Validating the spatial variability of the semidiurnal internal tide in a realistic global ocean simulation with Argo and mooring data

Gaspard Geoffroy[1], Jonas Nycander[1], Maarten C. Buijsman[2], Jay F. Shriver[3], and Brian K. Arbic[4]

[1]Dept. of Meteorology, Stockholm University, Stockholm, Sweden
[2]School of Ocean Science and Engineering, University of Southern Mississippi, Stennis Space Center, MS, USA
[3]Naval Research Laboratory, Stennis Space Center, MS, USA
[4]Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA

Correspondence: Gaspard Geoffroy (gaspard.geoffroy@misu.su.se)

Abstract. The autocovariance of the semidiurnal internal tide (IT) is examined in a 32-day segment of a global run of the HYbrid Coordinate Ocean Model (HYCOM). This numerical simulation, with 41 vertical layers and 1/25° horizontal resolution, includes tidal and atmospheric forcing allowing for the generation and propagation of IT to take place within a realistic eddying general circulation. The HYCOM data are in turn compared with global observations of the IT around 1,000 dbar, from Argo
5   float park phase data and mooring records. HYCOM is found to be globally biased low in terms of the IT variance and decay ~~rate of the autocovariance of the semidiurnal IT~~ of the IT autocovariance over timescales shorter than 32 days. Except in the Southern Ocean, where limitations in the model causes the discrepancy with in situ measurements to grow poleward, the spatial correlation between the Argo and HYCOM tidal variance suggests that the generation of low-mode semidiurnal IT is globally well captured by the model.

10   ## 1   Introduction

Internal tides (IT) are internal waves generated by the interaction of tidal currents with rough bathymetry. The radiated wave beams can travel thousands of kilometers (e.g., Zhao et al., 2016; Buijsman et al., 2020) over which they undergo dissipative processes as they interact with the eddying ocean and bottom topography, to eventually break (MacKinnon et al., 2017). The dissipation of IT represents a major source of vertical mixing in the ocean interior (de Lavergne et al., 2020). As such, IT have
15   a key influence on the ocean state (Melet et al., 2016), and therefore on the global climate system (see Melet et al., 2022, for a review on the subject).

At any given position in a stationary medium, tidally forced waves would have a constant phase difference to the astronomical forcing. However, since they propagate within the time varying ocean circulation, IT are subject to a variety of mechanisms that cause their phase difference to the tidal forcing at their generation site to shift with time (Rainville and Pinkel, 2006; Shriver
20   et al., 2014; Zaron and Egbert, 2014; Buijsman et al., 2017). In other words, IT lose *coherence* by interacting with the eddying ocean. They decorrelate: the autocovariance of a time series representing the internal tide variability at a fixed position (away from the source) inevitably decays with time lag (Caspar-Cohen et al., 2022; Geoffroy and Nycander, 2022).

Only the *coherent* fraction of the IT energy decays with time lag, that is the energy carried by waves with a constant phase difference to the astronomical forcing. Conversely, the *incoherent* fraction of the energy grows, so that the *total* IT energy, or variance, at a given location (the sum of the coherent and incoherent components) is unaffected by the decorrelation. For very long time lags, the coherent and the complementary incoherent asymptotic limits are often called *stationary* and *nonstationary* variance, respectively. While the global field of the stationary (low-mode) IT is widely considered to be well constrained by multiyear satellite altimetry observations (Dushaw et al., 2011; Ray and Zaron, 2016; Zaron, 2019; Zhao, 2019), the nonstationary component is not.

Some current high-resolution global ocean circulation models enable the estimation of the barotropic and internal tides concurrently with the ocean circulation (Arbic et al., 2010; Shriver et al., 2012; Buijsman et al., 2020). Such fully nonlinear numerical simulations also incorporate the interactions of the generated IT with the eddying ocean and its boundaries. The model utilized in this study is the HYbrid Coordinate Ocean Model (HYCOM; Chassignet et al., 2006) with 41 vertical layers and 1/25° horizontal resolution. The literature on the validation of HYCOM with observations is already vast (see Buijsman et al., 2020; Arbic, 2022, for recent accounts). In the latest developments, surface drifters have been used to validate the geographical variability of the kinetic energy in various frequency bands at the global scale (Arbic et al., 2022). While the global coverage of drifter data is comparable to that of satellite altimetry, the contribution from the baroclinic tides to the kinetic energy observed by drifters has not been determined yet. Hence, until recently, only altimetry could unveil the geographical variability of the IT at the global scale.

The empirical mapping of IT from altimetry remains challenging for various reasons (Egbert and Ray, 2017). Most notably, the long sampling intervals of altimeters and the low signal-to-noise ratio preclude any direct estimation in the time domain of the total IT variance at a single location. Notwithstanding, Zaron (2017) ~~analysed~~ analyzed along-track wavenumber spectra of the sea surface height to map the total and nonstationary semidiurnal IT variance (for the baroclinic mode 1 only). The author found a global mean ratio of nonstationary to total semidiurnal IT variance of 44%. He also outlined the spatial inhomogeneity of the tidal variability, with this ratio being larger than 50% in much of the equatorial ~~Pacific~~ Pacific and Indian Oceans (Zaron, 2017). In a comparison with data from HYCOM, Nelson et al. (2019) showed the 'k-space' methodology of Zaron (2017) to miss a large fraction of both the nonstationary and total variance. Nevertheless, the spatial correlation of the nonstationary fraction (i.e., the ratio of the nonstationary to total variance) between the model and altimetry suggests that the model at least qualitatively captures the generation of IT and their interactions with the background circulation (Nelson et al., 2019).

Recently, Geoffroy and Nycander (2022) used observations from Argo park phase data (Argo, 2000) to empirically map the variance of the semidiurnal IT at 1,000 dbar. The high sampling rate of the floats captures the total variance of the IT, i.e. the autocovariance at short time lags, before any significant loss of coherence occurs and after most of the ~~oceanic~~ noise has decorrelated. On the other hand, the Lagrangian sampling of the drifting floats results in decorrelating effects on top of the decorrelation of the IT itself (Zaron and Elipot, 2021; Caspar-Cohen et al., 2022; Geoffroy and Nycander, 2022). Following Caspar-Cohen et al. (2022), we call *apparent* decorrelation the effects of the Lagrangian sampling on the autocovariance. This apparent decorrelation cannot be disentangled from the decorrelation of the IT (i.e., the decorrelation due to interactions with the background circulation) using Argo data only. Thus, to gain insights into the IT decorrelation around 1,000 dbar, one

can instead apply the methodology of Geoffroy and Nycander (2022) to Eulerian observations from moorings. In the present work we will compare observations of the (total) variance and decorrelation of the semidiurnal IT around 1,000 dbar from Argo floats and moorings, respectively, to data from a global HYCOM run. Contrarily to other recent validations of HYCOM with mooring data (e.g., Ansong et al., 2017; Luecke et al., 2020), the Eulerian component of our analysis is not meant as a standalone point-to-point comparison. Rather, it is designed to bolster and extend the main analysis of the Lagrangian data.

The remainder of this work is organized as follows: In Sect. 2, the in situ datasets and the numerical simulation are presented. In Sect. 3, we review the methodology of Geoffroy and Nycander (2022) in an example location from the Lagrangian and Eulerian perspectives. In Sect. 4, the main results are presented. We compare the geographical variability of the variance and decorrelation of the semidiurnal IT obtained from the in situ data and numerical simulation. Then, the model data are used to quantify the decorrelation induced by the Lagrangian sampling. Finally, we outline the potential biases affecting the datasets. We conclude in Sect. 5, by discussing the results and giving a summary.

2   Data

The different comparisons in this work are all done in terms of vertical displacement of the isotherms. The measurable variables needed to compute vertical isotherm displacement at a fixed depth are the temperature anomaly and vertical temperature gradient at that depth. In this section we briefly describe the temperature time series from each dataset.

2.1   Argo Floats With Iridium Communication

We use data from a global collection of Argo Iridium floats deployed by the University of Washington as part of the National Ocean Partnership Program during the period 2004–2022. Between the descending and ascending profiling phases, these floats also record temperature and pressure with an hourly resolution while adrift at 1,000 dbar. This so-called park phase typically lasts 10 days. As in Hennon et al. (2014) and Geoffroy and Nycander (2022), we use the pressure records to correct for the small departures of a float from its drifting control level during a park phase.

Stitching together data from successive cycles, by filling the time between park phases (typically 6 hr) with NaNs, one can construct longer time series. In this study, we use segments of 32 days of data (i.e., the duration of the segment of numerical simulation we are using). The sampling period of the park phase can occasionally vary by more than a few seconds. To ensure evenly spaced time series, we linearly interpolate each concatenated record of 32 days onto a time axis with constant 1 h step. Any interpolated value lying between two original records that are more than 1.5 h apart is replaced by NaN.

The position of the floats can only be determined when they reach the surface. We assume straight trajectories between two successive surfacings (typically 10 days apart). This assumption has been shown reasonable by Geoffroy and Nycander (2022), especially since we discard segments of data for which the mean speed of the float is larger than 0.1 m s^{-1}. This criterion is primarily intended to avoid contamination by lee waves: a flow with speed $U \approx 0.1$ m s^{-1} passing over a bathymetric feature with horizontal length scale $\lambda \approx 10$ km will generate lee waves with angular frequency $\omega = (2\pi/\lambda)U$ rad s^{-1} of the order of the diurnal frequency.

90   The Argo dataset used in this work is an updated version of the one used by Geoffroy and Nycander (2022). After the base processing and quality controls, we are left with 22,414 valid 32-day segments from 891 individual floats. A more detailed description of the Argo data processing and quality controls we employed is available from Hennon et al. (2014) and Geoffroy and Nycander (2022).

2.2 Global Multi-Archive Current Meter Database (GMACMD)

95 The Global Multi-Archive Current Meter Database (GMACMD; Scott et al., 2011) compiles tens of thousands of oceanographic time series from moorings. Previous model-data validation efforts involving both HYCOM and mooring data notably include Luecke et al. (2020), where the authors compared the temperature variance and kinetic energy, over various frequency bands, in realistic global ocean simulations to more than 3,800 instrumental records.

Here, we extracted 331 temperature time series spanning 1972–2010 and meeting the following criteria:

100 1. The mooring lies in water deeper than 2,000 m

2. The record is longer than 64 days with a sampling interval shorter than 3 hr, for adequate resolution of the semidiurnal tidal signals

3. The instrument depth is within ± 200 m from 1,000 m

One particular mooring is used as an example in Sect. 3. It recorded during 366 days in the years 1982-83 at the position
105 $28.00°$ N, $151.95°$ W (north of the Hawaiian Ridge). As previously documented by Alford and Zhao (2007), we refer to the instrument at 1,119 m depth as mooring No. 2. This instrument sampled temperature with a 0.25 h resolution.

2.3 HYbrid Coordinate Ocean Model (HYCOM)

This study uses 32 days, from May 20 to June 20, 2019, of hourly output at 1,000 m depth from a global run of HYCOM, with 41 vertical layers and $1/25°$ horizontal resolution. The non data-assimilative simulation, designated 'GLBy190.04', includes
110 realistic tidal and atmospheric forcing enabling the generation and propagation of IT within the eddying general circulation. The high resolution 2D fields are complemented by monthly-mean 3D fields of temperature and salinity subsampled to $1°$.

A Lagrangian analysis of the simulation is used for a direct model-data validation with Argo floats. The Argo quasi-Lagrangian sampling is mimicked by releasing 41644 particles randomly across the world oceans. We let the particles be advected by the 2D velocity field at 1,000 m for 32 days while sampling temperature with an hourly resolution. This La-
115 grangian sampling of HYCOM is achieved using the software Parcels (Van Sebille et al., 2021). The Lagrangian simulation uses a classic Runge-Kutta method for computing the advection of the particles (with 5 minutes integration time step). As for the Argo data, we discard particles with a mean speed larger than 0.1 m s^{-1}. We also discard any particle crossing the 1,000 m isobath during the simulation.

3 Methods

120 In this section, we present a local example to introduce the methods that will be used in Sect. 4 for the global comparison. We start by comparing the Argo observations to the Lagrangian sampling of the HYCOM data. Next, we investigate the effects of the drift by comparing Lagrangian and Eulerian samplings of the numerical simulation. We end the section by introducing the comparison between the Eulerian HYCOM time series and mooring observations.

3.1 Lagrangian sampling of the isopycnal displacement at 1,000 dbar

125 As in Hennon et al. (2014) and Geoffroy and Nycander (2022), we define the vertical isotherm displacement observed by a Lagrangian particle η_{1000}^{L} as

$$\eta_{1000}^{L}(t) = \frac{T(t) - \overline{T}}{(\mathrm{d}T/\mathrm{d}P)_{1000}(t)}, \tag{1}$$

where $\overline{T}$ is the time average of the temperature measurements $T(t)$ over a particle trajectory, and $(\mathrm{d}T/\mathrm{d}P)_{1000}(t)$ is the temperature gradient at 1,000 dbar. Hence, the vertical isotherm displacement is simply computed as the temperature anomaly
130 recorded by a drifting particle at 1,000 dbar divided by the vertical temperature gradient at that depth. For HYCOM particles, we compute $(\mathrm{d}T/\mathrm{d}P)_{1000}$ as the mean, ~~within 100 m~~between 900 and 1100 m, of ~~1, 000 m, of~~ the vertical gradient calculated from the modeled monthly-mean 3D temperature field. We then linearly interpolate the temperature gradient in the horizontal to the successive particle positions (hence the time dependence). To avoid any spurious displacements, we set η_{1000}^{L} to NaN whenever the magnitude of the temperature gradient is smaller than 3×10^{-5} °C dbar^{-1}.

135 In the case of Argo floats, Eq. (1) is evaluated over each park phase. $\overline{T}$ then represents the time average of the temperature measurements $T(t)$ over a park phase, and the vertical temperature gradient is estimated from the average of the two neighboring ascending profiles. Specifically, $(\mathrm{d}T/\mathrm{d}P)_{1000}(t)$ is computed as the mean, within 100 dbar of the parking pressure, of the vertical gradient calculated from the average temperature profile. ~~To avoid any spurious displacements~~As for the HYCOM particles, we discarded the data from the whole park phase whenever the magnitude of the temperature gradient is smaller than
140 ~~3×10^{-5}~~3×10^{-5} °C dbar^{-1}. As explained in Sect. 2.1, η_{1000}^{L} time series from successive park phases are stitched together to constitute 32-day time series.

Both for Argo floats and HYCOM Lagrangian particles, the low frequency background activity is filtered out from η_{1000}^{L} using a ~~fourth order~~fourth-order Butterworth filter with a cut-off frequency of 0.3 cpd. The inertial peak is not removed by this filter poleward of about $\pm 10°$ of latitude.

145 ### 3.2 Averaging sample autocovariance series

The HYCOM derived Lagrangian time series η_{1000}^{L} can be analyzed in the same way as in Geoffroy and Nycander (2022). We illustrate the mapping process below. The geographical patch presented here is also described in Geoffroy and Nycander (2022). For this example area, Fig. 1 shows 8 32-day segments of Argo data and 13 HYCOM particles selected using their median position. The circular patch of radius 200 km containing the latter median positions is centered on the mooring No. 2.
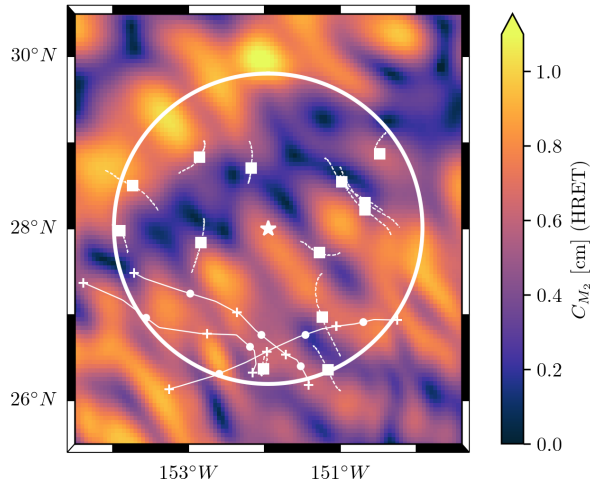
Figure 1. Example circular geographical patch of radius 200 km (the white star denotes mooring No. 2 at the center). The median position of the segments of Argo data and of the HYCOM particles are shown by white filled circles and white squares, respectively. The dashed white curves represent the trajectories of the HYCOM particles over the 32 days of numerical simulation. The binned Argo segments were all recorded by the same float, its trajectory is shown by the solid white curve with white crosses denoting the starting and ending points of the different 32-day segments. In the background we show the amplitude of the M_2 baroclinic sea level anomaly from the High-Resolution Empirical Tide model (HRET; Zaron, 2019).

150 From a finite time series $\eta(t)$ one can calculate the sample autocovariance $\hat{R}(\tau)$:

$$\hat{R}(\tau) = \frac{1}{N-\tau} \sum_{t=1}^{N-\tau} (\eta_t - \hat{\bar{\eta}})(\eta_{t+\tau} - \hat{\bar{\eta}}), \tag{2}$$

where N is the total number of observations, and $\hat{\bar{\eta}}$ is the sample mean of the series. Note that $\hat{R}(0)$ is the sample variance of the series. Here, N is taken as the number of non-missing observations, accounting for gaps in the Argo time series (during the descent, main profiling, and surface transmission phases of the float cycle, or because of failed quality controls). The sample
155 autocovariance as defined in Eq. (2) is an unbiased estimator of the "true" autocovariance (i.e., $\hat{R}(\tau)$ converges to the true value $R(\tau)$ for infinitely large N). In the remainder of this article we only compute the variance and autocovariance of vertical isotherm displacement time series. For the sake of brevity, we simply refer to them as the variance and autocovariance.

The sample autocovariances for all HYCOM particles within the circular patch shown in Fig. 1 are averaged to obtain a local mean autocovariance $\overline{R}_{\mathrm{HYCOM}}^{L}$. The local mean autocovariance from the Argo data, $\overline{R}_{\mathrm{Argo}}$, is computed in the same way. The
160 standard error of the local mean autocovariance is computed as

$$\mathrm{SEM}(\tau) = \frac{\mathrm{STD}(\tau)}{\sqrt{N_p}}, \tag{3}$$

where STD is the standard deviation of the sample autocovariances over the subset, and N_p is the number of particles in the subset. Fig. 2a demonstrates the good agreement of the two datasets until $\tau \simeq 200$ $\tau \approx 200$ h. Past this time lag limit, $\overline{R}_{\mathrm{Argo}}$

falls under its 95% confidence interval (95% C.I. ~~$\sim \pm 2\,\mathrm{SEM}$~~ $= \pm 2\,\mathrm{SEM}$) and thus cannot be considered significantly different from zero.

