022-07-01

DOI

10.1175/jtech-d-21-0034.1

Peer reviewed

Harmonic analysis of non-phase-locked tides with red noise using the red_tide package

Luke Kachelein*, Bruce D. Cornuelle, Sarah T. Gille, and Matthew R. Mazloff

Scripps Institution of Oceanography, University of California San Diego, La Jolla, California

*Corresponding author: Luke Kachelein, lkachele@ucsd.edu

Early Online Release: This preliminary version has been accepted for publication in *Journal of Atmospheric and Oceanic Technology* be fully cited, and has been assigned DOI 10.1175/JTECH-D-21-0034.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT

A novel tidal analysis package (red_tide) has been developed to characterize low-amplitude non-phase-locked tidal energy and dominant tidal peaks in noisy, irregularly sampled, or gap-prone time series. We recover tidal information by expanding conventional harmonic analysis to include prior information and assumptions about the statistics of a process, such as the assumption of a spectrally colored background, treated as non-tidal noise. This is implemented using Bayesian maximum posterior estimation and assuming Gaussian prior distributions. We utilize a hierarchy of test cases, including synthetic data and observations, to evaluate this method and its relevance to analysis of data with a tidal component and an energetic non-tidal background. Analysis of synthetic test cases shows that the methodology provides robust tidal estimates. When the background energy spectrum is nearly spectrally white, red_tide results replicate results from ordinary least squares (OLS) commonly used in other tidal packages. When background spectra are red (a spectral slope of -2 or steeper), red_tide's estimates represent a measurable improvement over OLS. The approach highlights the presence of tidal variability and low-amplitude constituents in observations by allowing arbitrarily configurable fitted frequencies and prior statistics that constrain solutions. These techniques have been implemented in MATLAB in order to analyze tidal data with non-phase-locked components and an energetic background that pose challenges to the commonly used OLS approach.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-00034.1.

1. Introduction

Tides are a major driver of oceanic variability. They are forced by the gravitational effects of the moon and sun and have a ubiquitous presence throughout the global ocean. The tide generating potential was described harmonically by Sir George Howard Darwin in 1883 (Doodson and Lamb 1921) and was further developed by Doodson, with the Darwin symbols for tidal constituents (e.g. O_1 and M_2) still in widespread use. The ubiquity of tides in oceanic data has motivated the development of techniques for determining tidal parameters. Two types of methods are frequently used to analyze tides: discrete Fourier transform-based methods and least squares-based harmonic analysis. Discrete Fourier transform-based methods use the energy contained in discrete frequency bands to diagnose the amplitude of tidal constituents. Least squares harmonic analysis has been used for decades (Munk and Hasselman 1964; Zetler et al. 1965) to estimate the amplitude and phase of tidal signals at known tidal frequencies.

Tidal signals can be separated into two components: the relatively predictable barotropic tide and the more variable baroclinic tide (Ray and Mitchum 1996a). The predictability of the barotropic tide is a consequence of its stable phase and amplitude due to its large scale, rapid propagation, and the regularity of the astronomical forcing. Classical harmonic analysis at tidal constituent frequencies is effective for analyzing time series of sea surface height or bottom pressure that are dominated by the barotropic tide and are characterized by sharp, narrow peaks in the frequency domain at tidal constituent frequencies. The interannual lunar nodal cycle (18.61 years) and lunar perigee (8.85 years), which are not directly resolvable in typically short tidal records, cause tidal modulations that affect the interpretation of tidal records (Haigh et al. 2011). The t_tide package, for example, accounts for these cycles using nodal corrections (Pawlowicz et al. 2002). However, other processes can modulate the tidal peaks.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

In the case of the baroclinic tide, which is more variable in amplitude and phase, propagation through varying stratification or non-linear interaction with other waves leads to amplitude and/or phase modulation via the transfer of energy to the internal wave continuum spectrum and a loss of coherence with the astronomical forcing (Chiswell 2002; Rainville and Pinkel 2006). Interaction with eddy fields, background currents, and the seasonal cycle in stratification all cause internal tides to vary in time (Ray and Zaron 2011). This variability spreads the tidal energy across a band of frequencies centered at the tidal forcing frequency, forming a tidal cusp (Munk et al. 1965); the spreading of energy in the frequency domain can pose challenges to describing the predictable tidal component. The component of tidal energy resulting from interaction with other processes in the ocean appears has been referred to by different names in the literature, including the incoherent tide (e.g. Eich et al. 2004), the non-stationary tide (e.g. Ray and Zaron 2011), and more recently the non-phase-locked tide Zaron (2019). We have chosen the term "non-phase-locked" because it is associated purely with tides and their generating potential, whereas the other terms overlap with wave and statistics terminology. Nevertheless, these other terms are reasonable, as this component of the tides is incoherent with astronomical forcing and is non-stationary in time.

Throughout this paper, *model* refers to the series of basis functions (harmonics) at frequencies chosen to approximate (or *to model*) a given time series, while the coefficients of these harmonics are referred to as model parameters. Solving for these parameters to find the best estimate of the underlying tidal component of the observations is the goal of least squares tidal harmonic analysis. The choice of basis functions is central to the technique, and the designation of these functions as a model for observations follows standard least squares terminology (e.g. Wunsch 1996). Conventional least squares tidal harmonic analysis models a time series as a sum of sinusoids at tidal frequencies, with amplitudes and phases optimized to best fit observations. This contrasts with analysis via the discrete Fourier transform in several important ways: Fourier analysis

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

requires evenly sampled time series and decomposes a signal into components at evenly spaced frequencies determined by the record length and sample rate. The Fourier transform is periodic at the record length ("fundamental") and is band-limited at the Nyquist frequency. Additionally, the Fourier transform does not allow for a separate component of noise. Harmonic techniques for analyzing tidal time series overcome these limitations of Fourier analysis. They allow for estimates that do not exactly match all observations and that provide a unique solution in the presence of noise and potential nonorthogonality between basis functions, for example when the frequencies of basis sinusoids differ by less than the fundamental frequency. This is solved as an inverse problem in which the fit is expected to differ from the observations by some residual. Data may be irregularly spaced and fit to arbitrary basis functions, including sinusoids of any frequency, unconstrained by periodicity over the record length and not band-limited by the sampling rate. Additionally, harmonic analysis allows for a noise component with prior statistics, generally mean and autocovariance. We denote the autocovariance and its matrix representation with the standard terms "covariance" and "covariance matrix", respectively, as we do not discuss any cross-covariance quantities for which such shorthand might be confusing.

At frequencies outside the tidal bands, ocean data tend to be spectrally red, with greater power at lower frequencies (Munk et al. 1965). When finite duration records with steep spectra are analyzed, high-amplitude low-frequency processes with periods that do not match the series length can alter estimated spectral power at higher frequencies. Spectral leakage is problematic for characterizing internal tides, especially in regions where mesoscale variability is much stronger than the internal tide (Ray and Zaron 2016). Low-frequency variability, tidal cusps, and high-frequency noise must all be accounted for either explicitly or as a residual term. In other words, total signal variability is modeled as sinusoids at given frequencies added to a residual broadband background that has power at all frequencies, which is characterized by a residual (or noise) covariance matrix. Our approach

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

is to choose basis functions and prior statistical assumptions about the signal and noise components (quantified in covariance matrices) that match the expected variability in the observations as well as computational cost allows; such constraints bias the model parameters by reducing their variance, which is appropriate when tidal constituents and related components are estimated from limited sampled data with noise. We seek to avoid overfitting and we argue that this produces better results than estimators obtained from methods that use less prior knowledge.

Least squares tidal harmonic analysis has drawbacks. Pawlowicz et al. (2002) identify some challenges including record length requirements for distinguishing some tidal frequencies, the lack of distinction between true tidal lines and background energy at tidal frequencies, and the broadening of spectral lines from estuarine tidal responses and stratification-dependent internal tides. Nevertheless, harmonic analysis has been widely adopted for tides because it is well-suited for signals with a weak noise component relative to the tidal signal. Pawlowicz et al. (2002), expanding upon earlier code for tidal harmonic analysis (Foreman 1977; Foreman and Henry 1989; Foreman et al. 2009) and employing MATLAB, created the widely-used t_tide package, incorporating methods to mitigate known drawbacks to classical harmonic analysis. These methods include nodal corrections and inference of unresolvable constituents to account for the long record lengths required for resolution under classical harmonic analysis, as well as three algorithms to provide confidence intervals to account for non-tidal energy at tidal frequencies. Other authors have expanded the t_tide procedure (Leffler and Jay 2009; Codiga 2011) or have modified it for specific dynamical regimes, such as tides in the presence of river outflow (Matte et al. 2013).

The two-fold problem of accurately estimating tidal variability with a component that is not phase-locked to astronomical forcing in the presence of spectrally-colored noise while minimizing spectral leakage has motivated us to develop a new tidal harmonic analysis package, red_tide, that accounts for tidal cusps and red background spectra. The appendix provides information on the access

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI: 10.1175/JTECH-D-21-0034.1.

of the package, which may be modified to accommodate individual needs. We do not presently incorporate nodal corrections like those used in t_tide in order to emphasize features specific to our method, though these corrections may be incorporated in a future version. The primary scientific motivation for developing the package is to support a detailed level of tidal analysis of highly variable baroclinic tides. For example, such tides are expected to be an important part of the signal at spatial scales of $O(10)$ kilometers to be measured by the upcoming Surface Water and Ocean Topography (SWOT) swath altimeter, for which the interaction between internal tides and ocean mesoscale variability requires the most accurate predictions possible (Chavanne and Klein 2010; Ray and Zaron 2011). Analysis related to such an application is likely to focus on in-situ data and ocean model output sampled at smaller time intervals than semi-diurnal tidal periods. Therefore relevant test cases are presented here. Though harmonic analysis of tidal signals aliased in satellite data has been performed on the phase-locked tide (Ray and Mitchum 1996b; Zhao et al. 2011), this is not examined here. This method is compared with existing methods (ordinary least squares and t_tide, which implements ordinary least squares) using synthetic and observational time series to examine performance and practical limitations. The t_tide package requires fewer inputs and produces matching results for lower computational cost when analyzing time series that do not exhibit the complications that the red_tide method seeks to address. Therefore, our method does not supplant existing tidal analysis packages except in specific regimes where it is advantageous. Other tidal packages, including those that are built upon t_tide (e.g. Leffler and Jay 2009; Codiga 2011; Matte et al. 2013), are not examined in this study.

The rest of this study is divided into three sections. We begin by outlining the linear algebraic and statistical methods that underpin our tidal analysis in Section 2. We then apply these methods to synthetic time series, first by highlighting specific features of the method, including application to a step function to demonstrate aliased signals (Section 3a), and then by analyzing tide-like synthetic

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

series to show the effect of prior statistical assumptions on model parameters (Sections 3b-d). Two examples of observational data follow in Section 4, with comparisons to t_tide. Finally, summary and discussion follow in Section 5.

2. Methods

The basic framework of weighted least squares estimation used in red_tide is outlined in a number of references (e.g. Wunsch 1996; Menke 2018). Here we provide a review of this framework, formulated for tidal harmonic analysis of records with arbitrarily structured noise, with notation following that of Ide et al. (1997) with a few modifications. Note that tidally-driven components of a time series are considered "signal" while non-tidal processes are referred to as "noise" for the purpose of distinguishing them in the context of harmonic analysis. These so-called noise terms may include instrument error as well as non-tidal processes in the ocean, such as submesoscale eddies and the internal wave continuum.

A zero-mean, discretely sampled ocean time series written as a column vector $\mathbf{y}$ of length N is modeled as the sum of sinusoids of tidal and non-tidal frequencies

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{r}, \tag{1}$$

where $\mathbf{H}$ is an N by $2M$ regressor matrix (i.e. the model basis functions), $\mathbf{x}$ are the $2M$ model parameters, and $\mathbf{r}$ represents the N-element residual time series. The columns of $\mathbf{H}$ are sines and cosines of prescribed frequencies, ω_m, for $m = 1, 2, \ldots M$, such that equation (1) can be expressed as

$$\begin{aligned} \mathbf{y} &= \sum_{m=1}^{M} \left(a_m \sin(\omega_m \mathbf{t}) + b_m \cos(\omega_m \mathbf{t}) \right) + \mathbf{r}, \\ \mathbf{x} &= [a_1, b_1, a_2, b_2, \ldots, a_M, b_M]^{\mathrm{T}} \\ \mathbf{t} &= [t_1, t_2, \ldots, t_N]^{\mathrm{T}} \end{aligned} \tag{2}$$

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

The unknown model parameters $\mathbf{x}$ are estimated by $\hat{\mathbf{x}}$ such that the trace of $\langle(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^{\mathrm{T}}\rangle$, the expected value of the sum of the squares of the model parameter errors, is minimized. This is done by calculating $\hat{\mathbf{x}}$ as the Bayesian maximum a posteriori (MAP) estimate, by requiring that $\mathbf{x}$ and $\mathbf{r}$ have independent Gaussian distributions which satisfy the conditions under which the MAP estimate gives $\hat{\mathbf{x}}$ (Van Trees 2001). By Bayes' theorem, the posterior probability distribution of $\mathbf{x}$ given observations $\mathbf{y}$ is proportional to the product of the prior probability distribution of $\mathbf{x}$ and the likelihood of $\mathbf{y}$ given $\mathbf{x}$:

$$\mathrm{P}(\mathbf{x}|\mathbf{y}) = \frac{\mathrm{P}(\mathbf{x})\mathrm{P}(\mathbf{y}|\mathbf{x})}{\mathrm{P}(\mathbf{y})}. \tag{3}$$

Given the assumptions of Gaussian statistics and a linear model:

$$\mathrm{P}(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\mathbf{x}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{x}\right)\exp\left(-(\mathbf{y}-\mathbf{H}\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x})\right). \tag{4}$$

The denominator $\mathrm{P}(\mathbf{y})$ is not a function of $\mathbf{x}$ and therefore can be omitted, since it is not relevant to the optimization. The matrices $\mathbf{R} = \langle\mathbf{r}\mathbf{r}^{\mathrm{T}}\rangle$ (size N by N) and $\mathbf{P} = \langle\mathbf{x}\mathbf{x}^{\mathrm{T}}\rangle$ (size $2M$ by $2M$) are the covariance matrices of $\mathbf{r}$ and $\mathbf{x}$ respectively ($\mathbf{P}$ is thus a hyperparameter of the Gaussian prior for $\mathbf{x}$). This expression as well as its logarithm,

$$\ln\left(\mathrm{P}(\mathbf{x}|\mathbf{y})\right) = -\mathbf{x}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{x} - (\mathbf{y}-\mathbf{H}\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}) + \text{constant}, \tag{5}$$

are at a maximum for some value $\mathbf{x} = \hat{\mathbf{x}}$. Since the posterior probability $\mathrm{P}(\mathbf{x}|\mathbf{y})$ follows a Gaussian distribution, its mean (the Bayes estimator) equals its mode (the MAP estimator) and can therefore be solved as a maximization problem (Van Trees 2001). At the mode, the partial derivative of equation (5) with respect to $\mathbf{x}$ vanishes:

$$0 = \frac{\partial}{\partial\mathbf{x}}\ln\left(P(\mathbf{x}|\mathbf{y})\right)\bigg|_{\mathbf{x}=\hat{\mathbf{x}}} = -2\mathbf{P}^{-1}\hat{\mathbf{x}} + 2\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y} - 2\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\hat{\mathbf{x}}. \tag{6}$$

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

The vector $\hat{\mathbf{x}}$ contains the most probable model parameters given observations $\mathbf{y}$, and solving for $\hat{\mathbf{x}}$ gives the most probable solution to equation (1):

$$\hat{\mathbf{x}} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}. \tag{7}$$

An equivalent expression via the matrix inversion lemma (e.g. Wunsch 1996) is not used here but appears in the appendix.