A handy tool for monitoring the evolution of the autocovariance of the semidiurnal IT is complex demodulation. Here, it consists in the least squares fitting of $A\cos(\omega_{\mathrm{SD}}\tau) + B\sin(\omega_{\mathrm{SD}}\tau)$, where ~~$\omega_{\mathrm{SD}} = (\omega_{M_2} + \omega_{S_2})/2$~~ $\omega_{\mathrm{SD}} = (\omega_{M_2} + \omega_{S_2})/2$ is the semidiurnal frequency, to the sample autocovariance in 48-h windows with 75% overlap. This is equivalent to fitting $C\cos(\omega_{\mathrm{SD}}\tau + \Phi)$, where C and Φ are the amplitude and phase, respectively. We then define the complex demodulate as $C = \sqrt{A^2 + B^2}$. This positive definite amplitude follows the envelope of the modulated sinusoidal with frequency ω_{SD}. Note that this definition differs from the complex demodulation used in Geoffroy and Nycander (2022), where the authors fitted $C\cos(\omega_{\mathrm{SD}}\tau)$, i.e. with $\Phi = 0$. In practice, we compute the complex demodulate of the autocovariance following the harmonic analysis method of Cherniawsky et al. (2001). The demodulation over a given 48-h window is performed in two iterations. We first fit $A\cos(\omega_{\mathrm{SD}}\tau) + B\sin(\omega_{\mathrm{SD}}\tau)$ to the signal and compute the root mean square error (RMSE). We then repeat the fitting with a trimmed signal that excludes values outside of $\pm 2\,\mathrm{RMSE}$ from the previously fitted curve.

Complex demodulation is just a convenient way of finding the envelope of the sample estimate of an underlying true oscillating function at a given frequency. However, as an estimate of the envelope of the true oscillating function ~~(computed from a limited sample size)~~, the complex demodulate can be shown to be biased high (see Appendix A). This bias is greatly mitigated (*i*) at short time lags, and (*ii*) when the sample size is large (e.g., when demodulating regional or global mean autocovariances). Throughout this paper we only rely on complex demodulation in one case or the other. A ~~reasonable~~ conservative estimate of the uncertainty of the envelope of the sample function (~~as an estimate of the true function~~i.e., the uncertainty of the demodulate) is the uncertainty of the sample function itself. If this uncertainty range is larger than the envelope of the sample function, the conclusion is not that the envelope of the true function can be negative, but that it is not significantly different from zero. Here, we evaluate the uncertainty affecting the complex demodulate of a mean autocovariance series over a 48-h window by computing the median of the standard error over that window, hereinafter denoted $\widetilde{\mathrm{SEM}}$. The corresponding 95% confidence intervals are taken as $\pm 2\,\widetilde{\mathrm{SEM}}$.

Following Geoffroy and Nycander (2022), the semidiurnal IT variance can be estimated from the first 48-h demodulate at the semidiurnal frequency. In Appendix B we investigate the potential contamination of the first demodulate by the non-tidal variability (noise). We found that a background noise can contribute either positively or negatively (a consequence of filtering the time series) to the amplitude of the first demodulate, depending on its characteristic timescale. Notwithstanding, for typical signal-to-noise ratio and timescale of the background noise observed by Argo floats, the first demodulate was seen to remain a conservative estimate of the IT variance. The effects of the non-tidal variability on the IT variance estimate as computed in Geoffroy and Nycander (2022) thus do not put their results into question.

Since HYCOM does not resolve the full spectrum of the oceanic variability, in particular the variability associated with short timescales, the corresponding stochastic noise is expected to differ from the one captured by the in situ data (both in terms of variance and characteristic timescale). The contamination of the first demodulate by this noise would therefore account for a systematic bias when comparing the simulated data with observations. To limit this, we chose to consistently subtract an estimate of the autocovariance of the non-tidal variability from the sample mean autocovariance before computing the complex

demodulates. The way we obtain such an estimate is described in Appendix C. Namely, we fit a model of the autocovariance
200   of a tidal variability on top of a background noise to the sample autocovariance by nonlinear least squares. Besides the noise
parameters, this also provides us with an estimate of the tidal variance. Still, in this work, we chose to perform the comparisons
in terms of the first demodulate for it is a conservative and more robust estimate of the IT variance.

The result of the complex demodulation at the semidiurnal frequency applied to our two (noise corrected) mean autocovariance series is presented in Fig. 2b. The first 48-h complex demodulate is ~~taken as~~ a direct estimate of the ~~total~~ semidiurnal IT

205   variance. From the Argo data plotted in Fig. 2b we obtain ~~23.3~~ 24.9 m^2 with a 95% C.I. of ~~[15.0, 31.6]~~ [16.7, 33.2] m^2. For this local example, the first demodulate of $\overline{R}_{\mathrm{HYCOM}}^{L}$ is almost identical to the first demodulate of $\overline{R}_{\mathrm{Argo}}$. Apart from the first couple of demodulates, the HYCOM demodulate series appears consistently smaller than the Argo one.
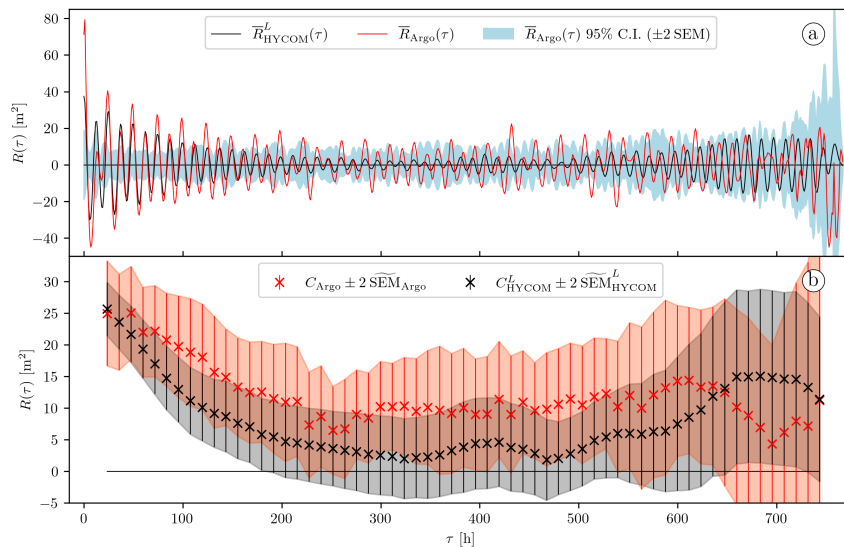
Figure 2. (a) Local mean autocovariance $\overline{R}_{\mathrm{HYCOM}}^{L}$ computed from the HYCOM particles (solid black curve), and local mean autocovariance $\overline{R}_{\mathrm{Argo}}$ computed from the 32-day segments of Argo data (solid red curve) and 95% confidence interval of $\overline{R}_{\mathrm{Argo}}$ (light blue shading). The data are from the geographical patch shown in Fig. 1. ~~The Argo variance lies above the figure scale, here $\overline{R}_{\mathrm{Argo}}(0) \simeq 121\,\mathrm{m}^2$.~~ (b) Complex demodulates at the semidiurnal frequency of the autocovariance series shown in (a) and their 95% C.I. The red and grey shadings highlight the 95% C.I. for the Argo and Lagrangian HYCOM data, respectively.

3.3   Eulerian perspective

The decay with time lag of the demodulates represented as red crosses in Fig. 2b mirrors the decorrelation captured by the
210   Lagrangian sampling of the Argo floats. The motion of the instruments results in decorrelating effects that cannot be isolated from the loss of coherence of the IT. However, by analyzing the HYCOM data within a Eulerian framework, one can directly monitor the decorrelation of the IT. The Eulerian HYCOM time series can then be compared with observations from moorings.

In the Eulerian framework, our methods remain practically unchanged. We now define the vertical isotherm displacement at a given location as

215
$$\eta^E_{1000}(x,y,t) = \frac{T(x,y,t) - \overline{T}(x,y)}{(\mathrm{d}T/\mathrm{d}P)_{1000}(x,y)} \tag{4}$$

where $\overline{T}(x,y)$ is the time average of the temperature field $T(x,y,t)$, and $(\mathrm{d}T/\mathrm{d}P)_{1000}(x,y)$ is the temperature gradient at 1,000 m, calculated from the modeled monthly-mean 3D temperature field. As for the Lagrangian time series, we apply a ~~fourth order~~ fourth-order Butterworth high-pass filter with a cut-off frequency of 0.3 cpd on η^E_{1000}.

For each HYCOM particle, we compute 32-day long time series of η^E_{1000} at the successive positions of the particle subsam-
220 pled at a 12-h rate. We then calculate the sample autocovariance from each of these time series and average the results over the particle's trajectory. As in the Lagrangian framework, the resulting averages can be averaged over different particles to compute local and global mean autocovariance series. By estimating the Eulerian autocovariance along the Lagrangian trajectories we account for the geographic variability of the IT. Also, our Eulerian sample autocovariance estimates contain many more degrees of freedom than their Lagrangian equivalents; As a result, the uncertainty affecting the Eulerian autocovariance
225 estimates is much smaller.

Eq. (4) is also used to compute vertical displacement time series from the mooring temperature records. Here, the vertical temperature gradient is computed from the annual mean climatology (WOA18; Boyer et al., 2018) as an average of the temperature gradient within 100 m of the instruments' depth. As for the HYCOM and Argo data, we discarded instruments for which the magnitude of the temperature gradient is smaller than ~~-3×10^{-5}~~ 3×10^{-5} °C dbar^{-1}. Each mooring time series is split
230 into successive 32-day segments. We then compute the sample autocovariance for each high pass filtered segment and average them to obtain a mean autocovariance.

The Eulerian sampling serves two main purposes: (*i*) validating the variance of the IT measured by the Lagrangian particles, and (*ii*) comparing the decorrelation of the IT in the HYCOM data ~~to~~ with mooring observations. We illustrate these two aspects for the local example introduced in Sect. 3.2 in Fig. 3 ~~,~~ and 4, respectively.

235 Fig. 3 shows the local mean autocovariance series at 1,000 dbar computed from both the Lagrangian η^L_{1000} (solid black curve) and Eulerian η^E_{1000} (solid red curve). As expected, the two autocovariance series demonstrate a close agreement at short time lags, before the motion of the particles causes the Lagrangian $\overline{R}^L_{\mathrm{HYCOM}}$ to decay faster: The first demodulate of the Eulerian $\overline{R}^E_{\mathrm{HYCOM}}$ (red crosses in Fig. 3b), reading ~~26.0~~ 28.4 m^2 with a 95% C.I. of ~~[25.3, 26.8]~~ [27.6, 29.1] m^2, is similar to the first demodulate of the Lagrangian $\overline{R}^L_{\mathrm{HYCOM}}$ (black crosses in Fig. 3b and in Fig. 2b). In contrast with $\overline{R}^L_{\mathrm{HYCOM}}$, $\overline{R}^E_{\mathrm{HYCOM}}$ is not
240 affected by the apparent decorrelation, and it remains close to the mean autocovariance computed from the (Eulerian) mooring No. 2 time series $\overline{R}_{\mathrm{Moor}}$ at all time lags (see Fig. 4). In conclusion, for this local example, the model agrees very well with the in situ observations, both in terms of variance and decorrelation of the semidiurnal IT at 1,000 dbar.
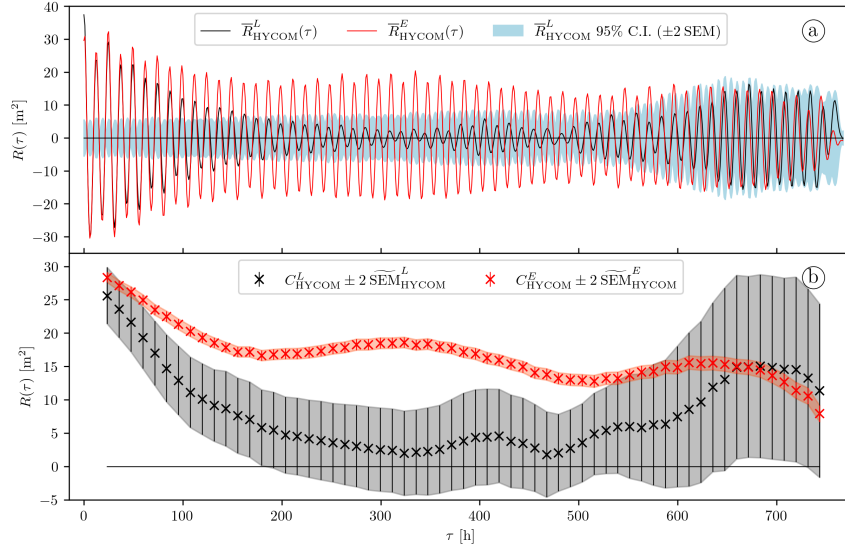
Figure 3. (a) Mean autocovariance computed from η_{1000}^L (solid black curve, same as the solid black curve in Fig. 2a) and 95% confidence interval of $\overline{R}_{\mathrm{HYCOM}}^L$ (light blue shading), and mean autocovariance computed from η_{1000}^E (solid red curve). The data are from the geographical patch shown in Fig. 1. (b) Complex demodulates at the semidiurnal frequency of the autocovariance series shown in (a) and their 95% C.I. The red and grey shadings highlight the 95% C.I. for the Eulerian and Lagrangian HYCOM data, respectively.

4 Results

4.1 The semidiurnal IT variance at 1,000 dbar

245 We bin the global collection of HYCOM particles based on their median position, using circular geographical patches of radius 200 km centered on a regular 2.5°×2.5° grid. Here, we use only particles for which (*i*) the mean speed is lower than 0.1 m s^{-1} (to avoid contamination by lee waves), and (*ii*) the variance of ~~η_{1000}~~ η_{1000}^L is lower than ~~2×10^4~~ 1×10^4 m^2. The ~~latter criteria accounts for the potential instability of the calculated η_{1000} values whenever the vertical temperature gradient,i.e. the denominator in Eq. (1) or (4), is small.~~ second criterion accounts for 41 discarded particles (out of 41,644) that we consider

250 as outliers, most of which are located in the Labrador Basin or in shallow water close to Antarctica. Including it instead would only marginally affect our results. We did not investigate more the cause of these extreme values.

As in Sect. 3.2 and 3.3, we compute an autocovariance series for each HYCOM particle, in the Lagrangian and Eulerian framework, separately. Then, we average these autocovariances over the corresponding patches to obtain local mean autocovariance series $\overline{R}_{\mathrm{HYCOM}}^L$ and $\overline{R}_{\mathrm{HYCOM}}^E$. For each geographical patch, we get local estimates of the ~~total~~ semidiurnal IT variance

255 from the first 48-h complex demodulate of $\overline{R}_{\mathrm{HYCOM}}^L$ and $\overline{R}_{\mathrm{HYCOM}}^E$.

We start by checking how the Lagrangian sampling affects the ~~total~~ semidiurnal IT variance. Fig. 5 shows the first 48-h complex demodulate of $\overline{R}_{\mathrm{HYCOM}}^E$ plotted as a function of the first demodulate of $\overline{R}_{\mathrm{HYCOM}}^L$ for our collection of geographical
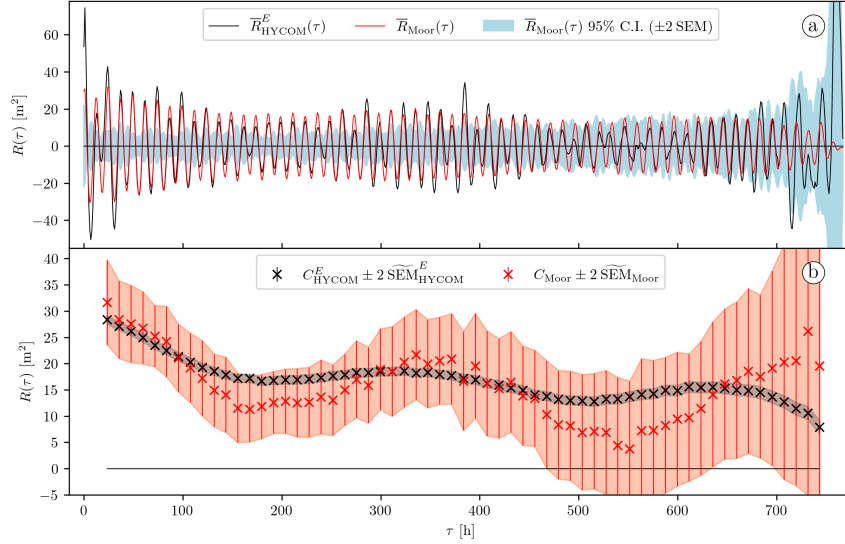
Figure 4. (a) Mean autocovariance computed from η_{1000}^{E} (solid black curve, same as the solid red curve in Fig. 3a), and mean autocovariance computed from eleven 32-day segments of the mooring No. 2 time series (solid red curve) and 95% confidence interval of $\overline{R}_{\mathrm{Moor}}$ (light blue shading). (b) Complex demodulates at the semidiurnal frequency of the autocovariance series shown in (a) and their 95% C.I. The red and grey shadings highlight the 95% C.I. for the mooring and Eulerian HYCOM data, respectively.

patches. The agreement is close to perfect (with a ~~Pearson's r squared, or~~ coefficient of determination ~~$r^2 \simeq 0.98$~~ $r^2 \approx 0.99$ and 0.76 in log-log ~~domain and $r^2 \simeq 0.74$ in linear domain~~ and linear domain, respectively). Thus, the motion of the Lagrangian

260   particles, and therefore of the Argo floats, have no significant impact on the measured ~~total~~ variance of the IT.

We can then map the semidiurnal IT variance (here taken as the first 48-h complex demodulate of $\overline{R}_{\mathrm{HYCOM}}^{L}$) and the associated $\widetilde{\mathrm{SEM}}$ on a 2.5°×2.5° grid (see Fig. 6a and 6b, respectively). In each figure we show only the bins which yield an IT variance larger than one $\widetilde{\mathrm{SEM}}$. Note that the latter criterion is less binding than the one used for the global maps in Geoffroy and Nycander (2022) since the uncertainties are smaller here, by definition. Fig. 6a can be directly compared with the global map

265   of the semidiurnal IT variance computed from Argo data (see Fig. 6c and 6d, updated from Geoffroy and Nycander, 2022). As documented in Buijsman et al. (2020), HYCOM is known to be subject to a thermobaric instability (TBI) in the north Pacific (dashed black rectangle in Fig. 6a). Since only a few patches of Argo data are available at the edge of this TBI area, we do not exclude it from the subsequent analysis.

The main patterns visible in Fig. 6a and Fig. 6c broadly agree, with energy radiating away from low-mode IT generation

270   hotspots, namely near Madagascar, Hawaii, and the tropical south and southwest Pacific. In Fig. 7a we show the Argo derived semidiurnal IT variance plotted as a function of the ~~simulated one~~ corresponding one from HYCOM. The two datasets are well correlated ($r^2 = 0.52$ and 0.38 in log-log ~~domain~~ and linear domain, respectively), but the semidiurnal IT variance is systematically smaller in HYCOM than in the Argo data.
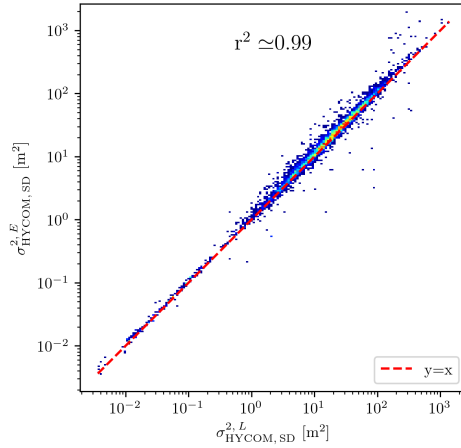
Figure 5. ~~First 48-h complex demodulate of $\overline{R}_{\mathrm{HYCOM}}^{E}$~~ Semidiurnal IT variance estimated from the Eulerian HYCOM data as a function of the ~~first demodulate of $\overline{R}_{\mathrm{HYCOM}}^{L}$~~ semidiurnal IT variance estimated from the Lagrangian HYCOM data for the unmasked bins in Fig. 6a. A warmer color indicates a ~~denser scatter~~ larger density.
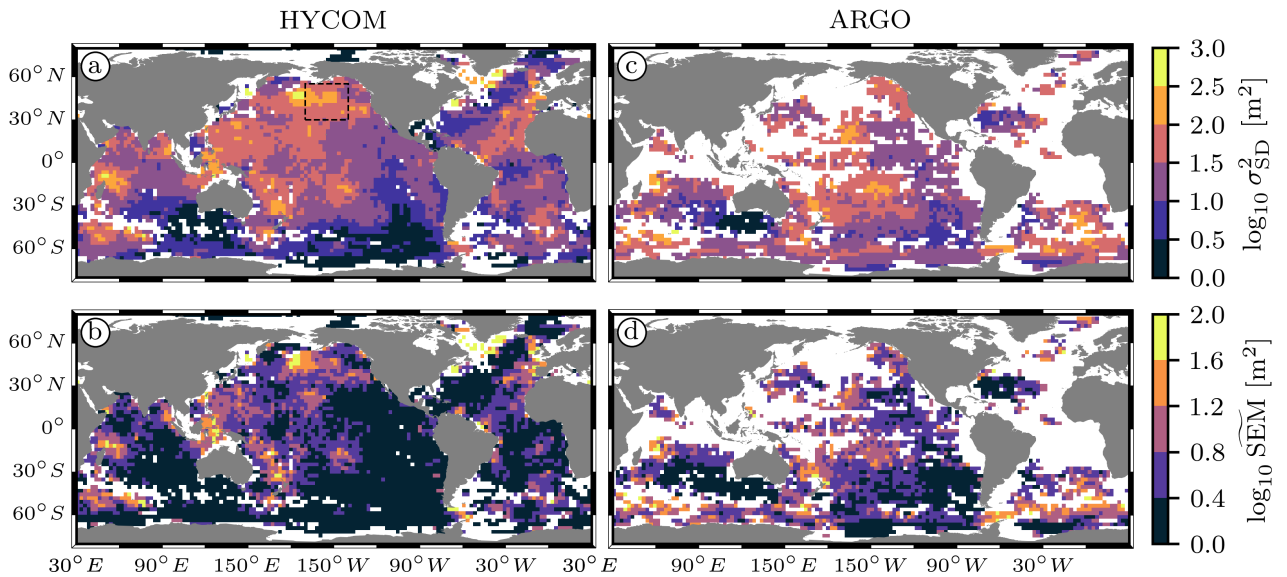
Figure 6. (a) Atlas of the ~~total~~ semidiurnal IT variance computed as the first 48-h complex demodulate of the local mean autocovariance series at 1,000 dbar ($\overline{R}_{\mathrm{HYCOM}}^{L}$), and (b) corresponding $\widetilde{\mathrm{SEM}}$. (c) and (d) Same as (a) and (b), respectively, but computed from the Argo data ($\overline{R}_{\mathrm{Argo}}$). (c) and (d) are updated from Geoffroy and Nycander (2022). The area where the simulation is affected by the TBI is shown by the dashed black rectangle in (a).

We investigate this bias by looking at the geographical distribution of the HYCOM to Argo semidiurnal IT variance ratio. For presentation purposes, instead of the latter ratio we plot the proxy $\sigma^2_{\mathrm{HYCOM}}/(\sigma^2_{\mathrm{HYCOM}} + \sigma^2_{\mathrm{Argo}})$ in Fig. 7b. This statistic is robust to outliers and maps a pair of variances to the closed range $[0,1]$. The ratio is relatively homogeneous globally (in the range $[0.25, 1.5]$, corresponding to $[0.2, 0.6]$ for our proxy), except over the Southern Ocean where the Argo inferred IT variance is significantly larger. Furthermore, we show the zonal mean of the Argo and HYCOM derived semidiurnal IT variance as a function of latitude in Fig. 7c. The zonal mean variance from Argo is generally larger, except around $40°$ N where the HYCOM inferred zonal mean variance peaks at ~~1.7~~ 1.6 times the Argo one. Discarding the data in the TBI area has virtually no effect on this peak (not shown). More noteworthy than this localized feature, the ratio of the HYCOM to Argo zonal mean variances ~~increases~~ decreases approaching the poles (poleward of $\pm 50°$ of latitude, not shown). In contrast, equatorward of $\pm 50°$, this ratio remains fairly constant, oscillating around a mean value of ~~0.74~~ 0.73 with a standard deviation of ~~0.27~~ 0.26. The global mean and standard deviation of the ratio are ~~0.62~~ 0.61 and 0.33, respectively.

The representativity of the zonal mean variances is smaller north of $40°$ N, because of the scarcity of available data (solid red curve in Fig. 7c). We therefore focus on the pronounced discrepancy affecting the Southern Ocean. We gather the data available over the unmasked bins in Fig. 7b into two groups, north and south of $50°$ S. For each group and the global collection of bins, we compute a mean autocovariance by averaging the corresponding Argo and HYCOM local mean autocovariance series (see Fig. 8). In Table 1, we summarize the ~~semidurnal~~ semidiurnal IT variance values computed as the first 48-h demodulate of the mean autocovariance for each group. A significant fraction of the divergence visible in the global mean autocovariance at short time lags (see Fig. 8a and 8b) can be explained by the larger discrepancy south of $50°$ S (see Fig. 8e and 8f). In this region, the Argo derived semidiurnal IT variance is close to ~~4~~ 3.5 times larger than the simulated one. The agreement is better north of $50°$ S, where the corresponding factor is 1.5 (see Fig. 8c and 8d). In a separate section we will discuss possible explanations for both the lower semidiurnal IT variance in HYCOM globally and the even lower simulated IT variance in the Southern Ocean.

Table 1. Semidiurnal IT variance from the autocovariance series plotted in Fig. 8.

		Argo	HYCOM	HYCOM/Argo
σ^2_{SD}	Global	~~35~~ 37 m^2	~~20~~ 22 m^2	~~0.57~~ 0.58
	North of $50°$ S	~~35~~ 38 m^2	~~24~~ 26 m^2	~~0.68~~ 0.69
	South of $50°$ S	~~34~~ 37 m^2	~~9~~ 11 m^2	~~0.28~~ 0.29

4.2 The decorrelation of the semidiurnal IT at 1,000 dbar

In contrast to Argo floats, the Eulerian sampling of moorings allows us to directly monitor the decorrelation of the IT. In a procedure similar to that in Sect. 4.1, we bin the global collection of HYCOM particles based on their median position, but this time using geographical patches (of radius 200 km) centered on the available moorings. We compute a sample autocovariance in the Eulerian framework for each particle, and average them within the corresponding patches to obtain local mean autoco-variance series $\overline{R}^E_{\mathrm{HYCOM}}$. The mooring time series in a given patch is split into successive 32-day segments. We then compute a
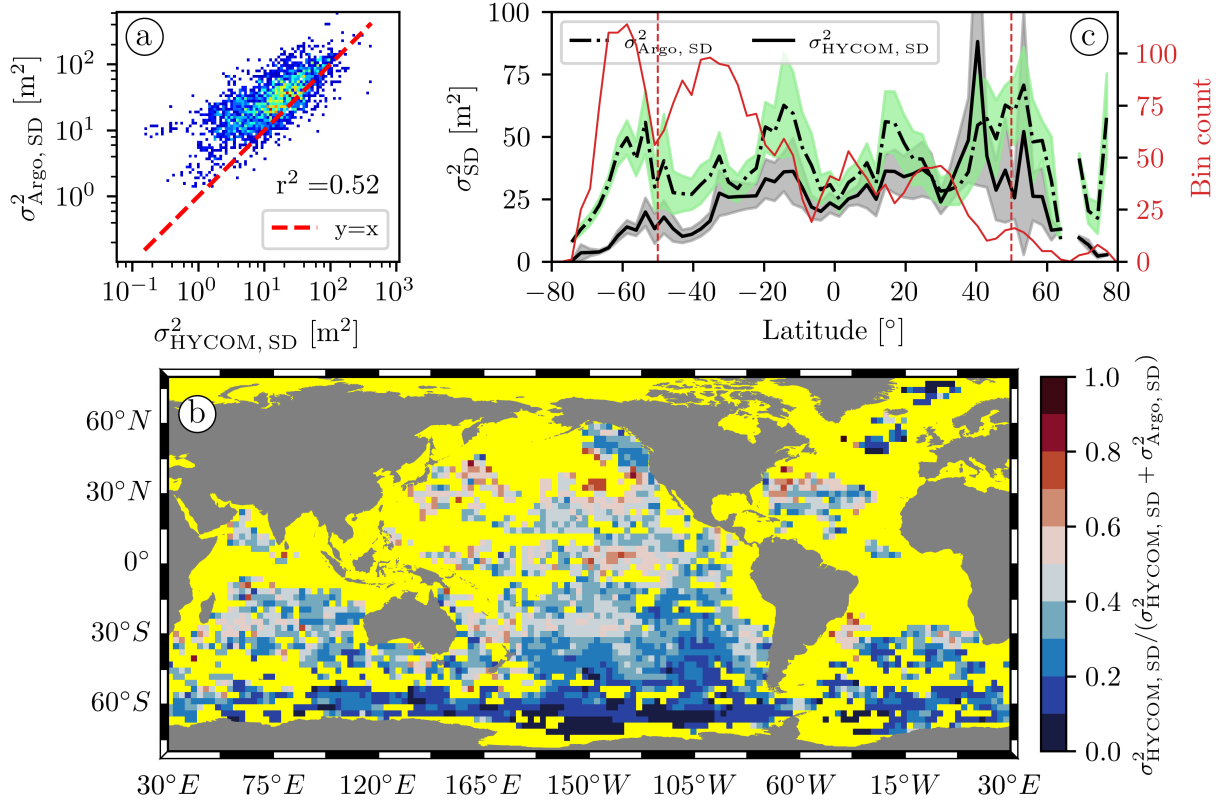
Figure 7. (a) Argo derived semidiurnal IT variance as a function of the ~~simulated~~ corresponding one from HYCOM for the collection of geographical bins shown in Fig. 7b. A warmer color indicates a ~~denser scatter~~larger density. (b) Atlas of $\sigma^2_{\text{HYCOM, SD}}/(\sigma^2_{\text{HYCOM, SD}} + \sigma^2_{\text{Argo, SD}})$, a proxy for the HYCOM to Argo semidiurnal IT variance ratio at 1,000 dbar. The value of this proxy is close to 0 where $\sigma^2_{\text{HYCOM, SD}} \ll \sigma^2_{\text{Argo, SD}}$, 0.5 where $\sigma^2_{\text{HYCOM, SD}} \sim \sigma^2_{\text{Argo, SD}}$, and 1 where $\sigma^2_{\text{HYCOM, SD}} \gg \sigma^2_{\text{Argo, SD}}$. The ocean mask ~~was~~ is colored in yellow for readability. (c) Zonal mean of the semidiurnal IT variance from Argo (dash-dotted black curve) and HYCOM (solid black curve) as a function of latitude and their respective 95% C.I. (light green and gray shading, respectively). The red curve represents the number of geographical bins used to compute the zonal mean at a given latitude (red axis on the right). The vertical dashed red lines are placed at 50° S and 50° N.

sample autocovariance for each segment and average them to obtain a mean autocovariance $\overline{R}_{\text{Moor}}$. Again, we use only particles for which (*i*) the mean speed is lower than 0.1 m s^{-1} (to avoid contamination by lee waves), and (*ii*) the variance of ~~η_{1000} is less than 2×10^4~~ η^L_{1000} is lower than 1×10^4 m^2 (outliers). After some additional quality controls, mainly discarding bins where either the mooring or HYCOM complex demodulates fall under one $\widetilde{\text{SEM}}$ within 15 days of time lag, we are left with ~~167~~172 moored instruments and the corresponding patches of simulated data.

For all the datasets used in this study, we found the probability distribution of the global collection of local mean autocovariance at any given ~~timelag~~ time lag to be skewed (not shown). This is not an issue when computing average statistics from
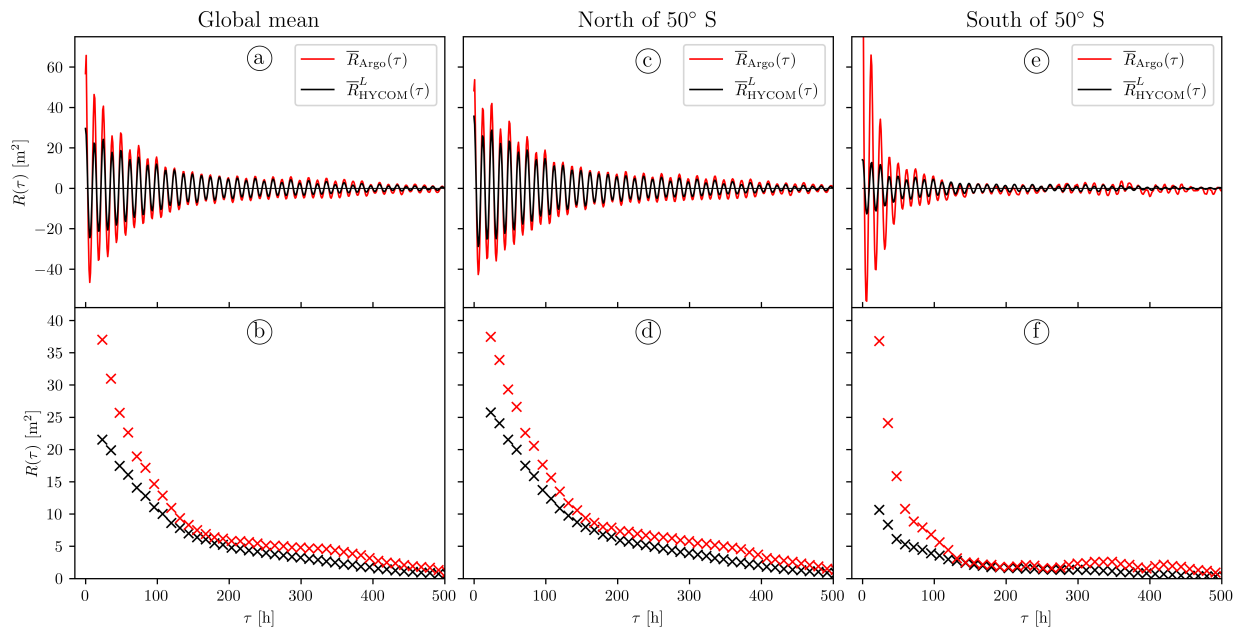
Figure 8. (a) Global mean autocovariance at 1,000 dbar computed from Argo data (solid red curve) and Lagrangian HYCOM particles (solid black curve), and (b) corresponding complex demodulates at the semidiurnal frequency. The uncertainties are vanishingly small (not shown). (c) and (d) Same as (a) and (b), respectively, but using only data north of $50°$ S. (e) and (f) Same as (a) and (b), respectively, but using only data south of $50°$ S. The Argo variance lies above the figure scale, here $\overline{R}_{\text{Argo}} \approx 173 \text{ m}^2$. In this figure, we truncated τ at 500 h since past this limit the autocovariance series are close to 0.

the Argo or the corresponding simulated data, since the number of samples (i.e., the number of geographical bins) is very large and the sample mean is therefore expected to be normally distributed, by virtue of the central limit theorem. In contrast,

310 geographical bins where mooring data are available are fewer. Thus, the influence of the tail of the distribution on the sample mean is larger when analyzing the relatively small collection of bins where mooring data are available than when considering the global collections of Argo and HYCOM data. This precludes the use of statistics that assume a normal distribution when describing regional or even global averages of the autocovariances computed from moorings. To limit the effects of skewness, we discard bins for which either the mooring or the simulated first 48-h demodulate is above the 95th percentile of its observed

315 distribution (here $P^{95}_{\text{Moor}} \approx 575 \text{ m}^2$ and $P^{95}_{\text{HYCOM}} \approx 330 \text{ m}^2$, respectively). The latter criterion accounts for 12 additional discarded bins, all located in moderate to strong mesoscale activity areas. Even after discarding these extreme samples, the data remain highly skewed.