The posterior covariance matrix of the difference between the estimated and true model parameters is

$$\langle(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^{\mathrm{T}}\rangle = \left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1}\right)^{-1}. \tag{8}$$

In red_tide, model parameters $\mathbf{x}$ are assumed to have a Gaussian probability distribution function (PDF) resulting from the Gaussian distributions of the prior and likelihood function. Therefore, the posterior PDF of x_m, the m-th element of $\mathbf{x}$, is a Gaussian with a mean given by the m-th element of $\hat{\mathbf{x}}$ from equation (7) and variance given by the m-th element of the diagonal of the matrix in equation (8).

If $\mathbf{x}$ was not expected to have a Gaussian distribution, a different expression for $\hat{\mathbf{x}}$ would need to be derived by similar Bayesian principles starting from equation (3), which is beyond the scope of this study. Quantities defined as nonlinear functions of Gaussian-distributed quantities, however, may have non-Gaussian distributions that can be estimated. In this paper, plotted uncertainty bounds for derived quantities are estimated using a Monte Carlo approach, based on an ensemble population of $\mathbf{x}$ with the posterior Gaussian PDF. This approach is implemented due to its flexibility and simplicity compared with analytical solutions, which are not always in closed form, or with more complicated approximations such as piecewise linear discretization of PDFs (e.g. Lourens and van Geer 2016). The tidal amplitude is one such quantity, where $\hat{A}_m = \sqrt{\hat{a}_m^2 + \hat{b}_m^2}$ is an estimator of the true amplitude A_m, which follows a noncentral χ distribution when the standard deviations of

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

a_m and b_m are equal ($\sigma_{a_m} = \sigma_{b_m} = \sigma_m$). Similarly, the tidal phase $\hat{\phi}_m$ is an estimator for the true phase shift ϕ_m, defined by $\tan(\phi_m) = -b_m/a_m$, whose statistics are also not Gaussian. Whenever uncertainty bounds are given for quantities derived from the Bayesian methods outlined above or for quantities that are functions of them, we refer to them as *credible intervals* in accordance with Bayesian terminology for the interval in which an estimated parameter lies with the stated probability (Lee 1997). These are analogous to *confidence intervals*, the uncertainty bounds on quantities derived from frequentist methods. Because Pawlowicz et al. (2002) uses the term *confidence intervals*, and because t_tide is not derived using a Bayesian framework, we refer to t_tide output and other non-Bayesian quantities as having confidence intervals when comparing it to red_tide output and its credible intervals.

Because **y** is modeled as the sum of sinusoids and a noise component, its expected power spectrum $S_{\mathbf{yy}}(f)$ may be interpolated to fitted frequencies and used to construct **P**. This approach, however, results in doubly counting energy, as **r** contributes to the variance of **y** at all frequencies, including frequencies modeled by **Hx**. Because the energy of a tidal peak and cusp are typically much higher than the background noise and may in fact be underestimated in $S_{\mathbf{yy}}(f)$ due to peak-broadening from spectral averaging, the double counting of energy will be small around most prominent tidal peaks. For cases where tidal signals are comparable in energy to **r** at tidal frequencies, **P** can be reduced by the appropriate amount. In sections 3c and 3d, the correct partition of energy into **P** versus **R** is possible because we have perfect knowledge of the underlying synthetic processes. For real data for which we lack perfect knowledge, such as those in section 4, the approach we use for convenience is to assume wide-sense stationary noise so that we can obtain the residual spectrum $S_{\mathbf{rr}}(f)$ from the Fourier transform of any column of **R**. We then subtract it from $S_{\mathbf{yy}}(f)$ to construct **P** in order to obtain a more accurate partition of signal versus noise energy. We also assume throughout that **y** is a wide-sense stationary time series; this means that the elements of **x**

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI: 10.1175/JTECH-D-21-00034.1.

are assumed to be uncorrelated and therefore that **P** is diagonal throughout (see Bendat and Piersol (2010) for nonstationary data analysis and double frequency spectra, the continuous analogue to a non-diagonal **P**).

Values along the main diagonal of **R** represent the expected variance of the misfits between fitted time series and observations. Off-diagonal elements indicate the covariance at lagged times, with values farther off the main diagonal corresponding to larger time lags. Beyond some time lag, the covariance may be approximated as zero if long-period energy is sufficiently small or explicitly represented in **H**. This approximation is useful, as it limits the memory requirement for large **R**. A diagonal **R** is a special case of this, with non-zero elements only along its main diagonal. This corresponds to an assumption of zero lagged noise correlation or equivalently an assumption of spectrally white noise. This approximation is often made for computational efficiency, as $\mathbf{R} = \sigma^2\mathbf{I}$ can be replaced in equation (7) with the constant σ^2, which is estimated as $\sigma^2 = \langle \mathbf{r}^{\mathrm{T}}\mathbf{r} \rangle / N$. In cases where the residual **r** may be better approximated as non-white noise, other procedures can be used to construct a non-diagonal **R**, including those described in section 2b and the appendix. In the examples that follow, we assume that all residuals have the same variance. This assumption of stationarity corresponds to constant elements along the diagonals of **R**, i.e. a Toeplitz matrix.

Time series with a nonstationary residual **r**, for example due to time-varying instrumental noise, may have their residual covariance approximated by a non-Toeplitz matrix **R** in order to reduce the impact of the affected segments on the calculation of $\hat{\mathbf{x}}$. Additionally, observational gaps have a non-trivial effect when treating background noise as correlated. The matrix **R** is also not Toeplitz in this case, though for the direct inversion of relatively small **R**, this does not seem to impact performance. Computation may be reduced when analyzing multiple time series if each time series $\mathbf{y}_n$ can be analyzed with the same **H**, **R**, and **P** such that all terms in equation (7) before **y** are evaluated once and multiplied by each $\mathbf{y}_n$.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

a. Relation to Least Squares

Ordinary least squares (OLS) has the same form as weighted least squares if $\mathbf{R} = \sigma^2\mathbf{I}$ and $\mathbf{P}^{-1} \to 0$ (Wunsch 1996), and is used in the t_tide package (Pawlowicz et al. 2002). OLS seeks to minimize $(\mathbf{y}-\mathbf{H}\mathbf{x})^{\mathrm{T}}(\mathbf{y}-\mathbf{H}\mathbf{x})$, the misfit between the fitted time series and observations. In this case, the solution is

$$\hat{\mathbf{x}}_{\mathrm{OLS}} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{y}. \tag{9}$$

Observation error is assumed to be white noise, and OLS places no constraint on the magnitude of the components of $\hat{\mathbf{x}}_{\mathrm{OLS}}$. Steep, non-constant background spectra would need to be modeled by including more parameters within this framework, which may be impractical. Without using a probabilistic framework, in order to constrain the magnitude of the model parameters $\hat{\mathbf{x}}$ to avoid overfitting, the quantity to be minimized is often written as $(\mathbf{y}-\mathbf{H}\mathbf{x})^{\mathrm{T}}(\mathbf{y}-\mathbf{H}\mathbf{x}) + \mathbf{x}^{\mathrm{T}}\mathbf{W}_{\mathbf{x}}\mathbf{x}$, where $\mathbf{W}_{\mathbf{x}}$ is a matrix that weights the relative importance of minimizing model parameter magnitude over misfit. This technique is called ridge regression and reduces overfitting at the expense of bias (Wunsch 1996). The resulting least squares equation is said to be *regularized*:

$$\hat{\mathbf{x}}_{\mathrm{TLS}} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \mathbf{W}_{\mathbf{x}}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{y}. \tag{10}$$

Similarly, the misfit $\mathbf{r}$ may be weighted by the matrix $\mathbf{W}_{\mathbf{r}}$ such that $(\mathbf{y}-\mathbf{H}\mathbf{x})^{\mathrm{T}}\mathbf{W}_{\mathbf{r}}(\mathbf{y}-\mathbf{H}\mathbf{x}) + \mathbf{x}^{\mathrm{T}}\mathbf{W}_{\mathbf{x}}\mathbf{x}$ is minimized, with the corresponding *weighted* regularized least squares equation

$$\hat{\mathbf{x}}_{\mathrm{WTLS}} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{W}_{\mathbf{r}}\mathbf{H} + \mathbf{W}_{\mathbf{x}}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{W}_{\mathbf{r}}\mathbf{y}. \tag{11}$$

This is equivalent to equation (7) when $\mathbf{W}_{\mathbf{x}} = \mathbf{P}^{-1}$ and $\mathbf{W}_{\mathbf{r}} = \mathbf{R}^{-1}$, though the derivation of that solution is distinct from these least squares approaches (see section 3.6.2 of Wunsch 1996, for details). A weighted least squares estimate (not regularized) is used in the UTide package (Codiga 2011), which implements an iteratively reweighted robust fit corresponding to equation (11) with $\mathbf{W}_{\mathbf{x}} = \mathbf{0}$

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

and $\mathbf{W_r}$ as a diagonal weighting matrix that de-emphasizes outliers and is updated iteratively, the details of which are beyond the scope of this study.

b. Power Law Noise

Away from energetic processes at tidal, seasonal, and inertial frequencies, spectra S of ocean time series tend to follow a power law of the form $S \propto f^{\gamma}$, where $\gamma \leq 0$ is the spectral slope and the negative value indicates more energy at lower frequencies (Agnew 1992). For computational efficiency in problems with many observations, we can assume wide-sense stationary noise and construct the residual covariance matrix $\mathbf{R}$ as a sparse, symmetric Toeplitz matrix with the diagonals calculated from the Fourier transform of $S \propto f^{\gamma}$ per the Wiener-Khinchin theorem, truncated at a user-defined time lag. Users of red_tide have the option of multiplying the covariance by a window function in order to reduce the spectral ringing that results from an abrupt drop to zero in the frequency domain. A low-amplitude white spectrum (spectral slope of 0) may also be added to account for observational error, such that the spectrum of $\mathbf{r}$ has a noise floor at all frequencies. This residual spectrum is not altered within fitted tidal frequency bands due to the impossibility of distinguishing tidal and non-tidal energy at the same frequency, which results in model parameter uncertainty. For the typical case of tides that are much more energetic than the background, this effect is small, while in cases where tidal constituents have low energy or a broadband cusp of interest, the relatively larger uncertainty estimates on model parameters reflect the fact that non-tidal variance is comparable to tidal variance. The approximation of the residual time series following some modified spectral power law with $\gamma < 0$ will hereafter be referred to as a *red noise* assumption, even when γ is not exactly -2. In addition to the special cases considered above, the noise can exhibit more complicated structure that allows $\mathbf{R}$ to be tractable while still representing some forms of red noise, for example noise as an autoregressive process (see Appendix).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

3. Application of Red_Tide to Illustrative Cases

In order to compare the performance of red_tide to other widely-used fitting procedures, we lay out several illustrative examples of its application to synthetic and physical data. Before more complicated cases are examined, it is worth reviewing a simplified case of tidal harmonic analysis that typifies the separate treatment of signal and noise and demonstrates the consequences of using an incomplete model. A typical ocean time series will have many tidal constituents present, with longer records better able to differentiate nearby frequencies. Because harmonic methods model the data using sinusoids of prescribed tidal frequencies, energy at unmodeled frequencies will remain in the residual time series. Figure 1 depicts the results of modeling a bottom pressure record (examined in greater detail in section 4a) at only two tidal frequencies: the principal lunar semidiurnal (M_2) and the principal solar semidiurnal (S_2). Figure 1A, with the interval in gray expanded in panel B, illustrates the misfit between the complicated data and the simple model that arises due to substantial residual energy at unmodeled frequencies. In the frequency domain (figure 1C), this is evident by the energetic tidal lines at frequencies that are not included in $\mathbf{H}$. For a data set with well-defined tidal lines with energy much greater than that of the background, conventional harmonic analysis like that employed by t_tide is well-suited to addressing this issue by simply including more tidal constituents. The red_tide package is designed for more complicated cases for which single tidal constituents are insufficient as well as cases where non-tidal variance is comparable to tidal variance.

a. Model Coefficient Covariance, the Gibbs Phenomenon, and Periodicity

One challenge to harmonic analysis stems from the accurate representation of model coefficient covariance matrix $\mathbf{P} = \langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle$. Before further examining realistic cases, we start with a familiar example that demonstrates the effect of the covariance on computed model coefficients for a discrete

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-00034.1.

step function. Modeling the step as a finite set of sinusoids leads to the Gibbs phenomenon, the tendency of the partial sum of a Fourier series to overshoot in the neighborhood of a discontinuity of the modeled function (Hewitt and Hewitt 1979). This phenomenon persists even with the addition of more terms in the partial sum, though the magnitude is reduced. While a discretely sampled time series may be fit well at observation points, band-limited Fourier coefficients do not adequately fit the step function between sampling times. The step function and resulting Gibbs phenomenon serve as a simple but extreme example of a situation present in real data: a process with variance at unmodeled frequencies has a prior that does not adequately describe the process. The variance of a step function is distributed across all frequencies (see equation (12) below), but when it is sampled coarsely and reconstructed as a band-limited process, some variance is aliased and the true underlying process is poorly reconstructed despite good agreement at observation times. Therefore, we incorporate assumptions and prior knowledge of a process, including cases in which there are fewer data than parameters that are suspected to be worth estimating.