The moorings may not offer as much spatial coverage as the Argo floats do, but they still provide an opportunity to validate the geographical variability of the semidiurnal IT variance in HYCOM. As in Sect. 4.1, we can get local estimates

320 of the semidiurnal IT variance from $\overline{R}^E_{\text{HYCOM}}$ and $\overline{R}_{\text{Moor}}$ in each geographical patch centered on a mooring. Fig. 9a shows the scatter plot of the first 48-h complex demodulate of $\overline{R}^E_{\text{HYCOM}}$ as a function of the

first demodulate of $\overline{R}_{\mathrm{Moor}}$ for our collection of geographical bins. As with the Lagrangian data (see Fig. 7a), the correlation is good (~~$r^2 = 0.53$~~ $r^2 = 0.51$ and 0.40 in log-log ~~domain~~and linear domain, respectively), but the semidiurnal IT variance is systematically smaller in HYCOM than in the mooring data. In Fig. 9b and 9c we show the geographical location of the patches along with a proxy for the HYCOM to mooring semidiurnal IT variance ratio, and the histogram of this proxy, respectively. The latter histogram shows that the distribution of the proxy is centered around ~~0.33~~0.37, corresponding to a ratio of ~~0.5~~0.6, approximately.
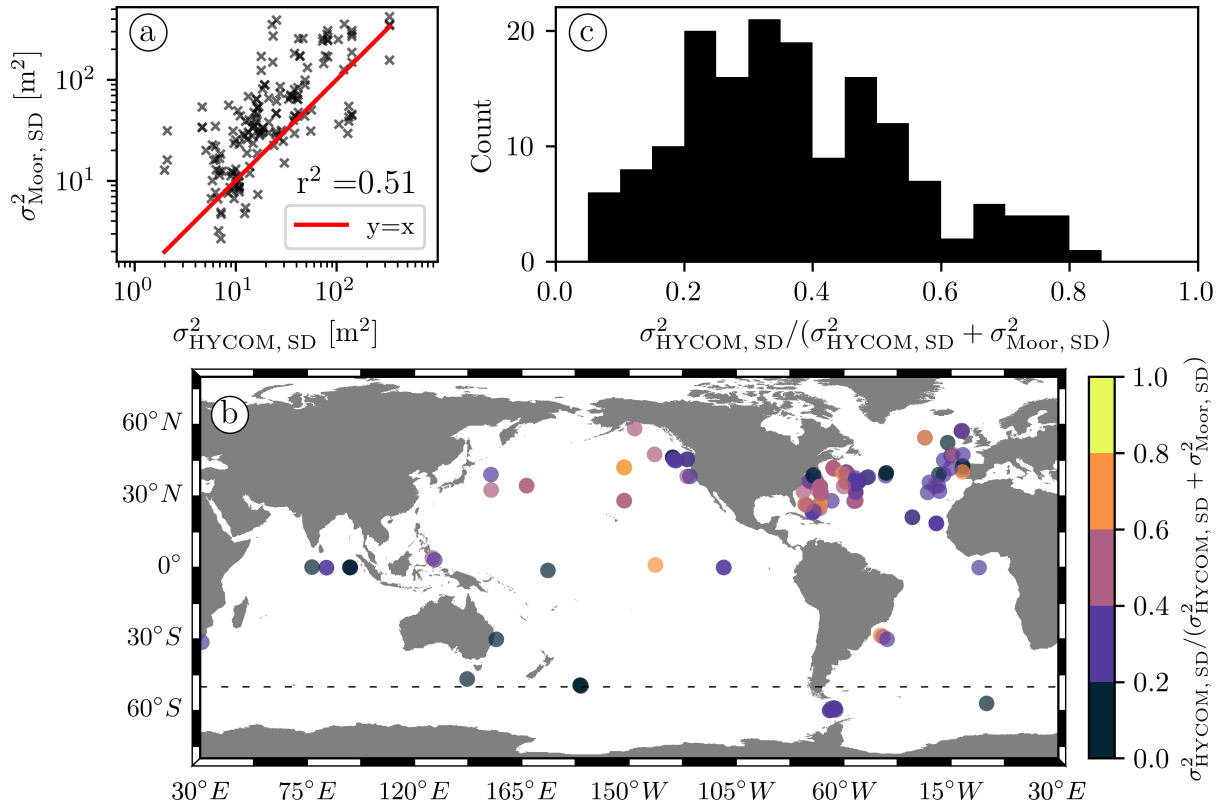
Figure 9. (a) Semidiurnal IT variance derived from mooring data as a function of the semidiurnal IT variance computed from the Eulerian HYCOM data. (b) Map of the ~~163~~ 160 geographical bins presented in (a) along with the proxy $\sigma^2_{\mathrm{HYCOM,\,SD}}/(\sigma^2_{\mathrm{HYCOM,\,SD}} + \sigma^2_{\mathrm{Moor,\,SD}})$ for the ratio of the semidiurnal IT variance computed from the Eulerian HYCOM and mooring data, and (c) the histogram of this proxy. An equivalent proxy was used in Fig. 7b. The dashed black line in (b) indicates $50°$ S.

To measure the strength of the decorrelation affecting the Eulerian mean autocovariances, we define the semidiurnal coherent variance fraction (SCVF_{15}) as the ratio between the 48-h demodulate at ~~$\tau \simeq 15$~~ $\tau \approx 15$ days (i.e., ~~half the~~ one spring-neap period) and the first demodulate. ~~Note that, due to the limiting duration of our time series, we cannot distinguish a strong but short from a weak but long decorrelating process.~~ The SCVF_{15} being calculated from the demodulate at relatively long

time lags, it is ~~meaningfull~~ meaningful only when estimated from a large enough sample size (i.e., with minimal uncertainty). Therefore, we start by considering all the data available over our global collection of geographical bins.

Fig. 10a displays the global mean autocovariances calculated from the mooring ($\overline{R}_{\mathrm{Moor}}$, solid red curve) and HYCOM data
($\overline{R}_{\mathrm{HYCOM}}$, solid black curve), as the average of the local mean autocovariance series over the bins presented in Fig. 9b. In Fig. 10b we show different statistics of the observed distribution of the demodulates of the local mean autocovariances in the form of boxplots as a function of time lag. Fig. 10b suggests that the mooring records exhibit both a larger semidiurnal IT variance and a stronger decorrelation on a global average: the mean of the first demodulates and the mean SCVF_{15} are ~~82~~ 76 m^2 and ~~0.50, and 43~~ 0.47, and 41 m^2 and ~~0.64~~ 0.57, for the moorings and HYCOM data, respectively. Using median values instead leads to the same conclusions (not shown).

We investigate a potential latitudinal dependence by plotting the mean autocovariance series computed separately from the geographical bins lying north (~~150~~ 149 instruments) and south (~~13~~ 11 instruments) of $50°$ S (see Fig. 10c and 10d, and Fig. 10e and 10f, respectively). The mean semidiurnal IT variance and the mean SCVF_{15} for each group are shown in Table 2. Again, the divergence between the two datasets appears enhanced south of $50°$ S. Moreover, the SCVF_{15} is larger in the Southern Ocean than elsewhere for HYCOM, but smaller for the moorings. This indicates a weaker, or slower, decorrelation of the IT in HYCOM in the Southern Ocean than in the global ocean.

The slower decorrelation of the IT in the simulation can be explained by some decorrelating processes, such as eddies or submesoscale variability, being weaker in HYCOM than in the real ocean. It could also be explained by the time variability of certain decorrelating processes. Our numerical simulation only spans May 20 to June 20, 2019. Therefore, it is potentially missing processes that would specifically occur or intensify at another time of the year (or in a different year). On the other hand, the mooring data span several decades. Thus, our single month of data from HYCOM may not be representative of the broader temporal sampling of the mooring data.

The spatial distribution of the moorings is sparse and tends to be denser in particular areas (e.g. the Gulf Stream region). This is all the more true south of $50°$ S , where much fewer moorings are available than in the northern region, and where they are mostly located in the Drake Passage. For both these reasons, the mean autocovariances computed from moorings north and especially south of $50°$ S cannot be considered truly representative of the IT in these vast regions. Nonetheless, they remain representative of the collection of geographical bins used to construct them.

4.3 Apparent decorrelation

In Sect. 4.1 we demonstrated that the motion of the floats has no significant impact on the measured total variance of the IT. However, the Doppler effect and spatial decorrelation induced by the Lagrangian sampling of the floats both act as decorrelating processes causing the autocovariance to decay with increasing time lag (Geoffroy and Nycander, 2022). Following Caspar-Cohen et al. (2022), we call this mechanism apparent decorrelation, as it is unrelated to the decorrelation of the IT. Geoffroy and Nycander (2022) estimated the apparent decorrelation timescale to be longer than that of the autocovariance function observed by Argo floats, on a global average. Thus, they concluded that the decay of the global mean autocovariance observed by Argo floats is primarily due to the decorrelation of the IT.
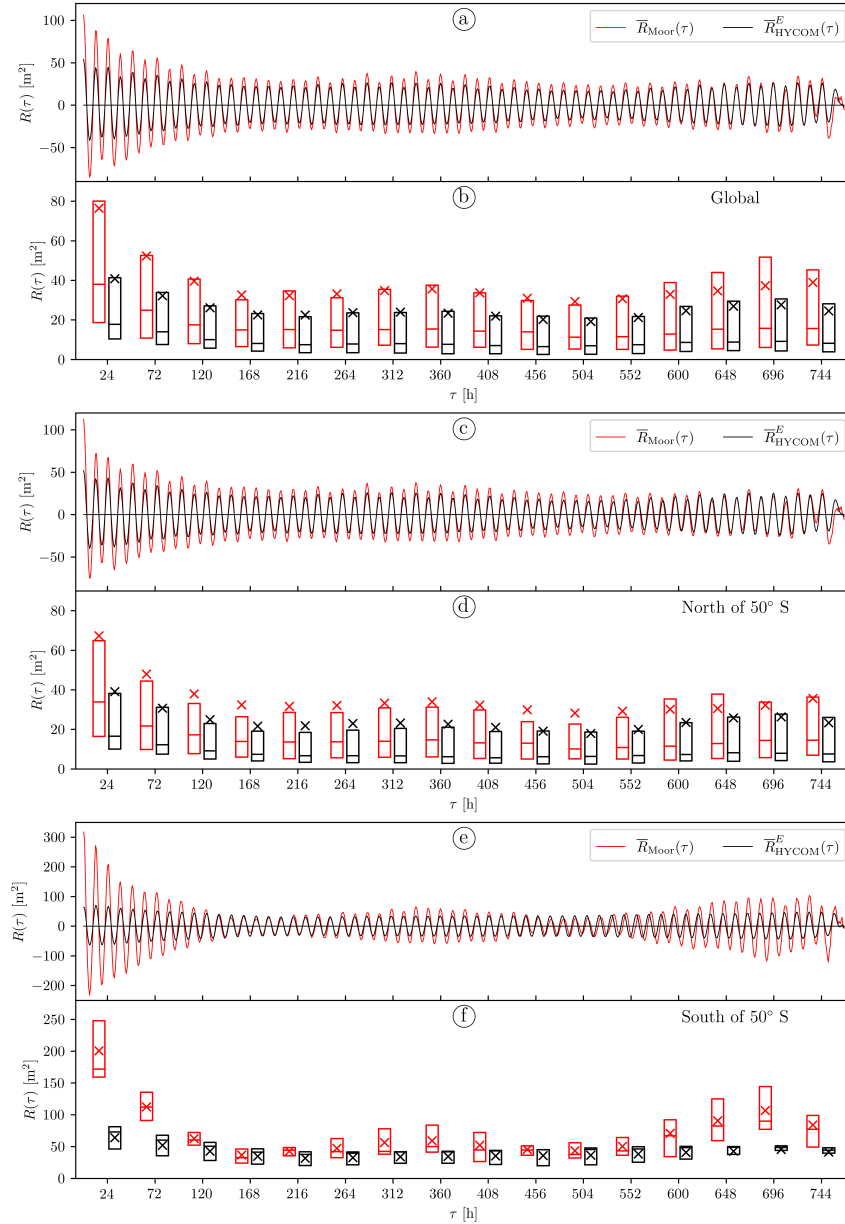
Figure 10. (a) Global mean autocovariance computed as the average of the local mean autocovariance series from the HYCOM particles over the geographical bins presented in Fig. 9b (solid black curve), and global mean autocovariance from the moored instruments computed in the same way (solid red curve); (b) Boxplots of the observed distribution of the complex demodulates at the semidiurnal frequency of the local mean autocovariances for the collection of geographical bins presented in Fig. 9b as a function of time lag. Each boxplot consists of a rectangle ~~extanding~~ extending from the first ~~quartile~~ to the third quartile of the data, with a line at the median and a cross at the mean. For a given time lag, the red and black boxplots, offset on either side of the time lag value, represent the distribution of the demodulates computed over the 48-h window centered on that time lag value from moorings and HYCOM, respectively. (c) and (d) Same as (a) and (b), respectively, but using only the ~~150~~ 149 geographical bins north of $50°$ S. (e) and (f) Same as (a) and (b), respectively, but using only the ~~13~~ 11 geographical bins south of $50°$ S.

Table 2. Summarizing numerics of the autocovariance series plotted in Fig. 10.

		Mooring	HYCOM	HYCOM/Mooring
Global	σ^2_{SD}	~~82~~ 76 m^2	~~43~~ 41 m^2	~~0.52~~ 0.53
	SCVF$_{15}$	~~0.50~~ 0.47	~~0.64~~ 0.57	~~1.27~~ 1.23
North of 50° S	σ^2_{SD}	~~69~~ 67 m^2	~~37~~ 39 m^2	~~0.54~~ 0.58
	SCVF$_{15}$	~~0.55~~ 0.50	~~0.62~~ 0.58	~~1.14~~ 1.15
South of 50° S	σ^2_{SD}	~~223~~ 201 m^2	~~103~~ 64 m^2	~~0.46~~ 0.32
	SCVF$_{15}$	~~0.34~~ 0.29	~~0.70~~ 0.54	~~2.05~~ 1.83

The HYCOM data allow to study this by directly comparing the global mean autocovariance computed in the Lagrangian and Eulerian frameworks (see Fig. 11a and 11b). As expected, the Lagrangian ($\overline{R}^L_{\mathrm{HYCOM}}$, solid black curve in Fig. 11a) and Eulerian ($\overline{R}^E_{\mathrm{HYCOM}}$, solid red curve) global mean autocovariances are virtually identical until ~~$\tau \simeq 50$~~ $\tau \approx 50$ h. Past this limit, the more rapid decay of $\overline{R}^L_{\mathrm{HYCOM}}$ can only be caused by the apparent decorrelation, while $\overline{R}^E_{\mathrm{HYCOM}}$ continues to solely reflect the decorrelation of the IT. Our aim in the present section is to find estimates for the characteristic timescales of the apparent decorrelation and decorrelation of the IT in HYCOM and compare it with observations.

Taking inspiration from Geoffroy and Nycander (2022) and Caspar-Cohen et al. (2022), we try a simple model for the complex demodulate at the semidiurnal frequency of the Eulerian autocovariance computed over a time lag window centered on τ:

$$\mathcal{C}^E_{\mathrm{SD}}(\tau) = \sigma^2_{\mathrm{SD}}\big(\alpha + (1-\alpha)\exp(-\tau/T)\big) + \sigma^2_{\mathrm{AM}}\big(\cos(\omega_{\mathrm{AM}}\tau) - 1\big). \tag{5}$$

Here, σ^2_{SD} is the total semidiurnal internal tide variance, α is the stationary fraction, T is the IT decorrelation timescale, and σ^2_{AM} and ω_{AM} are the variance and angular frequency of an amplitude modulating sinusoidal, respectively. A heuristic model for the demodulate of the corresponding Lagrangian autocovariance is then

$$\mathcal{C}^L_{\mathrm{SD}}(\tau) = \mathcal{C}^E_{\mathrm{SD}}(\tau)\exp(-\tau/T_{\mathrm{app}}), \tag{6}$$

where T_{app} is the apparent decorrelation timescale.

A constrained least squares fit of the model Eq. (5) to the complex demodulate series computed from $\overline{R}^E_{\mathrm{HYCOM}}$ (red crosses in Fig. 11b) yields ~~$T = 104$~~ $T = 94$ h. The fitted model as well as the exponential decay due to the IT decorrelation (i.e., only the term proportional to σ^2_{SD}) are represented in Fig. 11b as the solid and dashed red curves, respectively. On a global average, 95% of the decorrelation of the nonstationary IT in HYCOM is therefore achieved within 3~~$T \sim 300$~~ $T \approx 300$ h (for ~~$\exp(-3) \sim 0.05$~~ $\exp(-3) \approx 0.05$), i.e. well under the 32 days of data.

By dividing $\overline{R}^L_{\mathrm{HYCOM}}$ by $\overline{R}^E_{\mathrm{HYCOM}}$, we can isolate the effects of the apparent decorrelation on $\overline{R}^L_{\mathrm{HYCOM}}$. In Fig. 11c we plot this ratio after applying a median filter with a window of 18 h (solid black curve). A least squares fit of a simple decaying exponential to the obtained curve yields an apparent decorrelation timescale ~~$T_{\mathrm{app}} = 213$~~ $T_{\mathrm{app}} = 209$ h (dashed black curve).
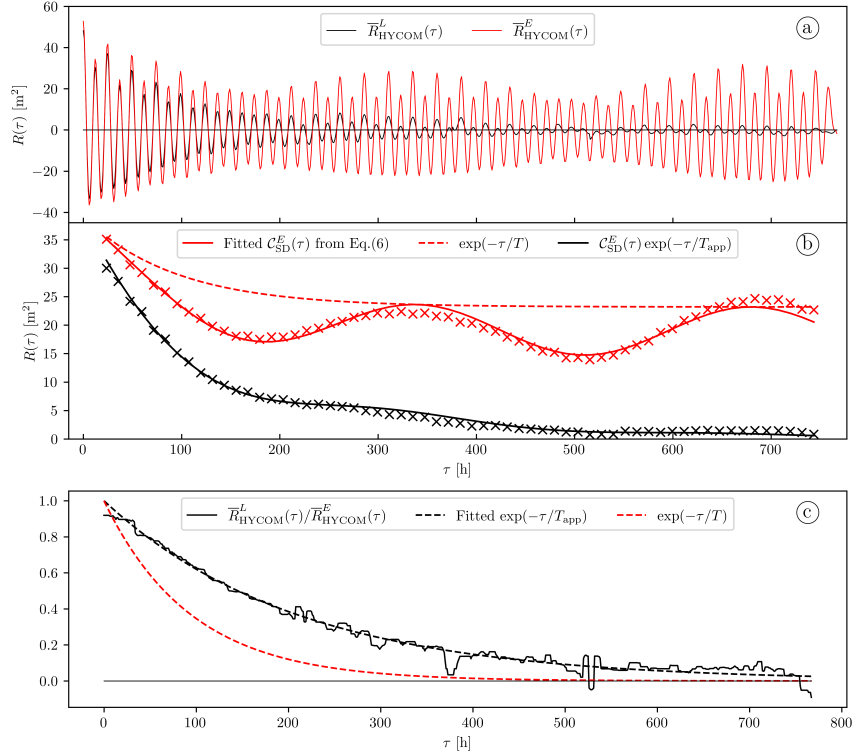
Figure 11. (a) Global mean autocovariance at 1,000 dbar computed from the global collection of HYCOM derived η^L (solid black curve) and η^E (solid red curve) time series, and (b) corresponding complex demodulates at the semidiurnal frequency. The uncertainties are vanishingly small (not shown). The solid and dashed red curves represent the result of fitting the model Eq. (5) to the Eulerian demodulates and the underlying exponential decay due to the decorrelation of the IT, respectively. The solid black curve corresponds to the model Eq. (6) with the different parameters set as described in the text. (c) Median filtered ratio of $\overline{R}_{\mathrm{HYCOM}}^L$ to $\overline{R}_{\mathrm{HYCOM}}^E$ (solid black curve) and fitted decaying exponential with the apparent decorrelation timescale (dashed black curve). For reference, we overlay a decaying exponential with the IT decorrelation timescale obtained by fitting the model Eq. (5) to the Eulerian demodulates (dashed red curve).

For verification, we compute the Lagrangian $\mathcal{C}_{\mathrm{SD}}^L$ from Eq. (6), by multiplying the fitted model of $\mathcal{C}_{\mathrm{SD}}^E$ (solid red curve Fig. 11b)

390   by the fitted exponential decay due to the apparent decorrelation (dashed black curve in Fig. 11c). The result (solid black curve in Fig. 11b) closely follows the demodulated Lagrangian autocovariance (black crosses in Fig. 11b). The different parameters obtained from the Eulerian and Lagrangian simulated data are summarized in Table 3. Note that the fitted amplitude modulating sinusoid is close to the spring-neap cycle, here ~~$\omega_{\mathrm{AM}} \sim |\omega_{\mathrm{M}_2} - \omega_{\mathrm{S}_2}| = 0.0177$~~ $\omega_{\mathrm{AM}} = |\omega_{\mathrm{M}_2} - \omega_{\mathrm{S}_2}| \approx 0.0177$ h^{-1}.

     ~~Lastly, we~~ We can compare the global mean values of T and T_{app} obtained from the simulation with the Argo observations.

395   The geographical coverage of the Argo data is less than that of HYCOM (roughly 55% in area). Hence, instead of the global collections of floats and Lagrangian particles, we now consider the intersection of the two datasets taken as the unmasked bins shown in Fig. 7b and the corresponding local mean autocovariances. These local mean autocovariances are averaged further

Table 3. Summary of the parameters estimated from the simulated Eulerian and Lagrangian data. These values are used to compute $\mathcal{C}^L_{\text{SD}}$ from the model Eq. (6) (solid black curve in Fig. 10b).

	σ^2_{SD}	α	T	σ^2_{AM}	ω_{AM}	T_{app}
~~37~~ Global	39 m^2	~~0.63~~ 0.60	~~105~~ 94 h	4 m^2	~~0.0186~~ 0.0185 h^{-1}	~~213~~ 209 h

to obtain a global mean autocovariance for each dataset that is representative of a common area. We then fit the model Eq. (6) and (5) to the demodulates of the Argo global mean autocovariance ~~computed from all the Argo floats available~~ (see Fig. 12,

400   updated from Geoffroy and Nycander, 2022). The ~~geographical coverage of the Argo data is less than that of HYCOM, but we are simply interested in a qualitative comparison of the two. The fitted parameters~~ parameters fitted to the Argo observations, as well as the parameters obtained by repeating the above procedure for the corresponding simulated data, are gathered in Table 4. The values of T_{app} and T ~~obtained~~ computed here, from Argo data only, are similar to those reported by Geoffroy and Nycander (2022), where the authors estimated it from a comparable collection of Argo floats and HRET (in particular,

405   the stationary limit was determined from HRET instead of by fitting). The fitted stationary fraction α, however, is about 3 times larger than previously reported. This could be explained by a biased low stationary variance (obtained by projecting HRET at 1,000 dbar) in Geoffroy and Nycander (2022). As from the simulated data, fitting our model to the Argo global mean demodulates yield a T shorter than T_{app}. While the Argo T_{app} is almost identical to the HYCOM value, T is about twice smaller.

410   ~~Parameters from the fitting of the model Eq. (6) to the demodulates of the global mean autocovariance computed from Argo data (see Fig. 12). σ^2_{SD} α T σ^2_{AM} ω_{AM} T_{app} 44 m^2 0.45 64 h 3 m^2 0.0182 h^{-1} 230 h~~
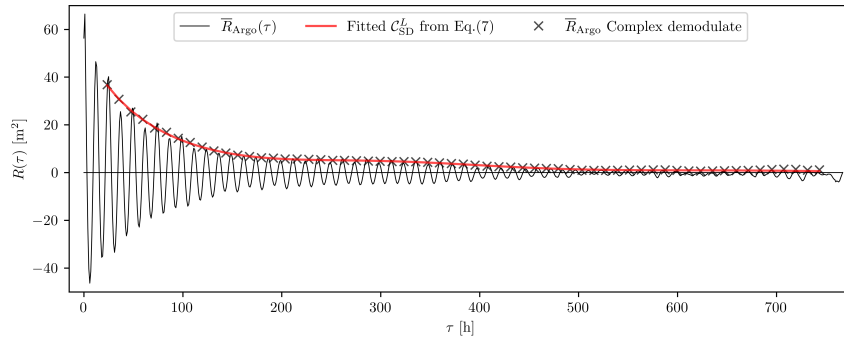
Figure 12. Global mean autocovariance at 1,000 dbar computed from ~~all~~ the Argo ~~floats available~~ data over the unmasked bins in Fig. 7b (solid black curve), and corresponding complex demodulates at the semidiurnal frequency (black crosses). The uncertainties are vanishingly small (not shown). The solid red curve represents the result of fitting the model Eq. (6) and (5) to the demodulates.

Similarly to Geoffroy and Nycander (2022), we conclude that the decorrelation of the IT is more rapid than the apparent decorrelation on a global average. Yet, according to the simulated data, it is only more rapid by a factor of 2, while this factor is 4 according to the Argo data. Hence, some decorrelating processes appear to be weaker in the simulation than in the real

415 ocean. As explained in Sect. 4.2, the slower decorrelation of the IT in HYCOM could also be explained by certain decorrelating processes being weaker than usual in the May 20 to June 20, 2019 period of outputs used here. Nevertheless, the decorrelation of the IT typically is at least as important as the apparent decorrelation over the first few days of time lag. This result might not hold true everywhere, since the geographical variability of T and α is expected to be large.

Lastly, Zaron (2022) provides a valuable comparison point for the results above. Along the same lines as in this section, he
420 uses the spatially averaged frequency spectrum of sea level observations from satellite altimetry to measure properties of both the baroclinic tidal peak and continuum (i.e., the coherent and incoherent IT variability, respectively). For the semidiurnal band, in the latitude range from $30°$ S to $30°$ N, the author found that about 62% of the IT variance eventually becomes decorrelated within 34 d. Assuming an exponential decay as in the model Eq. (5), this would represent a stationary fraction $\alpha \approx 0.38$ and a decorrelation timescale $T \approx 163$ h (for 99% of the nonstationary IT to decorrelate in $5T$). Repeating the fitting procedure above
425 for data limited to the $\pm 30°$ latitude range, we recover values in reasonable agreement from the Argo observations: $\alpha \approx 0.29$ and $T \approx 144$ h (see Table 4). The HYCOM fitted $\alpha \approx 0.64$ and $T \approx 105$ h agree less well with the results of Zaron (2022), but still are in the same ballpark. In this latitude range, the bins used for the fitting represent roughly 45% of the oceanic area.

Table 4. Parameters from the fitting of the model Eq. (6) to the demodulates of (*i*) the global mean autocovariance computed from the Argo data over the unmasked bins in Fig. 7b (see Fig. 12), and (*ii*) the mean autocovariance computed from the same bins but limited to the latitude range from $30°$ S to $30°$ N. These parameters are also estimated from the simulated Eulerian and Lagrangian data in the same locations.

		σ_{SD}^2	α	T	σ_{AM}^2	ω_{AM}	T_{app}
Argo	Global	$52\ \mathrm{m}^2$	0.43	51 h	$4\ \mathrm{m}^2$	$0.0177\ \mathrm{h}^{-1}$	207 h
	$\pm 30°$	$51\ \mathrm{m}^2$	0.29	144 h	$4\ \mathrm{m}^2$	$0.0185\ \mathrm{h}^{-1}$	242 h
HYCOM	Global	$26\ \mathrm{m}^2$	0.62	94 h	$3\ \mathrm{m}^2$	$0.0185\ \mathrm{h}^{-1}$	199 h
	$\pm 30°$	$37\ \mathrm{m}^2$	0.62	105 h	$4\ \mathrm{m}^2$	$0.0185\ \mathrm{h}^{-1}$	188 h

4.4 Potential sources of bias

Why is the IT variance lower and IT decorrelation weaker in HYCOM than in the observations, particularly in the Southern
430 Ocean ? At the time of writing we cannot think of a particular reason for either the Argo or the mooring derived IT variance to be biased high globally. ~~We did investigate~~ In particular, the correction accounting for the non-tidal variability we systematically subtract from the sample autocovariance precludes any contamination of the first demodulate by the background noise (see Appendix C).

We also investigated whether the bias in the Southern Ocean could be related to the contamination of the first 48-h demodu-
435 late at ω_{SD} by near-inertial waves as we approach the M_2 critical latitude (where $f = \omega_{M_2}$, at about $74°$ S). To check this, we can map the semidiurnal IT variance from the Argo data anew (as shown in Fig. 6c), this time fitting an additional $\cos(f\tau)$, where f is the local Coriolis frequency, to the local mean autocovariances. The result of the fit becomes unstable at $74°$ S, but the zonal mean of the demodulates at f does reach a maximum around $60°$ S while the zonal mean of the demodulates at ω_{SD}

remains unaffected (not shown). We conclude that the first 48-h demodulate at the ~~semidurnal~~ semidiurnal frequency is not affected by near-inertial waves, at any latitude.

As for the model, ~~equatorward of $\pm50°$ latitude and for seafloor depths deeper than 1250 m,~~ the horizontal grid spacing limits the number of vertical modes correctly resolved to the first five modes, equatorward of $\pm50°$ latitude and for seafloor depths deeper than 1250 m (Buijsman et al., 2020). Approaching the poles, the number of model layers below the mixed layer, and hence the vertical resolution, decreases. This further limits the number of resolved modes in HYCOM south of $50°$ S: roughly, modes 3 and 2 are only partially resolved poleward of $60°$ S and $65°$ S, respectively. Although the bulk of the IT variance at 1,000 dbar is likely related to mode-1 waves on a global average (Geoffroy and Nycander, 2022), the contribution from higher modes can become significant locally. In principle, both Argo floats, moorings, and HYCOM data include the effect of higher modes, but these are less well resolved in HYCOM, particularly in the Southern Ocean. Together with the assumption that higher-mode IT are less coherent (Egbert and Ray, 2017), this may explain the lower IT variance and weaker IT decorrelation in HYCOM than in the observations, both on a global average and more specifically in the Southern Ocean. It is also in line with the larger mean SCVF_{15} computed from HYCOM data in the Southern Ocean compared with the rest of the globe (indicating a weaker decorrelation there), whereas for moorings the SCVF_{15} is smaller in this region (see Table 2).

The mode-m vertical structure of the isopycnal displacement $\Phi_m(z)$ is obtained by solving the Sturm-Liouville problem

$$\frac{\mathrm{d}^2\Phi_m(z)}{\mathrm{d}z^2} + \frac{N^2(z)}{c_m^2}\Phi_m(z) = 0, \tag{7}$$

with the boundary conditions $\Phi_m(0) = \Phi_m(-H) = 0$, where H is the ocean depth, $N(z)$ is the buoyancy frequency profile, and $-1/c_m^2$ is the eigenvalue corresponding to the eigenfunction $\Phi_m(z)$ for mode-m (Gill, 1982). The modal partitioning of the IT energy at a given location is mainly determined by the conversion rate (both local and remote) and lifetime of each mode (de Lavergne et al., 2019). Additionally, depending on the local stratification and ocean depth, the parking depth at 1,000 dbar can be more or less close to the anti-node (point of maximal displacement) and node (point of no displacement) of the different vertical modes. Therefore, the normalized contribution of mode-m relative to mode-1 waves to the variance recorded at this depth is weighted by a coefficient γ_{m1}. In Appendix B of Geoffroy and Nycander (2022), the authors derived an expression for this coefficient:

$$\gamma_{m1}(z) = \frac{\Phi_m^2(z)\int_{-H}^{0} N^2\Phi_1^2\mathrm{d}z}{\Phi_1^2(z)\int_{-H}^{0} N^2\Phi_m^2\mathrm{d}z}. \tag{8}$$

In Fig. 13a and 13b we plot the global maps of γ_{21} and γ_{31} at 1,000 dbar, respectively, computed from Eq. (8) after solving the eigenvalue problem Eq. (7) for the $1/4°$ WOA18 summer climatology. Here, we extrapolated the climatological data down to the reference bathymetry from the General Bathymetric Chart of the Oceans 2019 (GEBCO, 2019) wherever the bathymetry is deeper than the deepest valid climatological record. A visual comparison with Fig. 7b in the Southern Ocean suggests that low HYCOM to Argo IT variance ratios (darker blue pixels in Fig. 7b) spatially coincide with large γ_{31}, mostly, or γ_{21} to some extent (especially in the Weddell Sea). South of $60°$ S, either γ_{31} or γ_{21} is larger than unity, hence the contribution from mode-3 or mode-2 waves to the IT variance recorded at 1,000 dbar is magnified compared to the contribution from mode-1 waves. However, modes 2 and higher are not well resolved in HYCOM at these latitudes.
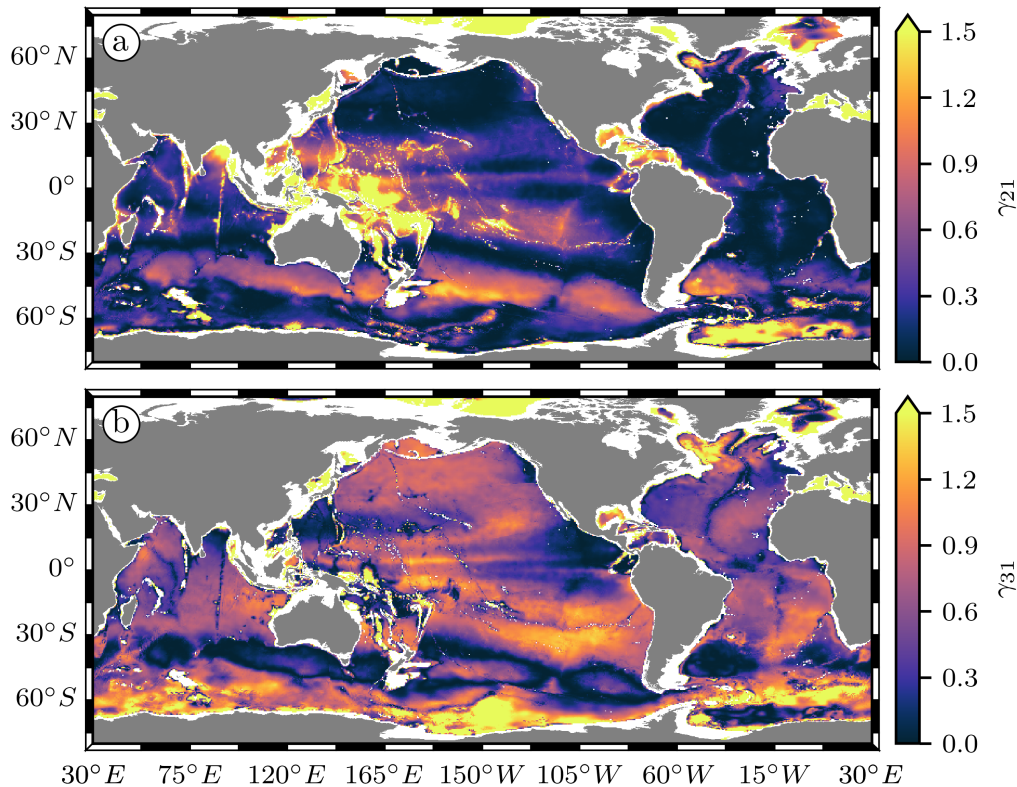
Figure 13. (a) Weight of the normalized contribution of mode-2 relative to mode-1 waves to the IT variance recorded at 1,000 dbar, computed from the WOA18 stratification. (b) Same as (a) but for the relative contribution of mode-3 waves.

The magnification of the contributions from modes 2 and 3 to the variance at 1,000 dbar in the Southern Ocean only affects how propagating IT are perceived at the Argo parking depth. This has no connection with the generation and dissipation processes that set the underlying modal partitioning of the IT energy. Additional explanations might be found by examining whether the main parameters affecting the generation of IT (namely the bottom topography, barotropic tidal forcing, and stratification) are less accurate in this region than in the rest of the globe.

The pattern of the enhanced discrepancies between Argo and HYCOM in the Southern Ocean (darker blue in Fig. 7) ~~could~~ might be correlated with the distribution of bathymetric features. For instance, the large values around $45°$ S, $105°$ W in the South Pacific are centered on the Chile Rise, a known IT generation area. To date, only about 19% of the global ocean seafloor has been mapped using shipborne techniques. Therefore, global bathymetric products widely rely on depth prediction from satellite gravimetry (Smith and Sandwell, 1994). Small scale features such as abyssal hills are not resolved by this technique.

In the recent SRTM15+ bathymetry (Tozer et al., 2019), gravity predicted depths were estimated to have root-mean-square (RMS) uncertainties and maximum error of the order of ±150 m and 1800 m, respectively (based on 50 cruises distributed

globally). With about 22% seafloor coverage by high quality direct measurements, the International Bathymetric Chart of the Southern Ocean v2 (Dorschel et al., 2022, IBCSO) is the state of the art bathymetric chart south of $50°$ S. A cell by cell comparison between IBSCO v2 and SRTM15+ v2.2 showed marked disparities, in particular for water depths between -4000 m and -1000 m with differences reaching up to 1700 m (Dorschel et al., 2022). Most of the bathymetric errors only affect the generation of high-mode IT which are bound to dissipate locally. Note that HYCOM may not resolve these high-mode IT anyway, as discussed in the present section. Still, the less frequent errors of the order of thousands of meters are likely to affect the generation of low-mode IT in the model.

Efforts have been made to improve the accuracy of the M_2 barotropic tides embedded in HYCOM. The simulation used in this study incorporates the framework presented in Ngodock et al. (2016) aiming at minimizing the tidal elevation RMS errors with respect to TPXO, a state of the art data-assimilative tide model (Egbert and Erofeeva, 2002). However, the tidal elevations in TPXO itself also have errors. Both Stammer et al. (2014) and Zaron and Elipot (2021) point at the imperfection of the modeled tidal elevations near Antarctica (using data from GRACE and CryoSat-2, respectively). On the other hand, it is not so much the sea surface height that matters here, but the tidal currents. Zaron and Elipot (2021) used surface drifter observations for evaluating TPXO predicted tidal currents throughout much of the global oceans. Unfortunately, the spatial density of observations is too poor to evaluate the model performance around much of Antarctica.

Lastly, to assess the stratification in HYCOM, we compare the phase speed of a mode-1 gravity wave in the model with the phase speed determined from climatology. The phase speed of a mode-m gravity wave c_m is directly related to the eigenvalue $-1/c_m^2$ obtained by solving the Sturm-Liouville problem Eq. (7). Having already solved Eq. (7) for the climatology, we now solve it for our simulated monthly-mean 3D fields of temperature and salinity (subsampled to $1°$). As for the climatology, the HYCOM data were extended down to the reference bathymetry from GEBCO 2019, wherever deeper, by appending the stored bottom value. We then linearly interpolated the climatological phase speed of a mode-1 gravity wave c_1^{WOA} at $1/4°$ resolution onto the coarser $1°$ HYCOM grid.