Another challenge stemming from the limitations of finite sampling and fitting is the inherent periodicity of solutions when a finite record is modeled as the sum of periodic functions; this is not a problem for tidal processes, which have periods much shorter than those of typical observations, but it will affect estimates for explicitly modeled low-frequency processes that may also be of interest. To demonstrate the effect of the estimator on both of these related issues (Gibbs phenomenon and artificial periodicity), we analyze a finite, uniformly sampled record of a step function using the method described in Section 2.

Because the step function, here represented continuously by the sign (signum) function, has an analytic Fourier transform, the expected model parameter covariance matrix $\mathbf{P}_{\text{step}} = \langle \mathbf{x}_{\text{step}} \mathbf{x}_{\text{step}}^{\text{T}} \rangle$ may

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

be constructed from the absolute value squared of the Fourier transform:

$$y_{\text{step}}(t) = \text{sgn}(t - t_0) = \begin{cases} -1, & t < t_0 \\ 0, & t = t_0 \\ 1, & t > t_0 \end{cases}$$

$$|\mathcal{F}\{\text{sgn}(t - t_0)\}(f)|^2 = \frac{1}{\pi^2 f^2}, \tag{12}$$

where $y(t)$ is the underlying continuous time series to be analyzed after discrete sampling, $\text{sgn}(t - t_0)$ is the sign function with jump discontinuity at t_0 and $\mathcal{F}\{\cdot\}(f)$ denotes the Fourier transform operation to the frequency domain. The square of the Fourier transform of this function is proportional to frequency^{-2}, therefore it has a spectral slope of -2 .

In the example, a step function of record length T is sampled symmetrically about the jump discontinuity at 1000 evenly-spaced times such that all Fourier frequencies, from the fundamental frequency $\Delta f = 1/T$ to the Nyquist frequency f_{Ny}, could be computed. Different choices of basis functions and model parameter covariance are used to evaluate the sensitivity of red_tide output to these inputs. If the model includes only frequencies greater than or equal to Δf, spaced at increments of Δf, then the model will have a fundamental periodicity of T, the record length, because this is the longest period represented in the model. To reduce this effect, we incorporate frequencies less than the fundamental frequency into $\mathbf{H}$, starting at $\Delta f/2$ and increasing by intervals of $\Delta f/2$. This extends the model periodicity to $2T$, twice the record length (Figure 2A), and also reduces Gibbs-like behavior at the beginning and end of the fitted time series.

Separately, we examine the effects of different model parameter covariance matrices on the behavior of the fitted time series near a jump discontinuity. Two choices for the model parameter covariance matrix $\langle \mathbf{x}\mathbf{x}^{\text{T}} \rangle$ are tested: (1) an assumption that the covariance is constant at all frequencies (a spectrally white process), and (2) an assumption that its power spectrum is proportional

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

to f^{-2} (a spectrally red process representing the true spectrum in equation (12)). Two sets of basis functions are used: one that is comprised of sinusoids at the Nyquist frequency f_{Ny} and lower, and one that is comprised of sinusoids at frequency $2f_{\mathrm{Ny}}$ and lower. Both basis sets are spaced in frequency by $\Delta f = 1/T$. Discretely sampled sinusoids of frequencies greater than f_{Ny} are indistinguishable from sinusoids at the frequencies lower than f_{Ny} to which they are aliased. The larger but linearly dependent bases set is used in order to demonstrate the effect that **P** in equation (7) has in constraining an underdetermined system; in this case, describing the behavior of a time series near a step discontinuity at times between the sampling interval requires that frequencies greater than f_{Ny} be represented, constrained by the expected spectral power of the signal. The residuals are assumed to be uncorrelated (white noise), and the expected residual variance can be calculated from the total spectral power at frequencies above the highest-frequency basis. The expected fraction of residual variance is calculated from the integral of the true spectrum over frequencies not explicitly modeled:

$$\frac{\mathrm{var}(\mathbf{r})}{\mathrm{var}(\mathbf{y})} = \frac{\int_{f_{\mathrm{high}}}^{\infty} f^{-2}\mathrm{d}f}{\int_{f_{\mathrm{low}}}^{\infty} f^{-2}\mathrm{d}f} = \frac{f_{\mathrm{low}}}{f_{\mathrm{high}}}, \tag{13}$$

where f_{low} and f_{high} are the lowest and highest frequencies in **H**. Frequencies less than f_{low} are not included in the integral due to the singularity of f^{-2} at $f = 0$ and the approximation of very low-frequency variance as a mean and trend. The total residual variance in this case is small, and hence the impact of the residual covariance matrix is expected to be negligible and is not examined in this example. The effects of different assumptions of residual covariance are examined later with data for which these effects are noticeable. Both fitted time series in figure 2A, periodic on T and on $2T$, incorporate a covariance matrix **P** constructed from a spectrum proportional to f^{-2}.

Near the jump discontinuity, there are two basis choices that we will consider: distinguishable basis sinusoids up to the Nyquist frequency and indistinguishable sinusoids up to twice the Nyquist

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI: 10.1175/JTECH-D-21-0034.1.

frequency, which are aliased for $f > f_{\mathrm{Ny}}$. We also consider two choices for model parameter covariance: a constant value and a covariance proportional to the known spectral power of the data. Together, these form four regimes that can demonstrate the effects of basis choice and model parameter covariance on the resulting fitted approximations to the same step data. The OLS approach models the step function as a series of sinusoids up to the Nyquist frequency with no assumptions about the model parameter covariance. This results in the Gibbs phenomenon near the jump discontinuity (Figure 2B). This result is nearly identical to the partial sum of the Fourier series for this function (not shown), with a slight difference due to least squares' allowance of non-zero $\mathbf{r}$. Using the same basis functions but making the assumption that the main diagonal of the parameter covariance matrix $\mathbf{P}$ corresponds to a spectral slope of -2 as does the true spectrum, the red_tide procedure reduces the Gibbs phenomenon at the expense of greater misfit in the immediate vicinity of the jump discontinuity (Figure 2C).

The Gibbs phenomenon may be reduced by including additional frequencies. Sinusoids at frequencies greater than the Nyquist frequency will be aliased, however, so the effect of the model parameter covariance is more pronounced than in the former examples. Figure 2D shows the result of fitting a finite record to indistinguishable (resolved and aliased) frequencies while naively assuming a spectrally white process. Without the regularization provided by a sufficiently accurate $\mathbf{P}$, the duplicate bases in $\mathbf{H}$ render it rank deficient and therefore the interpolated fitted time series is unrealistically large in amplitude at unobserved times. Modeling aliased frequencies with the assumption of an accurate (spectrally red) covariance, however, results in a fit that has further reduced the Gibbs phenomenon while also reducing the misfit at sampled times immediately before and after the jump discontinuity (Figure 2E). This approach (regularizing according to prior statistics) yields the smallest root-mean-square error of the four regimes, both at observation times and at the interpolated higher temporal resolution (Table 1). Though the solution in panel

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

E does not pass through the observations like those in panels B and D, it nevertheless is the best representation of the underlying sampled process due to its reduction of the Gibbs phenomenon, despite fitting to the same data. Note that the accuracy of these interpolations is only quantifiable here due to knowledge of the true underlying function from which the data are perfectly drawn, which is not possible with real observations.

These refinements to the standard OLS approach demonstrate three advantages of the methods used in red_tide. First, the use of a fundamental frequency lower than that suggested by the record length reduces the effect of periodicity imposed by the model on the solution by lengthening the time scale of this periodicity. Second, the choice of model parameter covariance matrix **P** impacts the solution: a choice of **P** that is more representative of the true behavior of the data reduces the magnitude of the Gibbs phenomenon. Third, frequencies inaccessible to a discrete Fourier transform may be included to more realistically account for variance. The ambiguity of aliased frequencies (those exceeding the Nyquist frequency but at which variance is present) is reduced by using an accurate **P**. We do not examine aliased signals elsewhere in this study.

b. Synthetic Time Series

Real ocean time series are substantially more complicated than the idealized step function above, with variance distributed across a range of frequencies, driven by a variety of physical processes. In order to simulate data with tidally-driven components, we generate synthetic time series from smooth spectra of predefined slope (the continuum), with sharp peaks in frequency space that simulate tides. In this way, we can exactly control and separate the signal and noise components. In order to simulate tidal cusps, we consider a modulation of the tidal amplitude following the description of amplitude-modulated baroclinic tides off Hawai'i by Chiswell (2002). By the convolution theorem, multiplying (modulating) the tidal component in the time domain by

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

a function is equivalent to convolving the Fourier transform of the tidal component with that of the modulating function in the frequency domain. For example, a purely sinusoidal high-frequency tide $\cos(\omega_T t + \phi)$ may be multiplied by an amplitude envelope $A(t)$ such that:

$$\eta(t) = A(t)\cos(\omega_T t + \phi), \tag{14}$$

the Fourier transform of which is

$$\begin{aligned} \mathcal{F}\{\eta(t)\}(\omega) &= \mathcal{F}\{A(t)\}(\omega) \otimes \mathcal{F}\{\cos(\omega_T t + \phi)\}(\omega) \\ &= \tilde{A}(\omega) \otimes \left(\mathrm{e}^{-i\phi}\delta(\omega - \omega_T) + \mathrm{e}^{i\phi}\delta(\omega + \omega_T)\right)\sqrt{\frac{\pi}{2}}, \end{aligned} \tag{15}$$

where $\mathcal{F}\{\cdot\}(\omega)$ is the Fourier transform to angular frequency (ω) domain, $\tilde{A}(\omega)$ is the Fourier transform of $A(t)$, and $\otimes$ denotes convolution. The constant $\sqrt{\pi/2}$ is a matter of convention such that the inverse Fourier transform is symmetric with the forward transform.

To generate the synthetic time series used in this study, a spectrally-red modulating function is used to simulate the cusps observed around tidal lines in the spectra of tidally-dominated ocean time series, and a random phase is assigned at each frequency before an inverse Fourier transform is applied to produce a synthetic time series with known spectral power. Samples from a red background are added to the time series to simulate broadband non-tidal ocean variability. Figure 3 shows an example of three synthetic spectra (upper panel) and corresponding hourly time series computed with random phase and truncated at 500 hours (bottom panel). The top panel also shows the power spectrum of observations in gray for comparison (these are discussed in Section 4a). The phase at each frequency is identical for each time series, so only the spectral power differs. All three have peaks of equal magnitude at the M_2 frequency. The red background spectrum is proportional to $1/\omega^2$ while the modulating spectrum is proportional to $1/(\omega^2 + \omega_0^2)$, where ω_0 is a small frequency introduced to eliminate the singularity at $\omega = 0$. In the red-background time series, there is more variability at low frequencies as in the real ocean, while the modulated one imitates

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-00041.

the interaction of the tide with low-frequency processes. The time series with the spectrally white background has comparatively more power at super-tidal frequencies, which is evident in the time domain from the short time-scale noise that is less noticeable in the data with a red background. The data displayed are sampled hourly to be consistent with the choice of the Nyquist frequency, while the length of the time series is set by the choice of the fundamental frequency.

c. Effect of Noise Covariance on Amplitude, Phase, and Uncertainty

The estimated amplitude and phase of the tidal constituents are affected by the spectrum used to construct **R**, also referred to here as the noise covariance or residual covariance. At each frequency represented in **H**, the data are represented twice: explicitly in the frequency domain as model parameters ($\hat{\mathbf{x}}$) and implicitly in the residual ($\hat{\mathbf{r}}$) in the time domain, which has energy at all frequencies in general. Here, "explicit" versus "implicit" refers to whether energy is described by the model sines and cosines or by a residual time series with the fitted tides removed. Choosing the residual covariance matrix **R** to approximate the true covariance of the non-tidal component of the data improves the accuracy of the estimated coefficients. This is important because geophysical time series generally do not have flat spectra but rather spectra that decrease with increasing frequency. For example, if data with a red noise term ($S_{\text{noise}} \propto f^{-2}$, like the red and blue curves in Figure 3) were modeled using a spectrally white **R** (equivalent to the black curve), the variance would be treated disproportionately as signal at low frequencies (where the true noise is actually more energetic than what is given by **R**) and disproportionately as noise at high frequencies (where the true noise is actually less energetic than what is given by **R**). This is the case even if the spectrum corresponding to **R** has the same total energy as the true background.

Here we analyze a 1001-point synthetic record sampled hourly. If there is a low-amplitude, high-frequency signal present in the data, a high assumed noise level at that frequency would

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

limit the detection of that constituent, as shown by the large blue intervals in Figure 4A at $f > 2$ cpd. A colored spectrum that matches the frequency dependence of the true noise component is an improvement over the simpler assumption of an uncorrelated (spectrally flat) white noise that does not match the true noise component. The former gives a relatively constant ratio of assumed spectrum to true spectrum across frequencies, while the latter gives a frequency-dependent ratio. This results in the variance of **Hx** being overrepresented at low frequencies, because the noise covariance is too low, and underrepresented at high frequencies, because the noise covariance is too high. Because variance is represented twice at modeled frequencies, the covariance matrices **R** and **P** serve as constraints on the partition of energy between **Hx** and **r** (recall that **R** represent covariance in time of a wide-sense stationary noise component and therefore has power at all frequencies, including those to which data are fitted).

Figure 4 shows the effects of these constraints in the frequency domain ("model space"): as described in section 3b and illustrated in Figure 3, a synthetic time series is constructed by adding a random background process with a spectral slope of -2 to a modulated semidiurnal (M_2) tidal process. In Figure 4A, the total spectrum (solid black) and its tidal component (signal, dashed black) and background (noise, dotted black) are plotted along with red_tide squared amplitude estimates ($\hat{a}^2 + \hat{b}^2$) calculated using a spectrally white noise covariance (blue), an approximate red noise covariance (red), and a "true" red noise covariance that matches the spectral power of the noise component (orange). All assume the same model parameter covariance matrix **P** constructed from the true tidal spectrum, not the total spectral density. Shading indicates 90% credible intervals of amplitude. These intervals are calculated using Monte Carlo sampling from a 10,000-member population obeying the posterior error of $\hat{\mathbf{x}}$ (equation (8)) to infer the distribution of $\hat{A}$. Figure 4B shows again the spectrum of the non-tidal (noise) component of the time series (dotted black line), the spectra corresponding to the noise covariance matrices (solid lines), and the posterior spectra

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI: 10.1175/JTECH-D-21-0034.1.

of the residual time series (connected dots) calculated using Welch's method for spectral estimation (overlapping windowed segments of length one fifth that of the record). For each noise covariance, the total noise energy is approximately 42% of the total signal energy integrated from $f = \Delta f$ to f_{Ny}. This matches the true noise-to-signal ratio. Both red noise analyses (red and orange lines in Figure 4B), which more closely approximate the true noise component, result in more accurate amplitude estimates at low frequencies where the signal-to-noise ratio is low (red or orange lines vs black dashed line in Figure 4A at frequencies less than 0.2 cpd), and greater precision at higher frequencies, where the signal-to-noise ratio is high. The calculated spectra of the residual time series are lower for increasingly accurate noise spectra, indicating that more variance is allocated to the model parameters.