In Fig. 14a we plot the climatology to HYCOM phase speed ratio. Most of the visible differences fall in the range of expected interanual variability of less than 10% (Chelton et al., 1998). However, in a few areas around Antarctica there are larger departures of the ratio from unity. Fig. 14b shows the zonal mean of the mode-1 phase speed from Argo and HYCOM as a function of latitude. Equatorward of $\pm 60°$, the two curves agree almost perfectly, and they also visually agree with the results of Chelton et al. (1998) (not shown). Poleward of $60°$ S and $60°$ N, however, differences steadily grow. In the Southern Ocean, the zonal mean climatology to HYCOM phase speed ratio peaks at 1.7 (see Fig. 14c). Typically, weaker stratification (smaller phase speed) results in smaller energy conversion.

~~In terms of wavelength (roughly $\propto 1/c$) the latter ratio is reversed, with HYCOM being biased high in the Southern Ocean, on a zonal mean. The energy conversion from the barotropic tide to the mode-m baroclinic tide mostly occurs where the bathymetry horizontal wavelength is comparable to the mode-m wavelength. Hence, a biased mode-m wavelength in HYCOM results in a biased energy conversion from the barotropic tide to the mode-m IT.~~
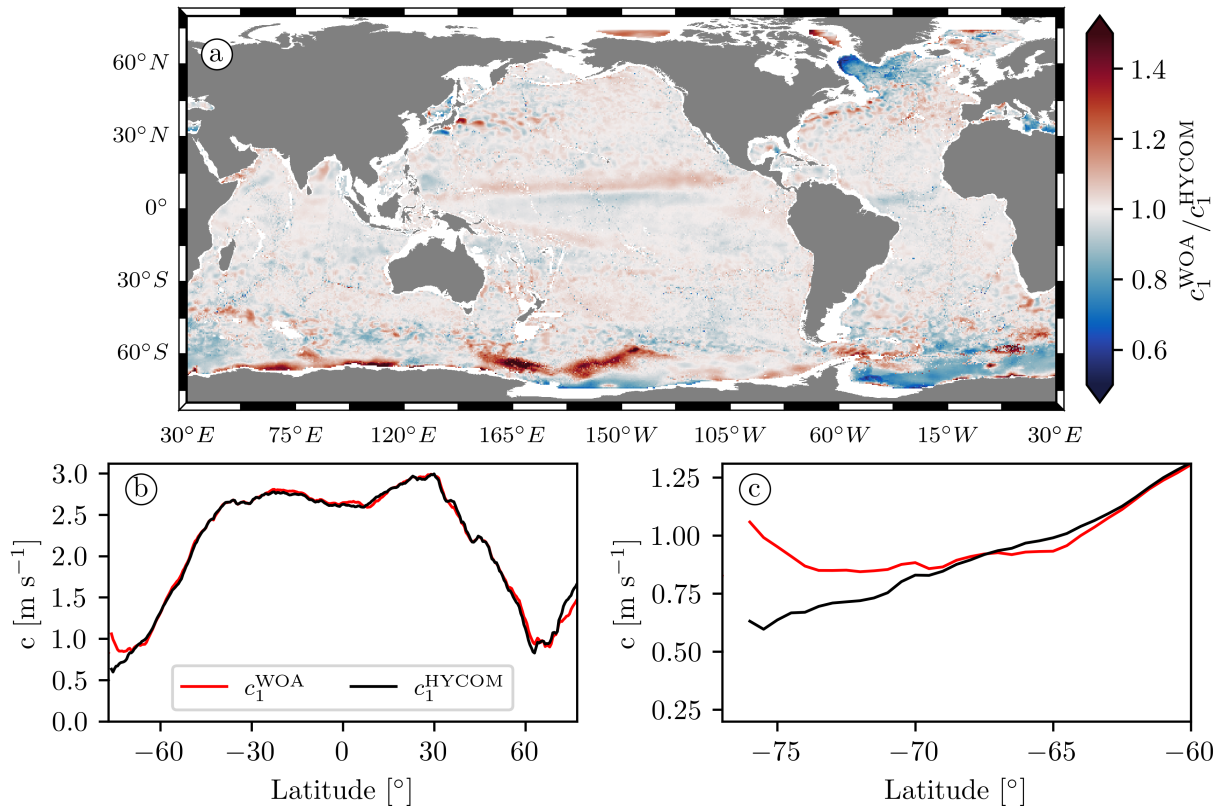
Figure 14. (a) Ratio of the phase speeds of a mode-1 gravity wave computed from the WOA18 and HYCOM stratification profiles. (b) Zonal mean of the phase speed of a mode-1 gravity wave from WOA18 (solid red curve) and HYCOM (solid black curve) as a function of latitude. (c) Same as (b) but zoomed in between $77°$ S and $60°$ S.

5 Conclusions

In this work we compared a 32-day segment of a global run of the HYCOM model, including realistic tidal and atmospheric forcing, with in situ observations of the semidiurnal IT around 1,000 dbar. First, a Lagrangian sampling of the simulation

520 was compared to park phase data from Argo floats to validate the geographical variability of the semidiurnal IT variance in HYCOM (see Sect. 4.1). Then, the Eulerian simulation outputs were directly compared to geographically sparser mooring records, in terms of variance and decorrelation of the IT (see Sect. 4.2).

The main spatial patterns of the simulated IT variance at 1,000 dbar broadly agree with Argo observations, with energy radiating away from low-mode IT generation hotspots (see Fig. 6). Nonetheless, on a global average, the HYCOM data exhibit

525 a smaller semidiurnal IT variance than observed by Argo floats, by a factor ~~0.57~~ 0.58 (see Table 1). This is in line with the results of Ansong et al. (2017) and Luecke et al. (2020), who found HYCOM values to be biased low by a similar factor when comparing the simulated M_2 IT energy flux and temperature variance to historical mooring observations.

While the difference between the model and Argo data appears reasonably homogeneous across most of the world ocean, it steadily increases towards the poles (see Fig. 7). Because of the scarcity of Argo floats available in the ~~northmost~~ northernmost region, we focused on the Southern Ocean. On average, south of $50°$ S, we found that the simulated semidiurnal IT variance is smaller than the variance observed by Argo by a factor ~~0.28~~0.29. North of $50°$ S, this factor is ~~0.68~~ 0.69 (see Table 1).

The mooring data support the above results for the semidiurnal IT variance. Additionally, we found that the decorrelation affecting the semidiurnal IT in HYCOM over a 32-day window is weaker than observed in the mooring records, on average (see Fig. 10 and Table 2). In other words, over timescales shorter than 32 days, the IT in HYCOM are more coherent than in observations. This weaker decorrelation of the IT in HYCOM can be explained by some decorrelating processes being weaker in the model than in the real ocean. It could also be explained by certain decorrelating processes being unusually weak in the May 20 to June 20, 2019 period of outputs used in this work. Depending on the location, the complete decorrelation of the nonstationary IT is not systematically observable in a 32-day duration. Longer time series are thus needed to accurately describe the decorrelation of the IT. Nonetheless, we found that the semidiurnal IT autocovariance in HYCOM actually reaches its stationary limit within approximately 300 h on a global average, i.e. well under our 32 days of simulated data (see Fig. 11). Put together, these results support the conclusions of Buijsman et al. (2020), who found that the stationary (i.e., the longterm coherent) M_2 IT solution from HYCOM is too energetic compared with altimetry. Note that their comparison was based on one-year simulated time series corrected for the duration difference with 17-year long altimetry records.

We also investigated the effects of the Lagrangian sampling inherent to the Argo floats. When comparing autocovariances computed from the HYCOM data sampled in the Lagrangian and Eulerian frameworks, respectively, we found the ~~total~~ IT variance to be unaffected in the mean (see Fig. 5). Moreover, the simulated apparent decorrelation (the decorrelation due to the motion of a Lagrangian particle) agrees very well with the apparent decorrelation experienced by Argo floats, on a global average (see Sect. 4.3). The ~~IT decorrelation~~ (Eulerian) decorrelation of the IT in the simulated data, on the other hand, typically is half as rapid as ~~observed by Argo floats~~the one inferred from Argo observations. This would make the IT decorrelation at least as important as the apparent one over the first few days of time lag. However, the geographical variability of the duration and strength of the IT decorrelation is expected to be large. Limiting the comparison to the latitude range from $30°$ S to $30°$ N, the Argo and HYCOM data lead to an IT decorrelation timescale of about 4.5 and 6 d, respectively (see Table 4), in reasonable agreement with the results of Zaron (2022).

Finally, we discussed the potential sources of bias. We could not think of a particular reason for the IT variance obtained from either the Argo or the mooring data to be biased high, particularly in the Southern Ocean. However, HYCOM is subject to various limitations. First and foremost, the model can only correctly resolve vertical modes up to 5 in most of the global oceans. Approaching the poles, the reduced number of layers further limits the number of resolved modes. While mode-1 IT supposedly accounts for most of the tidal variability at 1,000 dbar on a global average (Geoffroy and Nycander, 2022), in situ instruments also record the contribution from higher modes that can become significant locally. This may explain why the IT variance is larger in the in situ data than in HYCOM, particularly in the Southern Ocean. Insufficient model stratification also seems to be a specific problem in that very region. We have not been able to quantitatively explain the overall smaller IT

variance in HYCOM than in the in situ data over the global ocean. In principle, it could be due to limitations in the bathymetry, barotropic tidal forcing, or the stratification.

Code and data availability. Argo data were obtained from U.S. GDAC (ftp://usgodae.org/pub/outgoing/argo). These data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (https://argo.ucsd.edu, https://www.ocean-ops.org). The Argo Program is part of the Global Ocean Observing System. An Argo Iridium float list is maintained by Stephen C. Riser (http://runt.ocean.washington.edu/argo/heterographs/rollcall.html). The code used to download and process the Argo data is available at https://doi.org/10.57669/geoffroy-2022-argoit-1.0.0 (Geoffroy, 2022a). A global map of the total semidiurnal internal tide variance at 1,000 dbar produced using the latter code is available at https://doi.org/10.17043/geoffroy-2022-argoit-1 (Geoffroy, 2022b). There is no long term availability plan for the HYCOM data used in this work. Climatological data were obtained from the World Ocean Atlas 2018 (https://accession.nodc.noaa.gov/NCEI-WOA18). Bathymetric data were obtained from the GEBCO Compilation Group (https://www.gebco.net/data_and_products/gridded_bathymetry_data/). The Global Multi-Archive Current Meter Database is is not publicly available but can be obtained through request (http://stockage.univ-brest.fr/~scott/GMACMD/gmacmd.html). Netcdf versions of the baroclinic tidal harmonic constants from the High Resolution Empirical Tide model are made available by Edward D. Zaron (https://ingria.ceoas.oregonstate.edu/~zarone/downloads.html).

Appendix A: Bias of the complex demodulate of a sample mean function

We consider the function $R(\tau)$ on a 48-h interval. This is fitted to the function $A\cos(\omega\tau) + B\sin(\omega\tau)$, and we then define

$$C = \sqrt{A^2 + B^2}. \tag{A1}$$

This procedure is a mapping from the space of functions R to the positive number C, i.e. a functional. We can write this symbolically as

$$C = \Phi[R], \tag{A2}$$

where Φ is the functional.

Φ obviously has the property

$$\Phi[aR] = |a|\,\Phi[R], \tag{A3}$$

where a is a constant. Φ also has the property (proof hereafter)

$$\Phi[R_1 + R_2] \leq \Phi[R_1] + \Phi[R_2], \tag{A4}$$

with equality only if $R_1(\tau)$ and $R_2(\tau)$ are proportional, i.e. perfectly correlated with positive correlation.

Let $\{R_i(\tau)\}$ be an infinite series of functions drawn from a stochastic distribution. We have access to only N of these (typically around 10). Define the sample average

$$\overline{R}_N(\tau) = \frac{1}{N} \sum_{i=1}^{N} R_i(\tau). \tag{A5}$$

$\overline{R}_N(\tau)$ is an unbiased estimate of the true average $\overline{R}(\tau)$:

$$\lim_{N \to \infty} \overline{R}_N(\tau) = \overline{R}(\tau). \tag{A6}$$

We also want an estimate of $\Phi\left[\overline{R}\right]$. Denote the true value by C_∞:

$$C_\infty = \Phi\left[\overline{R}\right]. \tag{A7}$$

A sample estimate might be

$$C_N = \Phi\left[\overline{R}_N\right]. \tag{A8}$$

However, (A4) implies that this is biased. For example, define

$$\overline{R}_M(\tau) = \frac{1}{N} \sum_{i=N+1}^{2N} R_i(\tau). \tag{A9}$$

We get

$$C_{2N} = \Phi\left[\overline{R}_{2N}\right] = \Phi\left[\frac{\overline{R}_N + \overline{R}_M}{2}\right] \leq \frac{1}{2}(\Phi\left[\overline{R}_N\right] + \Phi\left[\overline{R}_M\right]), \tag{A10}$$

with equality only if $\overline{R}_N$ and $\overline{R}_M$ are proportional. But $\overline{R}_N$ and $\overline{R}_M$ are stochastic functions drawn from the same distribution, and therefore almost never proportional. On average we also have $\Phi\left[\overline{R}_N\right] = \Phi\left[\overline{R}_M\right]$, hence Eq. (A10) implies

$$C_{2N} < C_N, \tag{A11}$$

so that C_N is a decreasing function of N. Thus, we expect C_N to be larger than C_∞. In other words, the complex demodulate of the sample function $\overline{R}_N(\tau)$, i.e. the fitting defined in Eq. (A2), is a biased high estimate of the envelope of the true function $\overline{R}(\tau)$.

We now show the inequality Eq. (A4). It resembles the triangle inequality for two vectors:

$$|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|. \tag{A12}$$

For 2D vectors $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$, this gives

$$\sqrt{(a_1 + b_1)^2 + (a_2 + b_2)^2} \leq \sqrt{a_1^2 + a_2^2} + \sqrt{b_1^2 + b_2^2}, \tag{A13}$$

which is valid for arbitrary numbers a_1, a_2, b_1, and b_2.

Define a scalar product $\langle f, g \rangle$ between the functions $f(\tau)$ and $g(\tau)$ as the integral of fg over the 48-hour interval:

$$\langle f, g \rangle \equiv \int f(\tau) g(\tau)\, d\tau. \tag{A14}$$

Suppose that the period of $\sin(\omega \tau)$ exactly fits this window (note that this is the case in the present work, where $4 \times 2\pi/\omega_{\mathrm{SD}} \approx 48.8\,\mathrm{h}$). Then $\sin(\omega \tau)$ and $\cos(\omega \tau)$ are orthogonal:

$$\langle \cos(\omega \tau), \sin(\omega \tau) \rangle = 0. \tag{A15}$$

Any function $f(\tau)$ can be expanded in a Fourier series on the interval. A least squares fit of $f(\tau)$ to $A \cos(\omega \tau) + B \sin(\omega \tau)$ is the same thing as projecting $f(\tau)$ on the basis functions $\cos(\omega \tau)$ and $\sin(\omega \tau)$. Thus, the best fit is given by

$$A = \frac{1}{E} \langle f(\tau), \cos(\omega \tau) \rangle, \text{ and } B = \frac{1}{E} \langle f(\tau), \sin(\omega \tau) \rangle, \tag{A16}$$

where $E = \langle \cos(\omega \tau), \cos(\omega \tau) \rangle = \langle \sin(\omega \tau), \sin(\omega \tau) \rangle$. Consider two functions $f(\tau)$ and $g(\tau)$, and denote

$$
\begin{aligned}
a_1 &= \frac{1}{E} \langle f(\tau), \sin(\omega \tau) \rangle, \\
a_2 &= \frac{1}{E} \langle f(\tau), \cos(\omega \tau) \rangle, \\
b_1 &= \frac{1}{E} \langle g(\tau), \sin(\omega \tau) \rangle, \\
b_2 &= \frac{1}{E} \langle g(\tau), \cos(\omega \tau) \rangle.
\end{aligned}
\tag{A17}
$$

We then have

$$\Phi[f] = \sqrt{a_1^2 + a_2^2}, \tag{A18}$$

$$\Phi[g] = \sqrt{b_1^2 + b_2^2}, \tag{A19}$$

$$\Phi[f + g] = \frac{1}{E} \sqrt{\langle f + g, \sin(\omega \tau) \rangle^2 + \langle f + g, \cos(\omega \tau) \rangle^2} = \sqrt{(a_1 + b_1)^2 + (a_2 + b_2)^2}. \tag{A20}$$

Putting Eq. (A18)-(A20) in Eq. (A13) gives Eq. (A4).

Appendix B: Effects of the non-tidal variability on the complex demodulates

Two aspects of the processing originally used in Geoffroy and Nycander (2022) are important in mitigating a potential contamination of the first demodulate by the non-tidal variability (noise). (*i*) The demodulation over a given 48-h window is performed in two iterations. $A \cos(\omega_{\mathrm{SD}} \tau) + B \sin(\omega_{\mathrm{SD}} \tau)$ is first fitted to the signal and the root mean square error (RMSE)

is computed. The fitting is then repeated with a trimmed signal that excludes values outside of $\pm 2\,\mathrm{RMSE}$ from the previously fitted curve. (*ii*) The time series are high-pass filtered prior to computing their sample autocovariance. Suppose the non-tidal variability $\rho(t)$ is a first order autoregressive process (AR1) characterized by zero mean, the timescale τ_ρ, and the white Gaussian noise $\varepsilon_\rho(t)$ with zero mean and variance $\sigma^2_{\varepsilon,\rho}$. The variance and autocovariance of an AR1 process are given by

$$\sigma^2_\rho = \frac{\sigma^2_{\varepsilon,\rho}}{1 - \exp(-2/\tau_\rho)}, \; R_\rho(\tau) = \sigma^2_\rho \exp(-\tau/\tau_\rho), \tag{B1}$$

respectively. For very short τ_ρ, $\rho(t)$ is close to a white noise, i.e. showing a flat power spectrum. As τ_ρ increases, $\rho(t)$ resembles a red noise with increasingly steep exponential decay in spectral space. Therefore, as τ_ρ increases, an increasing fraction of σ^2_ρ is filtered out by the high-pass filter.