Figure 5 compares the performance of the algorithm for both amplitude and phase under these three noise regimes. The ratio of median estimated amplitude ($\hat{a}^2 + \hat{b}^2$) to the true signal amplitude is displayed in Figure 5A, which shows that all configurations produce accurate estimates near the tidal peak (2 cpd). At frequencies higher than about 5 cpd, the assumed variance of noise and model parameters are similar in magnitude (Figure 4B). In this frequency range, the amplitude ratios assuming red noise in Figure 5A show high variability with average values closer to one, indicating that they are less biased than the white-noise assumption, which is always greater than 1 for f>0.5 cpd, but also less precise, because the amplitudes under the red noise assumption vary more above and below 1. Amplitude uncertainty is measured by the standard deviation of the estimated distribution of modeled amplitude, normalized by the true amplitude of the tidal component (Fig. 5B). The standard deviations of the amplitudes are smaller for the red noise assumptions than for the white noise assumption for all frequencies, and especially near the tidal peak and its cusp. The assumption of full, unapproximated red noise (orange dots in Fig. 5B) gives the lowest standard deviation. For the phase, all methods give comparable estimates (Fig.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

5C). At higher frequencies, both red noise results are more precise than the white noise result. The uncertainty of the phase (Fig. 5D), here measured by the standard deviation of the estimated distribution of the centered phase, is lower near the tidal cusp and at higher frequencies for the red noise results. The standard deviation is lower for the white noise result at frequencies lower than 0.4 cpd, but this implies unreasonable precision, as suggested by the inaccurate phase estimates at these frequencies in panel C resulting from a signal-to-noise ratio of less than 1; the corresponding precision is too high to be reasonable for such inaccurate estimates. At these low frequencies, the standard deviation of phase for the red noise analyses approaches $\pi/\sqrt{3}$, indicated by the horizontal dotted line, which is the standard deviation of a uniform distribution on the interval $(-\pi, \pi)$. This is expected for the phase of a low-amplitude signal in the presence of energetic noise that renders that signal's phase unrecoverable. At frequencies higher than the tidal cusp, the standard deviation of phase from the white noise analysis approaches this value because it assumes that noise is more energetic than the signal, while the red noise assumptions give results with lower variability, with the full red noise assumption giving the lowest variability.

The lessons learned from the idealized case of a short record with a tidal cusp and red noise also hold for a more complicated and realistic case in which a long record with a proportionately finer frequency resolution is modeled only at frequencies in limited bands. In Figures 4 and 5, all resolvable frequencies are modeled to illustrate the effect of noise representation on the partition of variance into signal and noise. In practice, however, tidal time series may be several years long, and so the number of data, and hence the number of resolvable frequencies, can be large enough such that modeling all frequencies explicitly is computationally prohibitive. Furthermore, in such cases the residual covariance matrix $\mathbf{R}$ cannot be practically constructed at all possible time lags, necessitating instead an approximation like that discussed earlier (truncated, in sparse form, and windowed to reduce spectral ringing). Figure 6 shows results analogous to those in

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

Figure 4 for an hourly-sampled seven-year-long synthetic time series. In this case, only limited frequency bands are modeled, with the rest of the total variance accounted for by the residual. An accurate model parameter covariance enforces realistic amplitudes and uncertainty on the model parameters by setting a signal-to-noise ratio that approximates the true ratio of tidal energy to non-tidal energy as a function of frequency and not only the total, frequency-integrated ratio. This corresponds to the red curve in Figure 6. The blue curve corresponds to an assumption of white noise: though the total frequency-integrated signal-to-noise ratio is the same for both estimates, the white noise assumption over-represents the variance of the non-tidal component in the semi-diurnal band (around 2 cpd), giving larger credible intervals resulting from the unnecessary uncertainty introduced by overestimating the non-tidal variance. Though the red noise covariance is a windowed, truncated approximation of the true non-tidal covariance, its spectrum (dotted red) closely matches the true underlying spectrum of non-tidal variability (dashed black). With only 500 frequencies modeled out of the over 60,000 that are fully resolvable, the reconstructed time series accounts for more than 45% of the total variance and estimates tidal and off-tidal amplitudes within error. This shows that modeling all resolvable frequencies directly in $\hat{\mathbf{x}}$ is not only computationally impractical but also unnecessary, even for data with substantial non-tidal energy like those in Figure 4.

d. Effect of Noise Covariance and Record Length on Constituent Estimates

In order to evaluate the impact of the choice of noise covariance and record length on estimated tidal coefficients, we ran ten Monte Carlo experiments. These ten experiments used varying combinations of tidal energy (two regimes) and background structure (five regimes), which are described below. The six runs with more negative spectral slopes ($\gamma = -3, -2.5, -2$, corresponding to panels A-C and F-H in Figures 7 and 8) used 10,000 sample time series, while the four runs

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

with less negative spectral slopes ($\gamma = -1.5, -1$, corresponding to panels D, E, I, and J) used 50,000 sample time series to obtain stable results. Each time series had 24,000 hourly samples but was only analyzed up to 1500 hours in order to investigate the effect of record length on estimated parameters, with the exception of the $\gamma = -1$ case, which was analyzed up to 5000 hours in order to examine convergence at longer record lengths. Three tidal constituents typically seen in observations (K_1, M_2, and S_2) are added onto a synthetic noise background. Analysis for only the M_2 constituent is shown, as other tidal constituents are qualitatively similar. Only phase varies from one time series to another within a run, with signal and noise amplitude constant for all ensemble members. The phase is varied by randomly selecting from a uniform distribution on the interval $[0, 2\pi)$, and is varied separately for the noise and signal components. Two quantities derived from model parameters are examined to determine the impact of noise covariance and record length: the bias of tidal amplitude estimates as a measure of accuracy, and the variance of model parameters (harmonic coefficients) about the true values as a measure of precision.

The observed bias of estimated tidal amplitude $\hat{A}$ is affected by the variance of the tide relative to the background time series, the record length, and the choice of noise covariance. Figure 7 shows the mean ratio of the estimated M_2 constituent amplitude to true amplitude across the ten Monte Carlo experiments:

$$\text{Amplitude Ratio} = \frac{\left\langle \sqrt{\hat{a}^2 + \hat{b}^2} \right\rangle}{\sqrt{a^2 + b^2}}, \tag{16}$$

where brackets indicate averaging across Monte Carlo simulations, a and b are respectively the sine and cosine coefficients of the tidal constituent in question (here M_2), and hats denote estimates. Note that a and b vary with each Monte Carlo simulation, but the sum of their squares does not, hence $\sqrt{a^2 + b^2}$ can be pulled out of the brackets. The amplitude ratio is a measure of the accuracy of tidal harmonic estimates in terms of a simple quantity of physical relevance, tidal amplitude.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

Each panel shows this quantity as a function of the assumed noise spectral slope and record length. The panels are organized with steep spectral slopes at the top and successively flatter spectral slopes below. The left column (A-E) represents a strong tide case meant to simulate bottom pressure, with tidal spectral power 100 times that of the noise background at tidal frequencies, and the right column (F-J) shows a weak tide case meant to simulate tidal velocities or baroclinic tides, with tidal spectral power 10 times that of the local background. A perfect tidal amplitude estimate would have a normalized value of 1, and cases for which the interquartile range of simulations do not include 1 are hatched in Figure 7, as they do not provide accurate estimates of the tidal amplitude. Records with steeper spectral slope background processes (Figure 7 A-B, F-G) have less accurately estimated amplitudes than those with less steep background spectral slope (Figure 7 D-E, I-J), while the latter required more Monte Carlo simulations to show convergence to a record length-dependent amplitude accuracy. Additionally, records with less spectrally-steep backgrounds are less sensitive to the choice of noise covariance. In all cases, the use of a noise spectral background that matches the spectral structure of the data yields a more accurate estimate of amplitude than the use of a noise spectral background with a substantially different slope. These results are in agreement with the expectation that a more energetic and structured background process is best described by a residual component with a covariance matrix **R** constructed from an appropriate power spectrum. Amplitude estimates also improve with increasing record length, though this effect is smaller than the effect of the noise covariance.

Estimated amplitudes are comparably accurate and in some cases more accurate for spectral slopes less than (steeper than) the true slope when compared to the amplitude accuracy when using the true slope to construct **R**. This is, however, a small effect that appears to diminish with increasing record length. For example, in Figure 7 E and J, an assumed spectral slope of -1.5 achieves more accurate results than the true slope of -1, which gives lower amplitude ratios, an

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

effect that is stronger at shorter record lengths. This may be due to the bias of the estimators $\hat{a}$ and $\hat{b}$: the squared error of $\hat{a}$ and $\hat{b}$ are minimized in the least squares procedure and these estimators are biased low due to regularization (Hoerl and Kennard 1970; Wunsch 1996). The reduction of parameter error at the expense of added bias inherent to this method is evident in the difference between amplitudes estimated from OLS (noise slope "None") and the regularized methods. In particular, the use of a white noise covariance (noise slope 0) differs from OLS only by its inclusion of the regularizing term **P** (noise slope of 0 means that **R** is a scaled identity matrix that can be treated as a constant and would be canceled out if **P** were 0). OLS overestimates tidal amplitudes consistently while the regularization of the otherwise identical white noise approach typically underestimates tidal amplitudes except for the case of strong tides and a spectrally steep background (Figure 7 A-B) for which such a white noise assumption is not suitable.

Model parameters $\hat{\mathbf{x}} = [\hat{a}_1, \hat{b}_1, \hat{a}_2, \hat{b}_2 ...]$ are estimators for the true parameters $\mathbf{x}$ and, like amplitude $\hat{A}$ derived from them, better estimate truth with increasing record length, a more accurate noise covariance, and a higher signal-to-noise ratio. As a measure of precision of these parameter estimates, we calculate their variance about the true mean, normalized by the true amplitude:

$$\text{Normalized Parameter Variance} = \frac{1}{2\sqrt{a^2+b^2}}\left(\left\langle(\hat{a}-a)^2\right\rangle + \left\langle(\hat{b}-b)^2\right\rangle\right). \tag{17}$$

The base-10 logarithm of this quantity is shown in Figure 8 in order to better compare results across orders of magnitude. Due to the random phase, the statistics of sine and cosine coefficients are identical, constrained by $a^2+b^2=A^2$, where A is constant across all Monte Carlo simulations. Therefore, both $\hat{a}$ and $\hat{b}$ are used in calculating the normalized parameter variance, effectively doubling the number of realizations.

As with amplitude, when the random background has a more negative spectral slope, estimates of tidal coefficients are more sensitive to noise covariance choice (Figure 8 A-B, F-G) than

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

they are when the background has a less negative spectral slope (Figure 8 D-E, I-J). Unlike amplitude, coefficients are primarily sensitive to record length when the background noise is relatively unstructured, with increasing sensitivity to the choice of noise covariance as the true noise spectrum becomes steeper. This can be seen in the gradient of the normalized parameter variance in each panel: in panels A-B and F-G, where γ is more negative, the variance decreases for more accurate noise slopes, whereas in panels D-E and I-J, where γ is less negative, the variance decreases more strongly with increasing record length. Filled dots are placed at elements corresponding to the lowest value (highest precision) in their respective rows. For all simulations with a background process of spectral slope $\gamma \geq -2$ (less steep), precision is highest for the noise assumption that matches the true spectral slope. For $\gamma < -2$ (steeper), this is also the case for sufficiently long records, and for short records it is still better to match the noise spectrum than to use OLS.

In almost all cases, using equation (7) (with **P** and **R**) instead of equation (9) (OLS without an explicit noise assumption, labeled "none" in the rightmost column of every panel in Figures 7 and 8) resulted in more accurate estimates for A (ratios closer to 1 in Figure 7) and more precise estimates for **x** (lower parameter variance in Figure 8), at least for the cases examined here in which tidal energy is 10 to 100 times greater than that of noise at the same frequency. The OLS approach, which is widely implemented in tidal harmonic analysis, was comparable to the red_tide approach for $\gamma = -1.5$ and $\gamma = -1$, indicating that it is most suitable for tidal analysis of data with a spectral background that is nearly white. Data with steeper background spectra benefit from treating the residual **r** as a spectrally red process by way of the covariance matrix **R**.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

4. Application to Oceanographic Data

a. Bottom Pressure

The methods implemented in red_tide seek to address several potential issues that can arise when harmonically analyzing real tidal time series, each of which are presented in isolation in section 3 by using synthetic data. The first demonstration of red_tide on observations uses a bottom pressure record, a data set with low background noise where record length is sufficient and non-phase-locked energy is much weaker than the phase-locked tide. Bottom pressure measurements from the NOAA Deep-ocean Assessment and Reporting of Tsunamis (DART) (NOAA, 2005) are dominated by the largely coherent barotropic tide. The time series examined here originates from site 51406 (8.48°S, 125.03°W) and spans 3 years and 6 months of observations, from February 12, 2011 to September 6, 2014. Many tidal constituents of amplitudes spanning several orders of magnitude are present in this record, as seen by tidal lines in Figure 9. Pressure measurements have a significantly lower noise level than coastal surface height gauges and have accurate harmonic constituents even over short records (Le Provost 2001). Bottom pressure is therefore useful for evaluating the accuracy of harmonic decomposition in the regime of tidally-dominated, low noise observations, which typically do not pose major problems when calculating constituents. Hourly averaging of 15-second sampled data further suppresses noise and instrumental artifacts, such as digitization, and does not alias major constituents.

The power spectrum of the bottom pressure time series (Figure 9) exhibits many prominent peaks, of which 22 are singled out for analysis. These 22 frequency bands together account for more than 99% of the time series' variance; therefore both the white noise assumption (not shown) and red noise assumption produce essentially identical results. Figure 10 shows the output from red_tide using only the red noise assumption for $\mathbf{R}$ with spectral slope $\gamma = -3/2$ alongside t_tide

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

output. Where t_tide models only tidal frequencies associated with astronomical parameters, this analysis includes those same frequencies and 30 additional frequencies per constituent in a band of 10 year^{-1} centered on that tidal frequency in order to capture modulation at annual and semiannual cycles and cusp-like spreading of peaks. These harmonic amplitudes are spaced at $\Delta f = 1/3$ year^{-1}, a frequency step smaller (and hence of higher resolution) than that of the power spectrum, which is coarser in resolution due to segmenting. This corresponds to a 3-year period, shorter than the record length of 3 years and 206.5 days in order to ensure the resolution of annually periodic modulation of the main tidal constituents, regardless of the exact record length. The amplitudes are normalized to have units of spectral power for comparison. The results of analysis by t_tide are also normalized and plotted with 90% confidence intervals (the t_tide code that defaults to 95% confidence intervals is modified).

The tidal amplitudes given by t_tide largely match the red_tide results for high-amplitude constituents and fall within credible intervals at low-amplitude constituents, while providing analysis at more frequencies. Focusing on a single cluster of constituents shows this more clearly (Figure 11). The semidiurnal band, centered about the energetic M_2 constituent, contains several other well-resolved tidal constituents (Darwin symbols 2"N_2, μ_2, N_2, ν_2, λ_2, L_2, S_2, and K_2) that result from the complicated gravitational tidal forcing potential, many of which are not resolvable when using segmenting methods or are not exactly mapped by a simple Fourier transform, which would result in spectral leakage. Additionally, the energy of the cusps is explicitly modeled, with intervals that reflect the uncertainty associated with a realistic noise level that is a function of frequency, which becomes important for these low amplitude components. The characteristics of the non-phase-locked component of the tides can therefore be diagnosed from these cusps. The high signal-to-noise ratio at energetic constituents, on the other hand, means that the noise level is less important at these frequencies, hence the OLS approach of t_tide works well for these constituents.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

b. High-Frequency Radar

Like OLS, the red_tide package effectively models data with weak noise and tidal constituents that are highly coherent with astronomical forcing. The data regime for which red_tide is designed includes higher levels of structured noise and tidal energy that cannot be predominantly described by a small number of frequencies. Surface currents, which are driven by wind, tides, eddies, and mean flow, fall under this category.

Observations of surface currents are obtained from a high-frequency radar network (HFRnet, see Terrill et al. 2006). The California Current System (CCS), a region that is well-sampled by this network (Roarty et al. 2019), is used to evaluate the harmonic decomposition technique. Radial velocities measured by antenna stations are mapped to a Cartesian grid of zonal and meridional velocities using a least squares fit (Ohlmann et al. 2007). Surface currents are driven by a wide range of dynamics: direct wind forcing, near-inertial motions, interannual variability of the local current system, and tides, including tidal currents and the surface expression of internal tides. This contrasts with bottom pressure, which is dominated by tides, with many more prominent tidal frequencies than appear in surface currents.

Figure 12 shows the averaged rotary power spectrum (spectral power partitioned by rotational polarization) over the grid points in a region of the CCS ranging from 33.7561°N to 38.1252°N out to approximately 100 km from the coast (for the formulation of rotary spectra, see Gonella 1972). This formulation is convenient for visualizing the spectral power of surface currents due to the polarized flow resulting from near-inertial oscillations, which at these latitudes occur at frequencies between 1.12 to 1.26 cycles per day. Tidal peaks in HFR surface currents are pronounced and are comparable to or greater than the energy in the inertial and low-frequency (<0.4 cpd) bands. These spectra are calculated using Welch's method, and the error bar denotes the ratio of high to low

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

estimates for the power spectrum for a confidence level of 95%; this ratio is constant on logarithmic axes and does not vary with frequency. It is calculated from a chi-squared distribution for degrees of freedom equal to twice the number of windowed segments whose periodograms are averaged to calculate the spectrum (e.g. Bendat and Piersol 2010), divided by 9 from the assumption that neighboring sites are correlated.

An hourly-sampled, 9 year and 3 month-long (January 1, 2012 to April 1, 2021) HFR surface current time series from 35.5361°N, 121.1776°W, roughly 7 kilometers off the coast of San Luis Obispo County, California, is analyzed in four frequency bands: low frequency (0.00732 cpd and less), S_1 solar diurnal (1 cpd), M_2 lunar semidiurnal (1.932 cpd), and S_2 solar semidiurnal (2 cpd). These frequency bands, shown in Figure 13A, account for roughly a fifth of the variance of the time series for both zonal velocity u and meridional velocity v. Therefore, 80% of the variance is included in the residual time series. Harmonic coefficients from red_tide are calculated using a model parameter covariance built from the domain-averaged power spectrum (solid black line) from 1191 grid points at latitudes ranging from 33.7561°N to 38.1252°N, out to about 100 kilometers offshore. The individual power spectrum calculated from the analyzed time series is shown in gray for comparison: smoothed peaks due to low frequency resolution and spectral leakage due to sampling and windowing result in an individual time series' spectrum that does not capture features that the least squares approaches can. The noise (residual) spectrum has a constant spectral slope of -1 and a corresponding covariance truncated at 300 hours lag, resulting in modest spectral ringing (dashed line). Amplitudes given by red_tide are calculated as half the sum of the sine- and cosine-coefficients squared, $\frac{1}{2}(a^2+b^2)$, and are shown in red with 90% credible intervals, while equivalent amplitudes and 90% confidence intervals given by t_tide are shown in blue. Filled blue circles are constituents that t_tide considers "significant", where the amplitude is greater than the amplitude error (Pawlowicz et al. 2002), while hollow circles are not

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

considered significant. Only results for u are shown, as results for v are qualitatively similar; the full 2-dimensional character of time series in this region is shown in Figure 12 to illustrate the frequency-dependent polarization of surface currents, which may be modeled with red_tide. The calculation of rotary coefficients from red_tide output for u and v is straightforward, though results are not shown.

As with bottom pressure, harmonic decomposition reveals sharp peaks at tidal frequencies that are not resolved in power spectra with lower frequency resolution due to averaging over time intervals shorter than the record length. Because the choice of basis functions for red_tide is arbitrary, the fundamental frequency of the data set does not necessarily limit the spacing of modeled frequencies, though in practice the tolerance of basis nonorthogonality and resulting uncertainty will limit the choice of frequencies. The high noise level in these data results in large uncertainty at frequencies around tidal peaks as in Figure 13B, as indicated by shaded intervals. Despite this, annual modulation of the surface current is evident by the second and third most energetic peaks in the M_2 band appearing at $f_{M_2} \pm f_{\text{annual}}$, indicated by vertical dashed lines. The output of t_tide is also able to discern these modulation peaks, though it does not examine broadband modulation.

5. Summary and Discussion

The methods outlined here and implemented in red_tide are intended to provide best estimates of tidal amplitudes for data with red background spectra and significant tidal cusps. Red_tide incorporates a red noise covariance and includes additional frequencies beyond those of the astronomical forcing to accommodate data with highly energetic and correlated non-tidal components, a weak tidal signal relative to non-tidal processes, or a modulated tidal component with energy distributed across a band of frequencies. Short records, for which long-period variance appears trend-like, also benefit from these methods because variance at fitted frequencies is allocated to model parameters

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-00034.1.

according to prior statistical assumptions. The spectrally colored noise covariance is constructed to approximate the spectral properties of the non-tidal component of the data, and may be truncated and represented as a sparse matrix for computational efficiency with a window function applied to off-diagonal elements to suppress spectral artifacts that result from truncation. On the other hand, highly coherent tidal records with well-defined peaks and small cusps, such as bottom pressure, are well described by OLS, as the background noise is orders of magnitude lower in amplitude than the tidal signal.

These methods also address time series for which the choice of model covariance impacts results. We demonstrate this with a step function, a simple case that exhibits an extreme mismatch between the fit and the data when an inappropriate model parameter covariance is used, resulting in the well-known Gibbs phenomenon. We have found that when fitting a discrete step function across resolvable frequencies, the assumption of a realistic covariance reduces the magnitude of the Gibbs phenomenon near the jump discontinuity when compared to an assumption of constant covariance matrix $\langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle = \sigma_{\mathbf{x}}^2 \mathbf{I}$, and prevents dramatic overshoot when fitting to frequencies greater than the Nyquist frequency, which would otherwise be aliased (Figure 2 B-E).

The accuracy and precision of model parameters, given explicitly as a posterior covariance matrix, are impacted by the choice of residual (noise) covariance. Spectrally colored time series may have a residual background that varies over orders of magnitude across tidal bands, necessitating an appropriate noise covariance matrix if all constituents are to be estimated optimally. When the user-specified residual spectral power is significantly lower than the total spectral power of the time series, red_tide allocates more variance to the model parameters. If the energy of the non-tidal component is well understood, variance can be realistically allocated between the estimated signal $\mathbf{H}\hat{\mathbf{x}}$ and noise $\hat{\mathbf{r}}$ at the same frequencies. Estimating tidal amplitudes can be computationally expensive for long records ($O(10^5)$ or more data) because of the memory requirements for constructing

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

and multiplying large matrices. In these cases, we have identified strategies for approximating the covariance matrix in order to reduce computational requirements, including truncating and windowing the matrix in sparse form.

In summary, red_tide was designed to estimate tidal coefficients while incorporating prior assumptions that accurately account for the spectral structure of underlying noise and allow flexibility in the choice of modeled frequencies, which is important for data with a modulated, non-phase-locked tidal component. Longer records and less strongly correlated noise benefit less from this flexibility. Ordinary least squares is comparatively less suited to computing tidal harmonics from data with spectrally colored noise, especially red noise with a steep spectral slope. The code is available for use and modification, the details of which are in the appendix.

Acknowledgments. This work has been supported by a Future Investigators in NASA Earth and Space Science and Technology award (80NSSC19K1342). In addition, Luke Kachelein, Bruce Cornuelle, Sarah Gille, and Matthew Mazloff acknowledge support from the NASA Surface Water and Ocean Topography Science Team (awards NNX16AH67G and 80NSSC20K1136).

We would also like to thank Edward Zaron for his guidance on the evolving nomenclature for modulated tidal phenomena, as well as the anonymous reviewers for their thoughtful comments and suggestions that improved the quality and readability of this study.

Data availability statement. Relevant data and scripts, including those used to generate all figures and results in this paper, are publicly available through the University of California, San Diego library digital collections at `https://doi.org/10.6075/J080515G`. Readers interested in accessing observational data may visit `https://doi.org/10.7289/V5F18WNS` for bottom pressure data and `https://dods.ndbc.noaa.gov/thredds/hfradar.html` for high-frequency radar surface current data.

APPENDIX

a. Downloading red_tide

The red_tide package is available for download as a GitHub repository at `https://github.com/lkach/red_tide` and in archived form (see Data Availability Statement). This package is written in the MATLAB language, but translation to other programming languages is welcome and encouraged. It has also been designed to work in the free software GNU Octave language and Octave-specific instructions are provided with the software release. Input for red_tide is flexible, with several options and default settings.

b. Noise as an Autoregressive Process

The residual time series $\hat{\mathbf{r}} = \mathbf{y} - \mathbf{H}\hat{\mathbf{x}}$ will have a power spectrum that is similar to that of $\mathbf{y}$, except in the bands of modeled frequencies. With the energetic tidal components removed, a spectrally-red residual can be modeled as an autoregressive process AR(p), where order p is the maximum number of time steps for which the system has memory (von Storch and Zwiers 2003):

$$r_t = \alpha_0 + \sum_{k=1}^{p} \alpha_k r_{t-k} + \epsilon_t, \tag{A1}$$

represents an order-p AR process r at time t with a white noise component ϵ. The AR parameters α_k can be estimated from the Yule-Walker equations (e.g. von Storch and Zwiers, 2003). From these, the spectral density of r can be estimated by

$$\hat{S}_{\mathbf{rr}}(f) = \frac{\hat{\sigma}_\epsilon^2}{\left| 1 - \sum_{\ell=1}^{p} \hat{\alpha}_\ell \mathrm{e}^{-2\pi i \ell f} \right|^2}, \tag{A2}$$

where hats indicate estimates of a true value, f is frequency, and σ_ϵ^2 is the variance of the white noise ϵ. The spectrum of the AR-modeled residual time series can be used as an estimate of the

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

spectrum of underlying non-tidal, non-wind-driven intrinsic variability in the ocean. If $\hat{S}_{\mathbf{rr}}(f)$ accurately models the spectrum of $\mathbf{y}$ at frequencies outside those in $\mathbf{H}$ and assuming white noise ϵ, the coefficients $\{\hat{\alpha}_k\}$ may be used to construct $\mathbf{R}$ for a second iteration of fitting using this new estimate for the noise covariance matrix. This can be done by taking the inverse Fourier transform of $\hat{S}_{\mathbf{rr}}(f)$, which is the covariance of $\mathbf{r}$ via the Wiener-Khinchin theorem. $\mathbf{R}$ is then constructed from the covariance as outlined in section .

c. Alternative Form of Equation 7

An alternative equation that solves for $\hat{\mathbf{x}}$ can be achieved using the matrix inversion lemma (e.g. Wunsch 1996):

$$\hat{\mathbf{x}} = \mathbf{P}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}\mathbf{y}. \tag{A3}$$

In practice, tidal records usually contain many more data points N than frequencies M that are of interest to model. Therefore the data-space inversion (equation (7)) is used in red_tide because $\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$ has lower computation and memory requirements ($4M^2$ elements) than $\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}}$ (N^2 elements). The $N \times N$ residual covariance matrix inverse, $\mathbf{R}^{-1}$, that appears in equation (7) does not need to be computed explicitly when using efficient linear system solution algorithms for matrix inversion, which instead directly calculate $\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$, which is $2M \times N$. Further, a sparse representation of $\mathbf{R}$ minimizes memory requirements in equation (7) compared to the more challenging requirements of the dense $N \times N$ matrix in equation (A3).

d. Non-dimensionalization

The assumption of uncorrelated noise is physically unrealistic when modeling processes with a correlated, energetic noise component, which would be better represented by a more complicated $\mathbf{R}$. For example, Coles et al. (2011) use a non-diagonal matrix expression for $\mathbf{R}$, the efficient

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

inversion of which is outlined here. The computational resources to directly invert a non-diagonal $\mathbf{R}$ are too high to be practical: $\mathbf{R}$ is $N \times N$, where N is the length of $\mathbf{y}$, which is much longer than $\mathbf{x}$ in practice. Though red_tide uses MATLAB's default linear equation solving method, a Cholesky lower triangle factorization of the residual covariance matrix $\mathbf{R} = \langle \mathbf{r}\mathbf{r}^{\mathrm{T}} \rangle$ may also be used. The residual covariance can be factored as $\mathbf{R} = \mathbf{R}^{1/2}\mathbf{R}^{\mathrm{T}/2}$, where $\mathbf{R}^{1/2}$ is lower triangular with the inverse $\mathbf{R}^{-1/2}$ used as a (non-unique) whitening transform:

$$\mathbf{y_w} = \mathbf{R}^{-1/2}\mathbf{y}, \qquad \mathbf{r_w} = \mathbf{R}^{-1/2}\mathbf{r}, \qquad \mathbf{H_w} = \mathbf{R}^{-1/2}\mathbf{H}. \tag{A4}$$

Noting that $\mathbf{R}^{\mathrm{T}/2}\mathbf{R}^{-1}\mathbf{R}^{1/2} = \mathbf{I}$, these can be substituted into equation (7), which simplifies to

$$\hat{\mathbf{x}} = \left(\mathbf{H_w}^{\mathrm{T}}\mathbf{H_w} + \mathbf{P}^{-1}\right)^{-1} \mathbf{H_w}^{\mathrm{T}}\mathbf{y_w}. \tag{A5}$$

The choice of $\mathbf{R}$, and thus $\mathbf{R}^{-1/2}$, relies on accurately estimating the residual covariance. This may be done by examining the calculated data covariance matrix $\mathbf{y}\mathbf{y}^{\mathrm{T}}$ or equivalently the power spectrum $S_{\mathbf{yy}}$ and estimating $\langle \mathbf{r}\mathbf{r}^{\mathrm{T}} \rangle$.