We can investigate this further using a simple model, adapted from Geoffroy and Nycander (2022), of a tidal variability on top of a background noise:

$$h(t) = \rho(t) + \sum_i A_i \cos(\omega_i t + \phi(t)). \tag{B2}$$

Here, ω_i and A_i are the angular frequency, and the amplitude of the tidal constituent i, respectively, where $i \in \{\mathrm{M_2, S_2}\}$. $\phi(t)$ and $\rho(t)$ are AR1 processes representing random phase modulations and a non-tidal variability, respectively. These AR1 processes can be defined by their variance and characteristic time. We will vary σ^2_ρ and τ_ρ (the variance and characteristic time of $\rho(t)$, respectively) with the rest of the variables set to realistic values: the tidal variance $\sigma^2_{\mathrm{SD}} = (A^2_{\mathrm{M_2}} + A^2_{\mathrm{S_2}})/2 = 40\ \mathrm{m^2}$, and $\sigma^2_\phi = 1.19\ \mathrm{rad^2}$ and $\tau_\phi = 150\ \mathrm{h}$ (the variance and characteristic time of $\phi(t)$, respectively). Since the tidal variance is fixed, varying σ^2_ρ is the same as varying the signal to noise ratio $\mathrm{S/N} = \sigma^2_{\mathrm{SD}}/\sigma^2_\rho$. For reference, in Appendix C we estimate the global distribution of the signal-to-noise ratio observed by Argo floats (not shown). This distribution is skewed, with median 0.8 and 5th percentile 0.2, approximately. We also estimate the median $\tau_\rho \approx 1\ \mathrm{h}$ from the global collection of floats.

We start by varying τ_ρ for a fixed variance value of $\sigma^2_\rho = 200\ \mathrm{m^2}$, corresponding to the extreme $\mathrm{S/N} = 0.2$. For a given value of τ_ρ, we generate 1000 32-day long synthetic time series $h(t)$ following the model Eq. (B2). The time series are high-pass filtered using a fourth-order Butterworth filter with a cut-off frequency of 0.3 cpd. We then compute a sample autocovariance from each filtered time series and average these to obtain a mean autocovariance. Finally we compute the 48-h demodulates of the mean autocovariance following the complex demodulation method described in Sect. 3.2.

In Fig. B1, we show the mean autocovariance and the corresponding 48-h demodulates for $\tau_\rho = 0.1$, 1, and 12 h. The value of the first demodulate for $\tau_\rho = 0.1$ is essentially the same as if there was no $\rho(t)$ at all, namely $C(\tau = [0, 48]) \approx 34\ \mathrm{m^2}$. For longer τ_ρ, the main difference with the previous curve can be easily visualized from the first few demodulates. In particular, the first demodulates $C(\tau = [0, 48]) \approx 38$ and $31\ \mathrm{m^2}$, corresponding to $\tau_\rho = 1$ and 12 h, are systematically biased high and low, respectively.

We repeat the same experiment for τ_ρ ranging from 0 to 48 h, and focusing on the effects of the noise on the first demodulate. In Fig. B2, we plot the first demodulate as a function of τ_ρ (solid curve). For any τ_ρ, the first demodulate is seen to remain lower than the true $\sigma^2_{\mathrm{SD}} = 40\ \mathrm{m^2}$. Moreover, two regimes can be identified: (*i*) for τ_ρ shorter than 5 h, roughly, the contribution
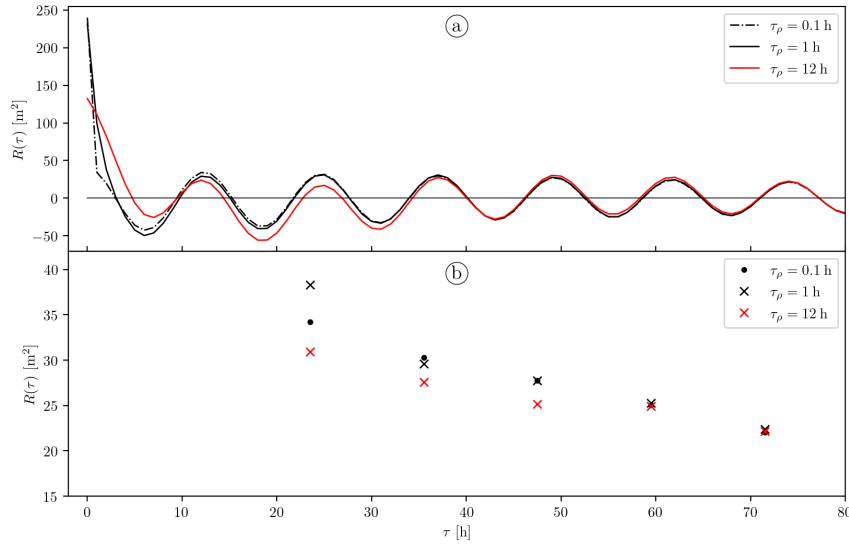
Figure B1. (a) Mean autocovariance computed from 1000 synthetic time series generated following the model Eq. (C8) for $\tau_\rho = 0.1$, 1, and 12 h (dash-dotted black, solid black, and solid red curve, respectively) and a fixed $S/N = 0.2$. (b) Complex demodulates at the semidiurnal frequency of the autocovariance series shown in (a).

of the non-tidal variability to the first demodulate is positive and peaks for $\tau_\rho = 1$ h where $C(\tau = [0, 48]) \approx 38$ m^2, and (*ii*) for longer τ_ρ, this contribution is negative with a minimum $C(\tau = [0, 48]) \approx 30$ m^2 at $\tau_\rho = 11$ h. For a more typical $S/N = 0.8$,

665   i.e. setting $\sigma_\rho^2 = 50$ m^2 in our synthetic time series, we found a similar dependence of the first demodulate on τ_ρ as previously but with a much smaller span in amplitude (dash-dotted curve in Fig. B2). The contribution of the non-tidal variability to the first demodulate again peaks for $\tau_\rho = 1$ h with $C(\tau = [0, 48]) \approx 35$ m^2, and the minimum occurs around $\tau_\rho = 8$ h where $C(\tau = [0, 48]) \approx 33$ m^2.

    Hence, for the typical $S/N = 0.8$ and $\tau_\rho = 1$ observed by Argo floats, we estimate that the non-tidal variability can lead to

670   a first demodulate being biased high by roughly 3%. For the same characteristic timescale but $S/N = 0.2$ (the 5th percentile of the global S/N distribution sampled by Argo floats), this bias can reach 10%. Notwithstanding, even for such extreme S/N value, the first demodulate is seen to remain a conservative estimate of the IT variance. The effects of the non-tidal variability on the IT variance estimate as computed in Geoffroy and Nycander (2022) thus do not put their results into question.

    It is not straightforward to understand why the contribution of the non-tidal variability to the value of the first demodulate

675   can be negative. Because of the way we defined $\rho(t)$, i.e. an AR1 process with positive parameter, its autocovariance should remain positive. Oscillations of $R_\rho(\tau)$ are actually a consequence of the high-pass filter applied to the time series. We show this in Fig. B3 where we plot the theoretical autocovariance of $\rho(t)$ with $\tau_\rho = 12$ h and $\sigma_\rho^2 = 200$ m^2 (solid black curve), and the sample mean autocovariance from 1000 non filtered and 1000 high-pass filtered synthetic time series (solid and dash-dotted red curve, respectively) constructed using the same parameters.
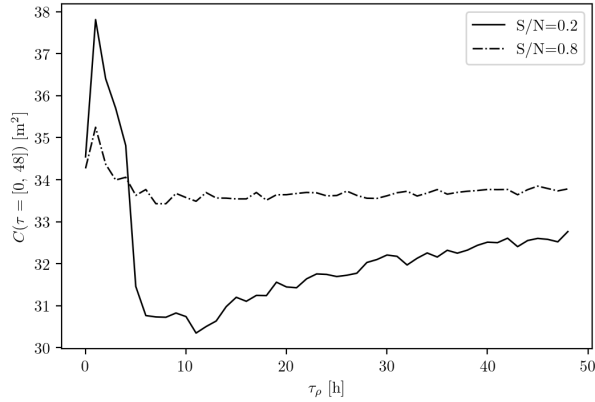
Figure B2. First 48-h demodulate of the mean autocovariance computed from 1000 synthetic time series generated following the model Eq. (B2) as a function of τ_ρ for fixed S/N $= 0.2$ (solid curve) and 0.8 (dash-dotted curve).
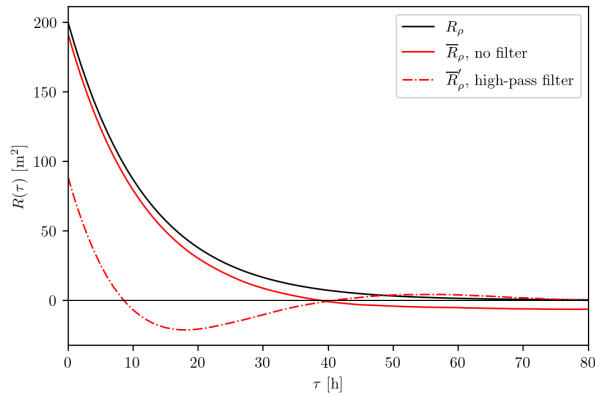
Figure B3. Theoretical autocovariance of the AR1 process $\rho(t)$ with $\tau_\rho = 12$ h and $\sigma_\rho^2 = 200$ m^2 (solid black curve), and sample mean autocovariance computed from 1000 non filtered and 1000 high-pass filtered synthetic time series of $\rho(t)$ (solid and dash-dotted red curve, respectively) constructed using the same parameters.

680   As mentioned in Sect. 3.2, we expect the stochastic background noise affecting the simulated data to be different from the one recorded by the in situ instruments. Therefore, to prevent any systematic bias in the comparisons presented in this work, we consistently correct the sample autocovariance by subtracting an estimate of the autocovariance of the non-tidal variability before performing the complex demodulation (see Appendix C). We illustrate the effects of this correction in Fig. B4. Here, we proceed in the same way as for Fig. B2, but this time varying S/N (from 4×10^{-2} to 4) for fixed $\tau_\rho = 1$ and 12 h. While the

685   first demodulates of the non corrected mean autocovariances diverge for small S/N values (solid curves), the first demodulates of the noise corrected autocovariances are virtually constant (dash-dotted curves).
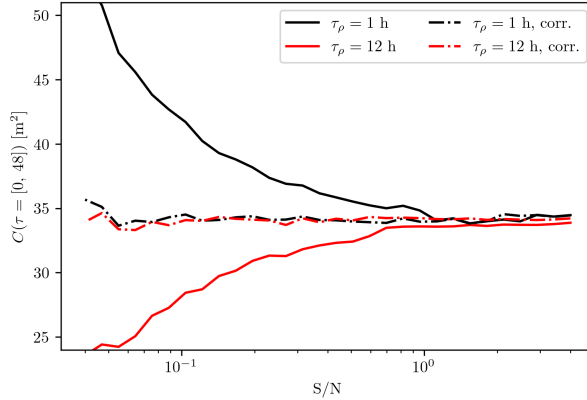
Figure B4. First 48-h demodulate of the mean autocovariance computed from 1000 synthetic time series generated following the model Eq. (B2) as a function of S/N for fixed $\tau_\rho = 1$ (solid black curve) and 12 h (solid red curve). The first demodulates of the corresponding noise-corrected autocovariances are shown as the black and red dash-dotted curves, for $\tau_\rho = 1$ and 12 h, respectively.

Appendix C: Estimating the autocovariance of the non-tidal variability

In the present appendix, we derive a model for the autocovariance of a tidal variability on top of a high-pass filtered stochastic noise. In Appendix B we saw that the autocovariance of the non-tidal variability is affected by the high-pass filter we apply on the original time series (see Fig. B3). This consequence of filtering the time series was not taken into account by Geoffroy and Nycander (2022) in the model of the background noise they used.

Assume that the non-tidal variability $\rho(t)$ is a first order autoregressive process (AR1) characterized by zero mean, the timescale τ_ρ, and the white Gaussian noise $\varepsilon_\rho(t)$ with zero mean and variance $\sigma^2_{\varepsilon,\rho}$. In the time domain, the filtered $\rho'(t)$, can be modeled as the convolution of the unfiltered $\rho(t)$ with the impulse response of the filter $h(t)$. We are, however, interested in the autocovariance of $\rho'(t)$, $R'_\rho(\tau)$, which is closely related to the power spectrum (one is the Fourier transform of the other). Luckily, it is much simpler to perform this operation in the frequency domain, where the power spectrum at the output of a linear filter, $\mathcal{P}'_\rho(z)$, is related to the power spectrum of the input stochastic process, $\mathcal{P}_\rho(z)$, by

$$\mathcal{P}'_\rho(z) = |H(z)|^2 \mathcal{P}_\rho(z). \tag{C1}$$

Here, $H(z)$ is the system or transfer function of the filter, and the z-domain is limited to the unit circle $z = \exp(j\omega\Delta t)$, i.e. the z-transform is equivalent to the discrete Fourier transform, where ω is the angular frequency, and Δt is the constant sampling interval (Kay and Marple, 1981). Moreover, the power spectrum of an AR1 process is a standard result:

$$\mathcal{P}_\rho(\omega) = \frac{\sigma^2_{\varepsilon,\rho}\Delta t}{1 + \exp(-2/\tau_\rho) - 2\exp(-1/\tau_\rho)\cos(\omega\Delta t)}. \tag{C2}$$

The Butterworth filter is linear in amplitude but not in phase. As a workaround, we applied a second order filter twice, once forward and once backward, to recover a 4$^\text{th}$ order filter with zero phase. The transfer function of a second order high-pass

Butterworth filter in the z-domain, $H_{\mathrm{B2}}(z)$, can be obtained from its Laplace transform:

$$H_{\mathrm{B2}}(s) = \frac{1}{(\omega_c/s)^2 + \sqrt{2}(\omega_c/s) + 1}, \tag{C3}$$

where ω_c is the cut-off angular frequency (Schlichthärle, 2000). We use the bilinear transform, namely setting

$$s = \frac{1}{\Delta t}\ln(z) \approx \frac{2}{\Delta t}\frac{z-1}{z+1} \tag{C4}$$

in Eq. (C3), to recover the expression

$$H_{\mathrm{B2}}(z) = \frac{4 - 8z^{-1} + 4z^{-2}}{(\omega_c^2\Delta t^2 + 2\sqrt{2}\omega_c\Delta t + 4) + (2\omega_c^2\Delta t^2 - 8)z^{-1} + (\omega_c^2\Delta t^2 - 2\sqrt{2}\omega_c\Delta t + 4)z^{-2}}. \tag{C5}$$

Hence, we can evaluate Eq. (C1) as

$$\mathcal{P}'_\rho(\omega) = \left(\left| H_{\mathrm{B2}}\big(z = \exp(j\omega\Delta t)\big) \right|^2 \right)^2 \mathcal{P}_\rho(\omega). \tag{C6}$$

Since ρ' is real-valued, the autocovariance $R'_\rho(\tau)$ is given by the inverse discrete Fourier cosine transform of its power spectrum $\mathcal{P}'_\rho(\omega)$ (computed numerically).

As mentioned in Sect. 3.2, to limit any systematic bias we consistently estimate and subtract the autocovariance of the non-tidal variability from the sample autocovariance before performing the complex demodulation. We obtain such an estimate from the least squares fit of the model adapted from Geoffroy and Nycander (2022):

$$R_m(\tau) = R'_{\rho_f}(\tau) + R'_{\rho_s}(\tau) + \sum_i \frac{A_i}{2}\cos(\omega_i\tau)\exp(-\sigma^2_{\phi,\,i} + R_{\phi,\,i}(\tau)), \tag{C7}$$

corresponding to the autocovariance of a tidal variability affected by a background noise:

$$m(t) = \rho'_f(t) + \rho'_s(t) + \sum_i A_i\cos(\omega_i t + \phi_i(t)). \tag{C8}$$

Here, ω_i and A_i are the angular frequency, and the amplitude of the tidal constituent i, respectively, where $i \in \{\mathrm{M_2, S_2, K_1, O_1}\}$. $\phi_i(t)$ is an AR1 process representing the random phase modulations affecting the constituent i. For the sake of simplicity, we assume that the decorrelating processes act in the same manner on neighboring tidal frequencies, that is, $\phi_{\mathrm{M_2}}(t) = \phi_{\mathrm{S_2}}(t)$ and $\phi_{\mathrm{K_1}}(t) = \phi_{\mathrm{O_1}}(t)$. Note that the high-pass filter used in this work was designed so as not to affect the tidal signal. $\rho'_f(t)$ and $\rho'_s(t)$ are high-pass filtered AR1 processes accounting for a fast and a slow non-tidal variability, respectively. All the AR1 processes are characterized by zero mean, the time scale τ_X, and the white Gaussian noise $\varepsilon_X(t)$ with zero mean and variance $\sigma^2_{\varepsilon,X}$, and $X \in \{\phi, \rho_f, \rho_s\}$. The variance and autocovariance of an AR1 process are

$$\sigma^2_X = \frac{\sigma^2_{\varepsilon,\,X}}{1 - \exp(-2/\tau_X)}, \quad R_X(\tau) = \sigma^2_X\exp(-\tau/\tau_X), \tag{C9}$$

respectively. The model Eq. (C7) does not take into account the effects of the drift of the floats or Lagrangian particles.

730   We fit the full model Eq. (C7), containing 12 parameters, to the sample autocovariance by nonlinear least squares, with weights inversely proportional to the corresponding SEM, and imposed bounds on the parameters (since we only fit positive timescales and variances). In particular, the timescales of the fast and slow non-tidal variability are restricted to the $[0, 5]$ and $[5, 48]$ h range, respectively. The autocovariance of the non-tidal variability is then computed as

$$R'_\rho(\tau) = R'_{\rho_f}(\tau) + R'_{\rho_s}(\tau) \tag{C10}$$

735 using the fitted parameters. Lastly, $R'_\rho(\tau)$ is subtracted from the sample autocovariance to correct for the effects of the filtered background noise.

    The inclusion of a tidal variability in the model Eq. (C7) primarily aims at limiting the projection of the observed tidal variability onto our model for the background noise when fitting. Nonetheless, the fitted parameters can be used to compute an estimate of the semidiurnal IT variance $\sigma^2_{\mathrm{SD,LS}} = (A^2_{\mathrm{M_2}} + A^2_{\mathrm{S_2}})/2$, and further to estimate the signal-to-noise ratio characterizing

740 the filtered Argo records as $\mathrm{S/N} = \sigma^2_{\mathrm{SD,LS}}/R'_\rho(0)$. In Fig. C1 we show maps of $\sigma^2_{\mathrm{SD,LS}}$, $R'_\rho(0)$, and S/N computed from the same collection of local mean autocovariances as used in Fig. 6c. Note that $\sigma^2_{\mathrm{SD,LS}}$ and $R'_\rho(0)$ are not correlated ($r^2 = 0.13$ and 0.009 in log-log and linear domain, respectively). The least squares fitted $\sigma^2_{\mathrm{SD,LS}}$ appear reasonable when compared with the first demodulates presented in Fig. 6c. The coefficient of determination $r^2 = 0.53$ and 0.29 in log-log and linear domain, respectively, indicates a good agreement between the two estimates. Moreover, the global mean and median of their

745 ratio are 0.8 and 0.7, respectively, accounting for the decorrelation of the IT taking place in the first 48 h. For such local mean autocovariances, however, the uncertainties are simply to high to reliably estimate the parameters of the decorrelating process $\phi_i(t)$. Finally, we chose to perform the comparisons presented in this work in terms of the first demodulate, for it is a conservative and more robust estimate of the IT variance.