Analogously, the model parameter covariance matrix $\mathbf{P}$ can be expressed as $\mathbf{P} = \mathbf{P}^{1/2}\mathbf{P}^{\mathrm{T}/2}$. Define $\mathbf{H'} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}^{1/2}$ such that equation (A5) can be written as

$$\hat{\mathbf{x}} = \mathbf{P}^{1/2}\left(\mathbf{H'}^{\mathrm{T}}\mathbf{H'} + \mathbf{I}\right)^{-1} \mathbf{H'}^{\mathrm{T}}\mathbf{y_w}. \tag{A6}$$

This is equivalent in form to the regularized least squares problem with a left-multiplied factor of $\mathbf{P}^{1/2}$ that allows $\hat{\mathbf{x}}$ to retain units of $\mathbf{y}$.

References

Agnew, D. C., 1992: The time-domain behavior of power-law noises. *Geophysical Research Letters*, **19 (4)**, 333–336, doi:10.1029/91GL02832.

Bendat, J., and A. Piersol, 2010: *Random Data*, chap. 12, 417–472. John Wiley & Sons, Ltd, doi:10.1002/9781118032428, ISBN:9781118032428.

Chavanne, C. P., and P. Klein, 2010: Can oceanic submesoscale processes be observed with satellite altimetry? *Geophysical Research Letters*, **37 (22)**, doi:10.1029/2010GL045057.

Chiswell, S. M., 2002: Energy Levels, Phase, and Amplitude Modulation of the Baroclinic Tide off Hawaii. *Journal of Physical Oceanography*, **32 (9)**, 2640–2651, doi:10.1175/1520-0485(2002)032<2640:ELPAAM>2.0.CO;2.

Codiga, D. L., 2011: Unified tidal analysis and prediction using the UTide matlab functions. Tech. rep., Graduate School of Oceanography, University of Rhode Island, Narragansett, RI. URL ftp://www.po.gso.uri.edu/pub/downloads/codiga/pubs/2011Codiga-UTide-Report.pdf, 59pp.

Coles, W., G. Hobbs, D. J. Champion, R. N. Manchester, and J. P. W. Verbiest, 2011: Pulsar timing analysis in the presence of correlated noise. *Monthly Notices of the Royal Astronomical Society*, **418 (1)**, 561–570, doi:10.1111/j.1365-2966.2011.19505.x.

Doodson, A. T., and H. Lamb, 1921: The harmonic development of the tide-generating potential. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **100 (704)**, 305–329, doi:10.1098/rspa.1921.0088.

Eich, M. L., M. A. Merrifield, and M. H. Alford, 2004: Structure and variability of semidiurnal internal tides in Mamala Bay, Hawaii. *Journal of Geophysical Research: Oceans*, **109 (C5)**, doi:https://doi.org/10.1029/2003JC002049.

Foreman, M. G. G., 1977: *Manual for tidal heights analysis and prediction. Pacific Marine Science Rep. 77-10*. Patricia Bay, Sidney, BC, Canada, Institute of Ocean Sciences, (2004 revision).

Foreman, M. G. G., J. Y. Cherniawsky, and V. A. Ballantyne, 2009: Versatile harmonic tidal analysis: Improvements and applications. *Journal of Atmospheric and Oceanic Technology*, **26 (4)**, 806–817, doi:10.1175/2008JTECHO615.1.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-00034.1.

Foreman, M. G. G., and R. F. Henry, 1989: The harmonic analysis of tidal model time series. *Advances in Water Resources*, **12**, 109–120, doi:10.1016/0309-1708(89)90017-1.

Gonella, J., 1972: A rotary-component method for analysing meteorological and oceanographic vector time series. *Deep Sea Research and Oceanographic Abstracts*, **19 (12)**, 833 – 846, doi:10.1016/0011-7471(72)90002-2.

Haigh, I. D., M. Eliot, and C. Pattiaratchi, 2011: Global influences of the 18.61 year nodal cycle and 8.85 year cycle of lunar perigee on high tidal levels. *Journal of Geophysical Research: Oceans*, **116 (C6)**, doi:10.1029/2010JC006645.

Hewitt, E., and R. E. Hewitt, 1979: The Gibbs-Wilbraham phenomenon: An episode in Fourier analysis. *Archive for History of Exact Sciences*, **21 (2)**, 129–160, doi:10.1007/BF00330404.

Hoerl, A. E., and R. W. Kennard, 1970: Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12 (1)**, 55–67, doi:10.1080/00401706.1970.10488634.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation : Operational, sequential and variational. *Journal of the Meteorological Society of Japan. Ser. II*, **75 (1B)**, 181–189, doi:10.2151/jmsj1965.75.1B_181.

Le Provost, C., 2001: Chapter 6 ocean tides. *Satellite Altimetry and Earth Sciences*, L.-L. Fu, and A. Cazenave, Eds., International Geophysics, Vol. 69, Academic Press, 267 – 303, doi: 10.1016/S0074-6142(01)80151-0.

Lee, P. M., 1997: *Bayesian statistics : an introduction*. 2nd ed., Arnold, London, ISBN: 0-471-19481-6.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI: 10.1175/JTECH-D-21-00034.1.

Leffler, K. E., and D. A. Jay, 2009: Enhancing tidal harmonic analysis: Robust (hybrid L1/L2) solutions. *Continental Shelf Research*, **29 (1)**, 78 – 88, doi:10.1016/j.csr.2008.04.011, physics of Estuaries and Coastal Seas: Papers from the PECS 2006 Conference.

Lourens, A., and F. C. van Geer, 2016: Uncertainty propagation of arbitrary probability density functions applied to upscaling of transmissivities. *Stochastic Environmental Research and Risk Assessment*, **30 (1)**, 237–249, doi:10.1007/s00477-015-1075-8.

Matte, P., D. A. Jay, and E. D. Zaron, 2013: Adaptation of classical tidal harmonic analysis to nonstationary tides, with application to river tides. *Journal of Atmospheric and Oceanic Technology*, **30 (3)**, 569–589, doi:10.1175/JTECH-D-12-00016.1.

Menke, W., 2018: *Geophysical data analysis : discrete inverse theory*. Fourth edition ed., Elsevier Ltd., London, United Kingdom.

Munk, W., and K. Hasselman, 1964: Super-resolution of tides. *Studies on Oceanograhy*, 339–334, (Hidaka volume).

Munk, W. H., B. Zetler, and G. W. Groves, 1965: Tidal cusps. *Geophysical Journal of the Royal Astronomical Society*, **10 (2)**, 211–219.

National Oceanic and Atmospheric Administration, 2005: Deep-Ocean Assessment and Reporting of Tsunamis (DART(R)). NOAA National Centers for Environmental Information, accessed: 2021-03-21, doi:10.7289/V5F18WNS.

Ohlmann, C., P. White, L. Washburn, B. Emery, E. Terrill, and M. Otero, 2007: Interpretation of coastal HF radar–derived surface currents with high-resolution drifter data. *Journal of Atmospheric and Oceanic Technology*, **24 (4)**, 666–680, doi:10.1175/JTECH1998.1.

Pawlowicz, R., R. Beardsley, and S. Lentz, 2002: Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Computers & Geosciences*, **28 (8)**, 929–937, doi: 10.1016/S0098-3004(02)00013-4.

Rainville, L., and R. Pinkel, 2006: Propagation of low-mode internal waves through the ocean. *Journal of Physical Oceanography*, **36 (6)**, 1220 – 1236, doi:10.1175/JPO2889.1.

Ray, R. D., and G. T. Mitchum, 1996a: Surface manifestation of internal tides generated near Hawaii. *Geophysical Research Letters*, **23 (16)**, 2101–2104, doi:10.1029/96GL02050.

Ray, R. D., and G. T. Mitchum, 1996b: Surface manifestation of internal tides generated near Hawaii. *Geophysical Research Letters*, **23 (16)**, 2101–2104, doi:10.1029/96GL02050.

Ray, R. D., and E. D. Zaron, 2011: Non-stationary internal tides observed with satellite altimetry. *Geophysical Research Letters*, **38 (17)**, doi:10.1029/2011GL048617.

Ray, R. D., and E. D. Zaron, 2016: M2 internal tides and their observed wavenumber spectra from satellite altimetry. *Journal of Physical Oceanography*, **46 (1)**, 3–22, doi: 10.1175/JPO-D-15-0065.1.

Roarty, H., and Coauthors, 2019: The global high frequency radar network. *Frontiers in Marine Science*, **6**, 164, doi:10.3389/fmars.2019.00164.

Terrill, E., and Coauthors, 2006: Data management and real-time distribution in the HF-radar national network. *OCEANS 2006*, 1–6, doi:10.1109/OCEANS.2006.306883.

Van Trees, H. L., 2001: *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*, chap. 2, 54–63. John Wiley & Sons, Inc., ISBNs: 0-471-09517-6 (Paperback); 0-471-22108-2 (Electronic).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-00034.1.

von Storch, H., and F. W. Zwiers, 2003: *Statistical Analysis in Climate Research*. Cambridge University Press, ISBN: 0-521-45071-3.

Wunsch, C., 1996: *The Ocean Circulation Inverse Problem*. Cambridge University Press, ISBN: 0-521-48090-6.

Zaron, E. D., 2019: Predictability of non-phase-locked baroclinic tides in the Caribbean Sea. *Ocean Science*, **15 (5)**, 1287–1305, doi:10.5194/os-15-1287-2019.

Zetler, B. D., M. D. Schuldt, R. W. Whipple, and S. D. Hicks, 1965: Harmonic analysis of tides from data randomly spaced in time. *Journal of Geophysical Research*, **70 (12)**, 2805–2811.

Zhao, Z., M. H. Alford, J. Girton, T. M. S. Johnston, and G. Carter, 2011: Internal tides around the Hawaiian Ridge estimated from multisatellite altimetry. *Journal of Geophysical Research: Oceans*, **116 (C12)**, doi:10.1029/2011JC007045.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

LIST OF TABLES

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

TABLE 1. Root-mean-square error for the four step function analyses shown in Figure 2(B-E). RMSE is calculated at observation times (top row) and at a set of times sampled at 100 times the observational resolution (bottom row).

	White **P** $f_{max} = f_{Ny}$	True **P** $f_{max} = f_{Ny}$	White **P** $f_{max} = 2f_{Ny}$	True **P** $f_{max} = 2f_{Ny}$
RMSE at observation times	0.0080	0.0084	0.0122	0.0038
RMSE of interpolated time series	0.0211	0.0201	0.5014	0.0171

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

LIST OF FIGURES

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-00034.1.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI: 10.1175/JTECH-D-21-0034.1.

Fig. 13. Harmonic amplitudes (normalized to units of the power spectrum) of a high-frequency radar time series (zonal component u), with the time series power spectrum S_{uu} and domain-averaged power spectrum (used in **P**) shown for comparison. Model coefficient uncertainty (90% credible interval) is represented by shaded intervals. The covariance used to build **R** is truncated at 300 hours time lag and is built from a spectrum of spectral slope -1 (dotted line). Amplitudes from t_tide, both significant and non-significant, are included for comparison. (B) Close-up of the M_2 band centered at the principal lunar semidiurnal frequency. The vertical dashed lines on either side of the central M_2 line are spaced by the annual frequency . 63

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

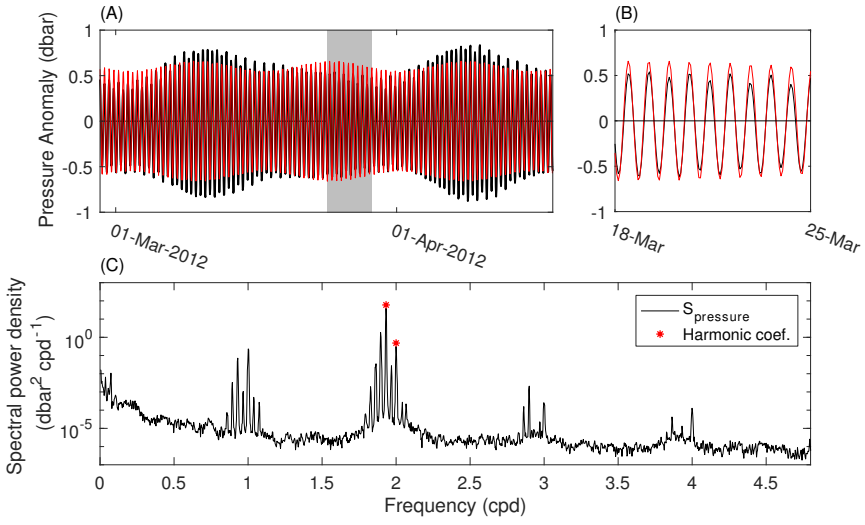

FIG. 1. A simplified example of tidal harmonic analysis on bottom pressure data (3.5 year record at 8.48°S, 125.03°W, DART array site 51406, further analyzed in section 4a). (A) A segment of the longer time series that was analyzed (black) and the resulting fit (red) to only two tidal constituent frequencies: M_2 and S_2. (B) A closeup of the time series over the gray interval in (A). (C) The power spectrum of the time series and the magnitude of the harmonic coefficients.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

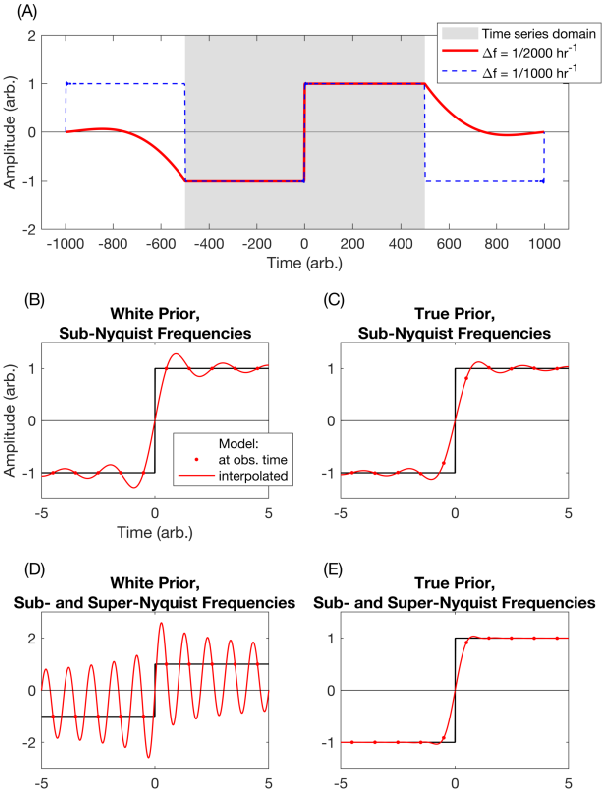

FIG. 2. Step function discretely-sampled at 1000 points. (A) Full record modeled with orthogonal sinusoids (blue dashed) and with non-orthogonal sinusoids corresponding to a record twice as long (red solid). The gray shaded region indicates the range of data that are fit, outside of which the time series is extrapolated. (B) Close view of discontinuity and red_tide output given spectrally white (incorrect) model parameter covariance, **P**, only modeling up to the Nyquist frequency f_{Ny}. (C) Discontinuity and output using true **P**, only modeling up to f_{Ny}. (D) Same as panel B but including model frequencies above f_{Ny} up to $2f_{\mathrm{Ny}}$. (E) Same as panel C but including model frequencies above f_{Ny} up to $2f_{\mathrm{Ny}}$. The residual is treated as white (uncorrelated) noise in all panels, with minimal contribution to the total variance (0.2% for the blue dashed curve in panel A, 0.1% for the red curve in panel A, 0.1% for panels B-C, and 0.05% for panels D-E).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

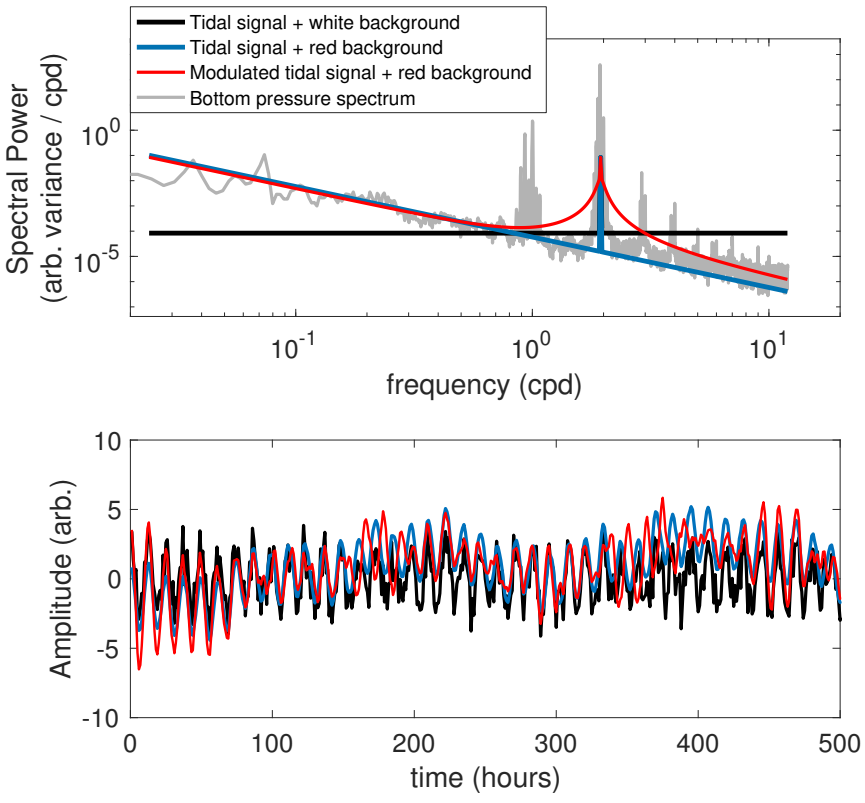

FIG. 3. (top) Example of idealized power spectra with a single constituent, here M_2, with the computed power spectrum of the bottom pressure record from Figure 1 included for comparison, and (bottom) segments of corresponding time series computed from the power spectra with random phase at each frequency (500 hourly synthetic data points). All three spectra share the same tidal peak magnitude, and all three time series share the same phase across frequencies.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

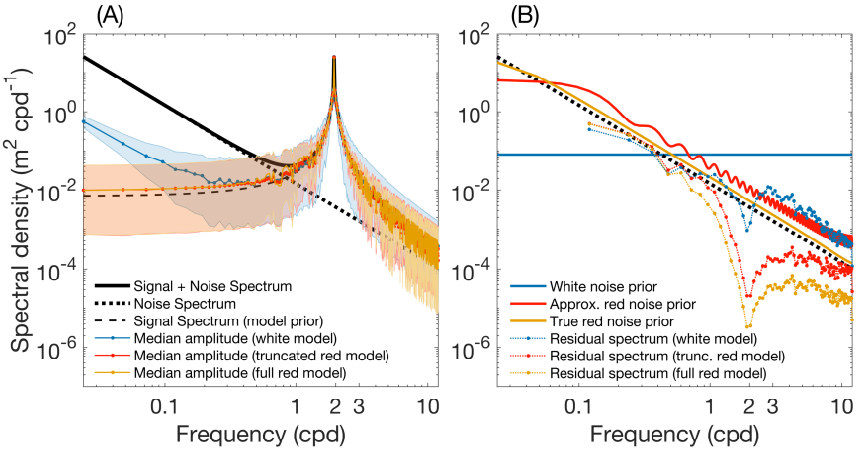

FIG. 4. (A) Synthetic true spectra (noise and signal) and harmonic amplitudes calculated using red_tide, and (B) assumed and calculated noise spectra. Shading indicates 90% credible intervals of amplitude, with thick lines indicating median amplitude based on the estimated distribution of coefficients. All quantities are scaled to spectral density units for comparison. The true spectra are used to construct the noise and signal components of the time series (1001 hourly data points) from which harmonic amplitudes are calculated. The noise spectra in (B) indicate the Fourier transforms of the residual covariance matrix **R** used in each application of the red_tide procedure. All noise spectra used to construct **R** have the same variance, which approximates the 42% contribution of true noise variance to the true total variance. The residual spectra are calculated using Welch's method with a Hanning window and 9 overlapping segments. The tidal peak is at the M_2 frequency and is convolved with a spectrum of slope -2 in order to simulate real tidal modulation that leads to tidal cusps.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

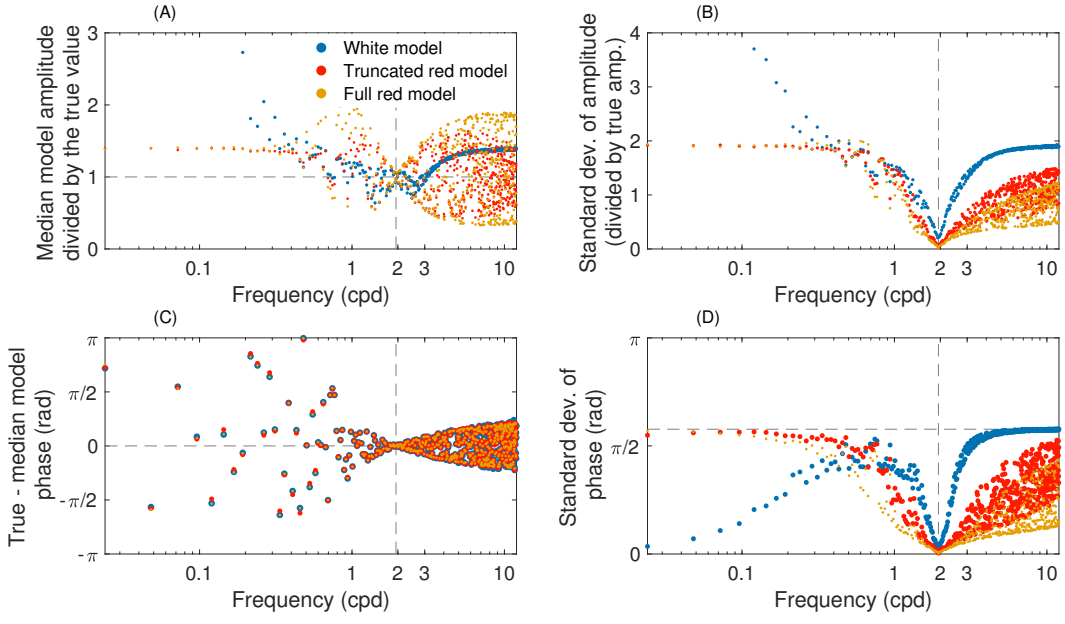

FIG. 5. For the same time series and analyses in Figure 4, (A) model amplitude fraction of truth, (B) standard deviation of the normalized amplitude, (C) phase difference between the true phase and the model estimates, and (D) standard deviation of the estimated phase after being centered on the interval $(-\pi, \pi)$.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

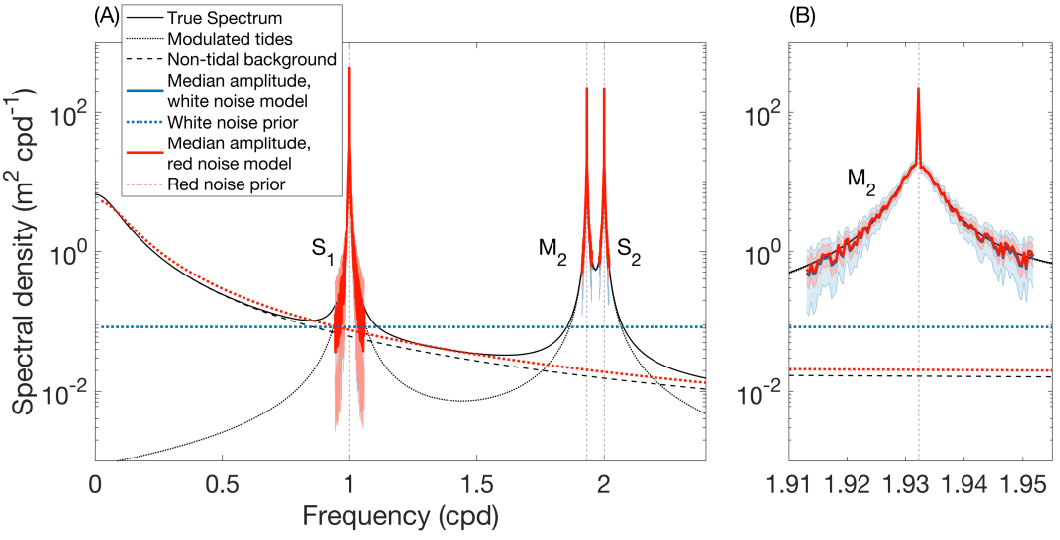

FIG. 6. Synthetic spectrum, harmonic amplitudes calculated from corresponding time series (7 years sampled hourly) using red_tide (scaled to spectral density units), and spectra used to construct **R**, shown (A) in the range of frequencies typically relevant to tidal harmonic analysis, and (B) at a close view about the M_2 constituent frequency. Shading indicates 90% credible intervals of amplitude. Dashed colored lines indicate the Fourier transforms of the residual covariance matrix **R** used in each estimation. Both of these noise spectra assume the same total signal to noise variance ratio of 1.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

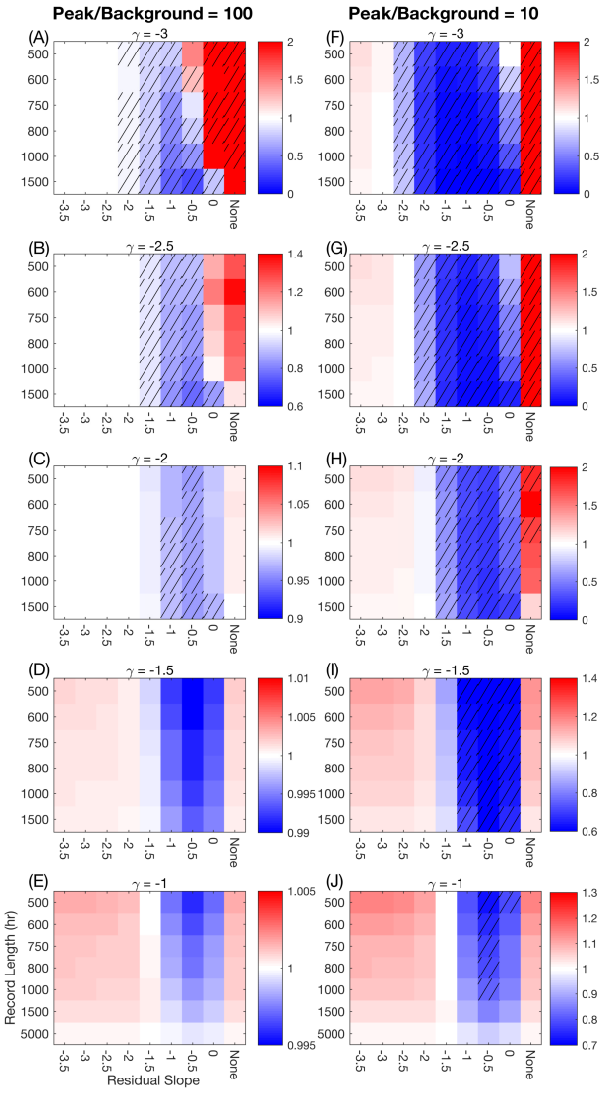

FIG. 7. Mean ratios of M_2 harmonic amplitudes to true amplitudes using red_tide and ordinary least squares. Panels (A-E) correspond to time series with tidal spectral power density 100 times greater than that of the random background process at the M_2 frequency. Panels (F-J) correspond to time series with tidal spectral power density 10 times greater than that of the random background process at M_2. The random background process to which tides are added has a spectral slope indicated by γ. Time series have 24,000 hourly samples but are only analyzed up to 1500 hours, except for $\gamma = -1$, which was also analyzed at 5000 hours. Each plot corresponds to N Monte Carlo simulations with constant spectra and random phase, where N = 10,000 for $\gamma \leq -2$ and N = 50,000 for $\gamma > -2$. The spectral slope of the noise term used in red_tide (horizontal axis) and the length of the record (vertical axis) are varied, with noise slope "none" corresponding to the OLS solution. Color bars are centered at unity and scale differently across panels for clarity. Hatched cells indicate combinations of record length and assumed noise spectral slope that resulted in the value 1 (estimated amplitude equal to true amplitude) falling outside the interquartile range for the 10,000 or 50,000 member ensemble.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

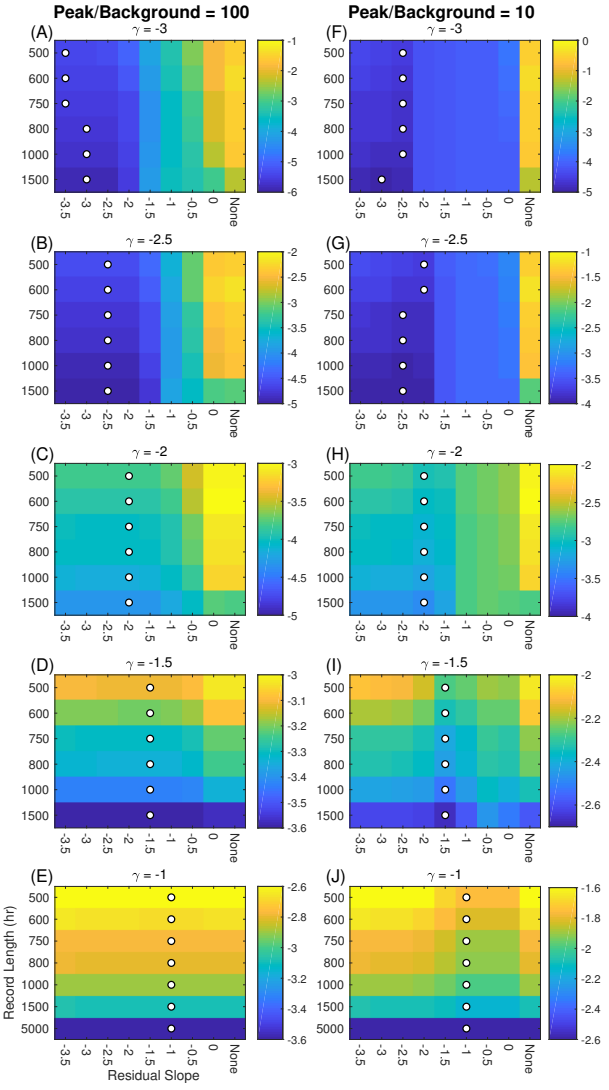

FIG. 8. Log_{10} of normalized model parameter variance for the M_2 constituent calculated from the same artificial data described in Figure 7 with the same figure layout. Filled dots indicate the minimum value of its row, and color bars scale differently across panels for clarity.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

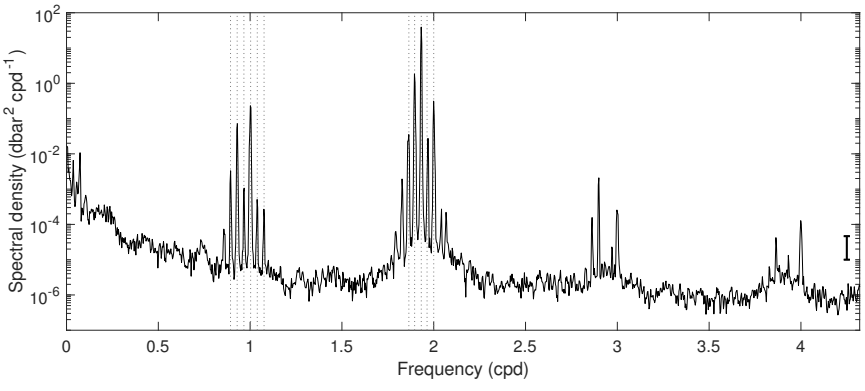

FIG. 9. Power spectrum of bottom pressure from site 51406 (8.48°S, 125.03°W), part of the DART array. Higher harmonics of the tidal frequencies are present at resolved frequencies higher than roughly 4 cpd (not shown). Vertical dashed gray lines denote prominent tidal constituents, which are the dominant sources of variance in bottom pressure time series. This time series of length 3 years and 6 months (Feb. 12, 2011 to Sept. 6, 2014), sampled every 15 seconds, is averaged over 1 hour intervals, with the power spectrum estimated using Welch's method. The 95% confidence intervals are indicated by the error bar in the lower right, plotted at frequency 4.25 cpd. This corresponds to a ratio valid at all magnitudes of spectral density.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

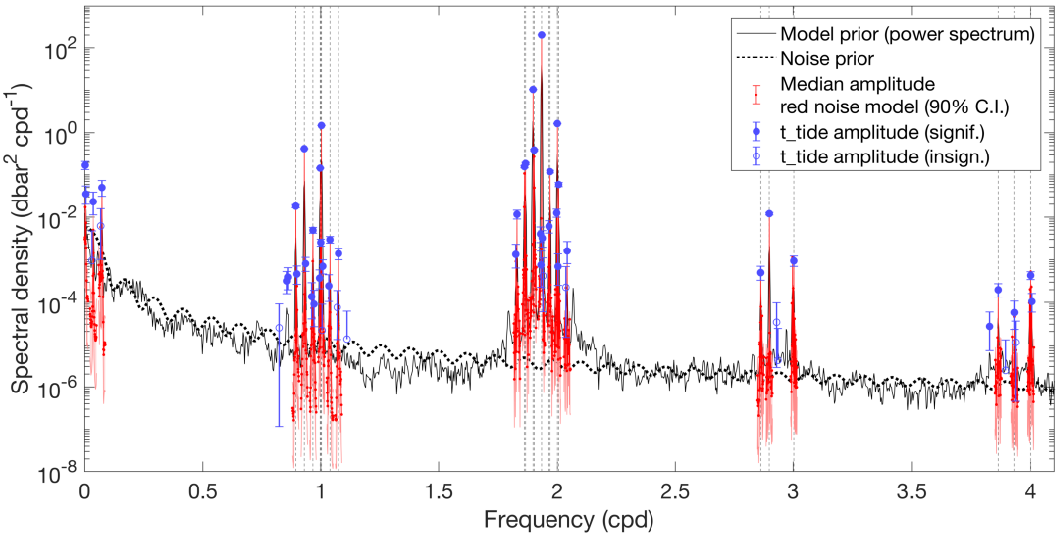

FIG. 10. Diagonal of model parameter covariance matrix (linear interpolated power spectrum, black solid line), harmonic amplitudes calculated in red_tide using a red noise covariance (spectral slope = -3/2 to empirically match observed spectrum), and results using t_tide. Shading indicates 90% credible intervals. Select diurnal and semidiurnal tidal frequencies are indicated by dashed gray vertical lines. The t_tide amplitudes are from constituents automatically chosen by the package, with amplitudes considered by t_tide to be "significant" filled in, otherwise unfilled, and with 90% confidence bars. All values are in units of spectral density in order to compare quantities defined on different frequency intervals.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

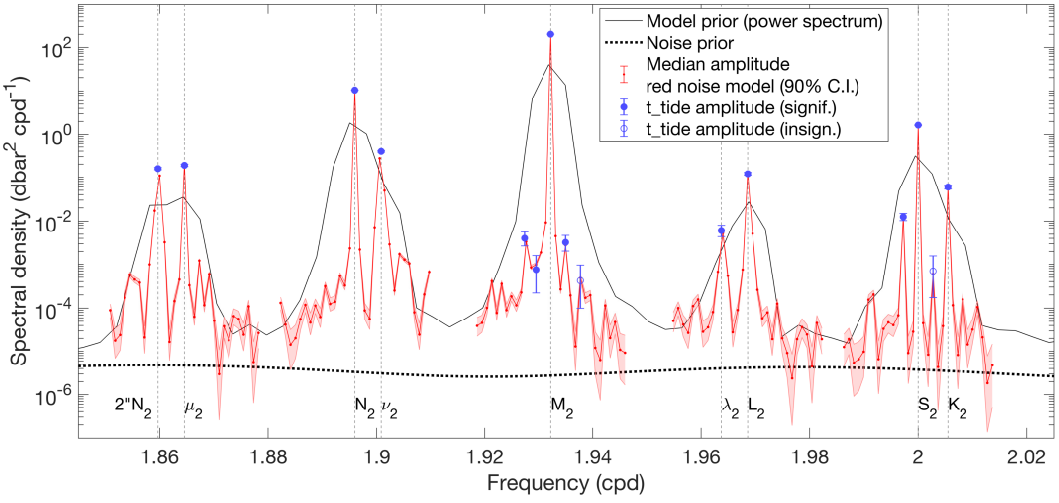

FIG. 11. Results from Figure 10 focused on the semidiurnal band. Prominent tidal constituent frequencies, including those not resolved in the power spectrum, are indicated by vertical dotted lines and labeled according to their conventional Darwin symbols.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

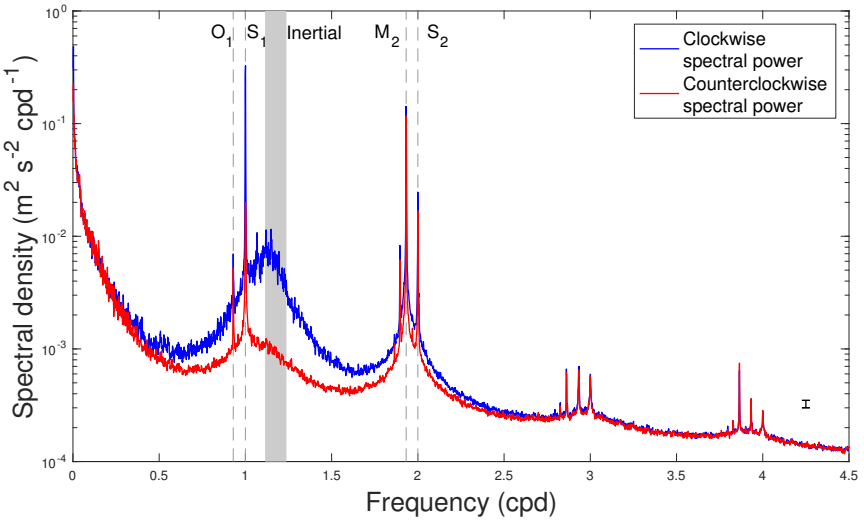

FIG. 12. Rotary power spectrum of surface current **u** from HFR stations along the central California coast, regionally averaged over 1191 grid points. Time series of length 9 years and 3 months with at least 50% data coverage are sampled hourly. Dashed gray lines denote, in order of increasing frequency, the O_1, S_1, M_2, and S_2 constituents, while the gray band denotes the range of inertial frequencies across the latitudes spanned by the data. Higher harmonics of the tidal frequencies are present at resolved frequencies higher than those shown. The 95% confidence ratio is indicated by the error bar in the lower right, plotted at frequency 4.25 cpd. This corresponds to a ratio valid at all magnitudes of spectral density.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.

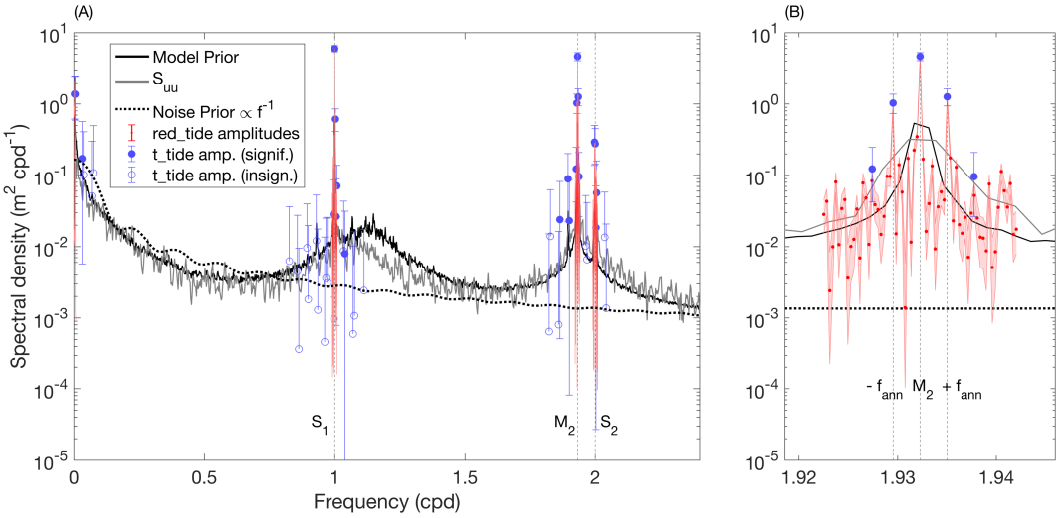

FIG. 13. Harmonic amplitudes (normalized to units of the power spectrum) of a high-frequency radar time series (zonal component u), with the time series power spectrum S_{uu} and domain-averaged power spectrum (used in **P**) shown for comparison. Model coefficient uncertainty (90% credible interval) is represented by shaded intervals. The covariance used to build **R** is truncated at 300 hours time lag and is built from a spectrum of spectral slope -1 (dotted line). Amplitudes from t_tide, both significant and non-significant, are included for comparison. (B) Close-up of the M_2 band centered at the principal lunar semidiurnal frequency. The vertical dashed lines on either side of the central M_2 line are spaced by the annual frequency

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0034.1.