Author contributions. Gaspard Geoffroy: Conceptualization, Formal analysis, Project administration, Software, Visualization, Writing - orig-
750 inal draft, Writing - review & editing. Jonas Nycander: Conceptualization, Funding acquisition, Writing - original draft, Writing - review & editing. Maarten C. Buijsman: Project administration, Writing - review & editing. Jay F. Shriver: Formal analysis, Resources, Software. Brian K. Arbic: Writing - review & editing.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. GG and JN are supported by grant number 2017-04623 from the Swedish Research Council. The computations and data
755 handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre (NSC) partially funded by the Swedish Research Council through grant agreement no. 2018-05973. MCB and BKA are supported by the Office of Naval Research (ONR) grant numbers N00014-19-1-2704 and N00014-19-1-2712, respectively, which both fall under the
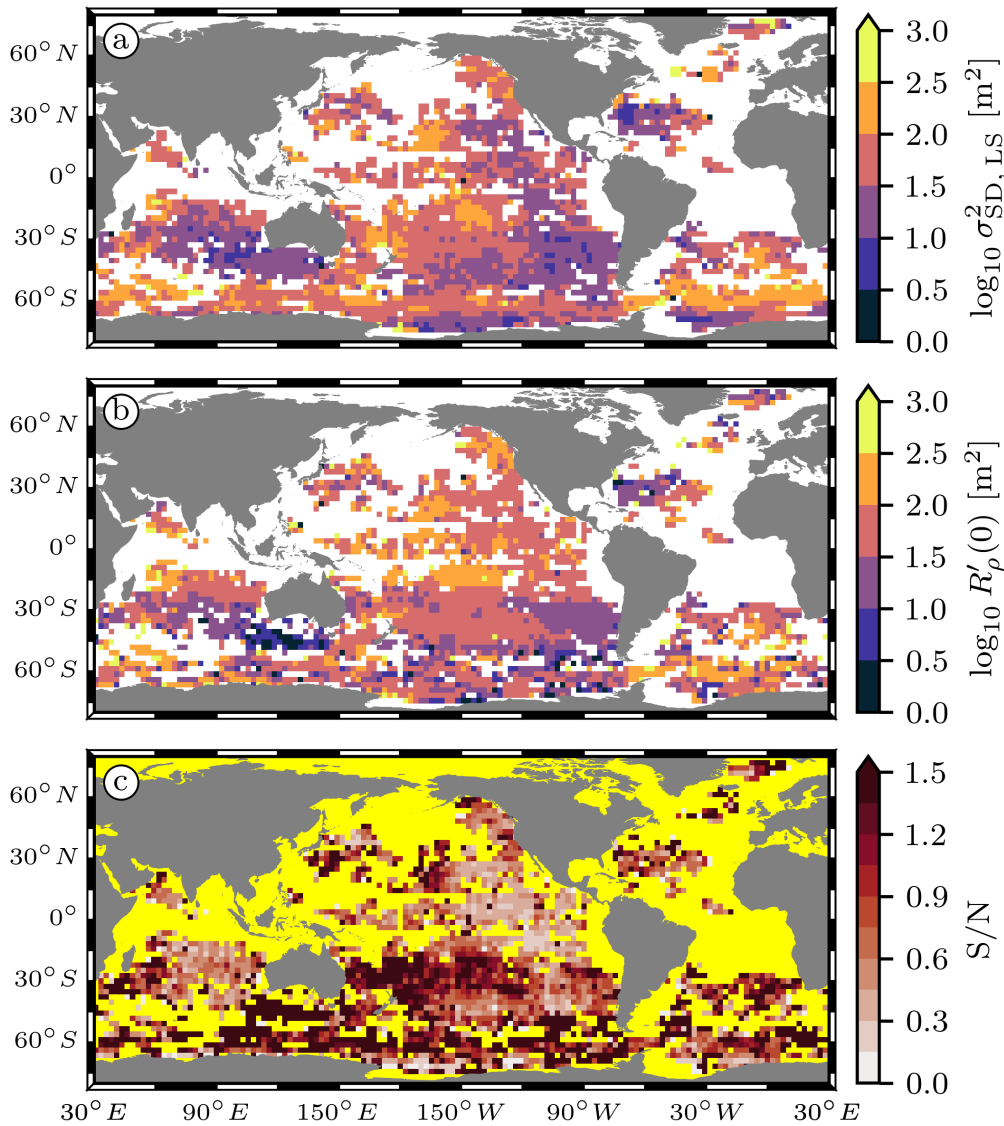
Figure C1. (a) Atlas of the semidiurnal IT variance at 1,000 dbar computed from the fitting of the model Eq. (C7) to the same local mean autocovariance series as used in Fig. 6c. (b) Non-tidal variance from the same fitting as in (a). (c) Signal-to-noise ratio computed as the ratio of (a) over (b). The ocean mask is colored in yellow for readability.

project "Modeling, characterizing, and predicting effects of internal gravity waves on acoustic propagation on basin to global scales". JFS was supported by ONR grant number N0001422WX01919, which falls under the project " Diagnosis and validation of the time and spatial variability of remotely generated internal waves in global ocean simulations". The authors would like to thank ~~the anonymous reviewers who~~ Shane Elipot and one additional anonymous reviewer who greatly helped improve this manuscript.

760

References

Alford, M. H. and Zhao, Z.: Global patterns of low-mode internal-wave propagation. Part I: Energy and energy flux, Journal of Physical Oceanography, 37, 1829 – 1848, https://doi.org/10.1175/JPO3085.1, 2007.

Ansong, J. K., Arbic, B. K., Alford, M. H., Buijsman, M. C., Shriver, J. F., Zhao, Z., Richman, J. G., Simmons, H. L., Timko, P. G., Wallcraft, A. J., and Zamudio, L.: Semidiurnal internal tide energy fluxes and their variability in a global ocean model and moored observations, Journal of Geophysical Research: Oceans, 122, 1882–1900, https://doi.org/https://doi.org/10.1002/2016JC012184, 2017.

Arbic, B. K.: Incorporating tides and internal gravity waves within global ocean general circulation models: A review, Progress in Oceanography, 206, 102 824, https://doi.org/https://doi.org/10.1016/j.pocean.2022.102824, 2022.

Arbic, B. K., Wallcraft, A. J., and Metzger, E. J.: Concurrent simulation of the eddying general circulation and tides in a global ocean model, Ocean Modelling, 32, 175–187, https://doi.org/https://doi.org/10.1016/j.ocemod.2010.01.007, the magic of modelling: A special volume commemorating the contributions of Peter D. Killworth – Part 2, 2010.

Arbic, B. K., Elipot, S., Brasch, J. M., Menemenlis, D., Ponte, A. L., Shriver, J. F., Yu, X., Zaron, E. D., Alford, M. H., Buijsman, M. C., Abernathey, R., Garcia, D., Guan, L., Martin, P. E., and Nelson, A. D.: Frequency dependence of near-surface oceanic kinetic energy from drifter observations and global high-resolution models, https://doi.org/10.48550/ARXIV.2202.08877, 2022.

Argo: https://doi.org/10.17882/42182, Argo float data and metadata from Global Data Assembly Centre (Argo GDAC), 2000.

Boyer, T. P., Garcia, H. E., Locarnini, Ricardo A.and Zweng, M. M., Mishonov, A. V., Reagan, J. R., Weathers, K. A., Baranova, O. K., Seidov, D., and Smolyar, I. V.: World Ocean Atlas 2018. Temperature, and salinity, Dataset, https://accession.nodc.noaa.gov/NCEI-WOA18, Accessed 11-10-2021, 2018.

Buijsman, M. C., Arbic, B. K., Richman, J. G., Shriver, J. F., Wallcraft, A. J., and Zamudio, L.: Semidiurnal internal tide incoherence in the equatorial Pacific, Journal of Geophysical Research: Oceans, 122, 5286–5305, https://doi.org/10.1002/2016JC012590, 2017.

Buijsman, M. C., Stephenson, G. R., Ansong, J. K., Arbic, B. K., Green, J. M., Richman, J. G., Shriver, J. F., Vic, C., Wallcraft, A. J., and Zhao, Z.: On the interplay between horizontal resolution and wave drag and their effect on tidal baroclinic mode waves in realistic global ocean simulations, Ocean Modelling, 152, 101 656, https://doi.org/https://doi.org/10.1016/j.ocemod.2020.101656, 2020.

Caspar-Cohen, Z., Ponte, A., Lahaye, N., Carton, X., Yu, X., and Gentil, S. L.: Characterization of internal tide incoherence: Eulerian versus Lagrangian perspectives, Journal of Physical Oceanography, 52, 1245 – 1259, https://doi.org/10.1175/JPO-D-21-0088.1, 2022.

Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Wallcraft, A. J., Metzger, E. J., Blanton, B. O., Lozano, C., Rao, D. B., and Hogan, P. J.: Generalized vertical coordinates for eddy-resolving global and coastal ocean forecasts, Oceanography, 19, 20, 2006.

Chelton, D. B., deSzoeke, R. A., Schlax, M. G., El Naggar, K., and Siwertz, N.: Geographical variability of the first baroclinic Rossby radius of deformation, Journal of Physical Oceanography, 28, 433–460, https://doi.org/10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2, 1998.

Cherniawsky, J. Y., Foreman, M. G. G., Crawford, W. R., and Henry, R. F.: Ocean tides from TOPEX/Poseidon sea level data, Journal of Atmospheric and Oceanic Technology, 18, 649 – 664, https://doi.org/10.1175/1520-0426(2001)018<0649:OTFTPS>2.0.CO;2, 2001.

de Lavergne, C., Falahat, S., Madec, G., Roquet, F., Nycander, J., and Vic, C.: Toward global maps of internal tide energy sinks, Ocean Modelling, 137, 52–75, https://doi.org/https://doi.org/10.1016/j.ocemod.2019.03.010, 2019.

de Lavergne, C., Vic, C., Madec, G., Roquet, F., Waterhouse, A. F., Whalen, C. B., Cuypers, Y., Bouruet-Aubertot, P., Ferron, B., and Hibiya, T.: A parameterization of local and remote tidal mixing, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 065, https://doi.org/10.1029/2020MS002065, 2020.

Dorschel, B., Hehemann, L., Viquerat, S., Warnke, F., Dreutter, S., Tenberge, Y. S., Accettella, D., An, L., Barrios, F.,
Bazhenova, E., et al.: The International Bathymetric Chart of the Southern Ocean version 2, Scientific Data, 9, 275,
https://doi.org/https://doi.org/10.1038/s41597-022-01366-7, 2022.

Dushaw, B. D., Worcester, P. F., and Dzieciuch, M. A.: On the predictability of mode-1 internal tides, Deep Sea Research Part I: Oceano-
graphic Research Papers, 58, 677–698, https://doi.org/10.1016/j.dsr.2011.04.002, 2011.

Egbert, G. D. and Erofeeva, S. Y.: Efficient inverse modeling of barotropic ocean tides, Journal of Atmospheric and Oceanic Technology, 19,
183 – 204, https://doi.org/10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2, 2002.

Egbert, G. D. and Ray, R. D.: Tidal prediction, Journal of Marine Research, 75, 189–237, https://doi.org/10.1357/002224017821836761,
2017.

GEBCO: Dataset, https://doi.org/10.5285/836f016a-33be-6ddc-e053-6c86abc0788e, GEBCO 2019 Grid, 2019.

Geoffroy, G.: Scripts for mapping the total semidiurnal internal tide variance at 1,000 dbar using Argo data, https://doi.org/10.57669/
geoffroy-2022-argoit-1.0.0, 2022a.

Geoffroy, G.: Global map of the total semidiurnal internal tide variance at 1,000 dbar from Argo data, Dataset, https://doi.org/10.17043/
geoffroy-2022-argoit-1, 2022b.

Geoffroy, G. and Nycander, J.: Global mapping of the nonstationary semidiurnal internal tide using Argo data, Journal of Geophysical
Research: Oceans, 127, e2021JC018 283, https://doi.org/https://doi.org/10.1029/2021JC018283, 2022.

Gill, A. E.: Atmosphere-ocean dynamics, Elsevier, academic press edn., 1982.

Hennon, T. D., Riser, S. C., and Alford, M. H.: Observations of internal gravity waves by Argo floats, Journal of Physical Oceanography, 44,
2370–2386, https://doi.org/10.1175/JPO-D-13-0222.1, 2014.

Kay, S. and Marple, S.: Spectrum analysis—A modern perspective, Proceedings of the IEEE, 69, 1380–1419,
https://doi.org/10.1109/PROC.1981.12184, 1981.

Luecke, C. A., Arbic, B. K., Richman, J. G., Shriver, J. F., Alford, M. H., Ansong, J. K., Bassette, S. L., Buijsman, M. C., Menemenlis, D.,
Scott, R. B., Timko, P. G., Voet, G., Wallcraft, A. J., and Zamudio, L.: Statistical comparisons of temperature variance and kinetic energy
in global ocean models and observations: Results from mesoscale to internal wave frequencies, Journal of Geophysical Research: Oceans,
125, e2019JC015 306, https://doi.org/https://doi.org/10.1029/2019JC015306, 2020.

MacKinnon, J. A., Zhao, Z., Whalen, C. B., Waterhouse, A. F., Trossman, D. S., Sun, O. M., Laurent, L. C. S., Simmons, H. L., Polzin,
K., Pinkel, R., Pickering, A., Norton, N. J., Nash, J. D., Musgrave, R., Merchant, L. M., Melet, A. V., Mater, B., Legg, S., Large, W. G.,
Kunze, E., Klymak, J. M., Jochum, M., Jayne, S. R., Hallberg, R. W., Griffies, S. M., Diggs, S., Danabasoglu, G., Chassignet, E. P.,
Buijsman, M. C., Bryan, F. O., Briegleb, B. P., Barna, A., Arbic, B. K., Ansong, J. K., and Alford, M. H.: Climate Process Team on
internal wave–driven ocean mixing, Bulletin of the American Meteorological Society, 98, 2429 – 2454, https://doi.org/10.1175/BAMS-
D-16-0030.1, 2017.

Melet, A., Legg, S., and Hallberg, R.: Climatic impacts of parameterized local and remote tidal mixing, Journal of Climate, 29, 3473 – 3500,
https://doi.org/10.1175/JCLI-D-15-0153.1, 2016.

Melet, A., Hallberg, R., and Marshall, D. P.: Chapter 2 - The role of ocean mixing in the climate system, in: Ocean Mixing, edited by
Meredith, M. and Naveira Garabato, A., pp. 5–34, Elsevier, https://doi.org/10.1016/B978-0-12-821512-8.00009-8, 2022.

Nelson, A. D., Arbic, B. K., Zaron, E. D., Savage, A. C., Richman, J. G., Buijsman, M. C., and Shriver, J. F.: Toward realistic non-
stationarity of semidiurnal baroclinic tides in a hydrodynamic model, Journal of Geophysical Research: Oceans, 124, 6632–6642,
https://doi.org/10.1029/2018JC014737, 2019.

Ngodock, H. E., Souopgui, I., Wallcraft, A. J., Richman, J. G., Shriver, J. F., and Arbic, B. K.: On improving the accuracy of the M2 barotropic tides embedded in a high-resolution global ocean circulation model, Ocean Modelling, 97, 16–26, https://doi.org/https://doi.org/10.1016/j.ocemod.2015.10.011, 2016.

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

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

Schlichthärle, D.: Digital filters: basics and design, Springer Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-04170-3, 2000.

845 Scott, R. B., Goff, J. A., Naveira Garabato, A. C., and Nurser, A. J. G.: Global rate and spectral characteristics of internal gravity wave generation by geostrophic flow over topography, Journal of Geophysical Research: Oceans, 116, https://doi.org/10.1029/2011JC007005, 2011.

Shriver, J. F., Arbic, B. K., Richman, J. G., Ray, R. D., Metzger, E. J., Wallcraft, A. J., and Timko, P. G.: An evaluation of the barotropic and internal tides in a high-resolution global ocean circulation model, Journal of Geophysical Research: Oceans, 117,
850   https://doi.org/https://doi.org/10.1029/2012JC008170, 2012.

Shriver, J. F., Richman, J. G., and Arbic, B. K.: How stationary are the internal tides in a high-resolution global ocean circulation model?, Journal of Geophysical Research: Oceans, 119, 2769–2787, https://doi.org/10.1002/2013JC009423, 2014.

Smith, W. H. F. and Sandwell, D. T.: Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry, Journal of Geophysical Research: Solid Earth, 99, 21 803–21 824, https://doi.org/https://doi.org/10.1029/94JB00988, 1994.

855 Stammer, D., Ray, R. D., Andersen, O. B., Arbic, B. K., Bosch, W., Carrère, L., Cheng, Y., Chinn, D. S., Dushaw, B. D., Egbert, G. D., Erofeeva, S. Y., Fok, H. S., Green, J. A. M., Griffiths, S., King, M. A., Lapin, V., Lemoine, F. G., Luthcke, S. B., Lyard, F., Morison, J., Müller, M., Padman, L., Richman, J. G., Shriver, J. F., Shum, C. K., Taguchi, E., and Yi, Y.: Accuracy assessment of global barotropic ocean tide models, Reviews of Geophysics, 52, 243–282, https://doi.org/https://doi.org/10.1002/2014RG000450, 2014.

Tozer, B., Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., and Wessel, P.: Global bathymetry and topography at 15 arc sec:
860   SRTM15+, Earth and Space Science, 6, 1847–1864, https://doi.org/https://doi.org/10.1029/2019EA000658, 2019.

Van Sebille, E., Kehl, C., Lange, M., Delandmeter, P., and contributors: Parcels, https://doi.org/10.5281/zenodo.7035503, 2021.

Zaron, E. D.: Mapping the nonstationary internal tide with satellite altimetry, Journal of Geophysical Research: Oceans, 122, 539–554, https://doi.org/10.1002/2016JC012487, 2017.

Zaron, E. D.: Baroclinic tidal sea level from exact-repeat mission altimetry, Journal of Physical Oceanography, 49, 193–210,
865   https://doi.org/10.1175/JPO-D-18-0127.1, 2019.

Zaron, E. D.: Baroclinic tidal cusps from satellite altimetry, Journal of Physical Oceanography, 52, 3123 – 3137, https://doi.org/10.1175/JPO-D-21-0155.1, 2022.

Zaron, E. D. and Egbert, G. D.: Time-variable refraction of the internal tide at the Hawaiian Ridge, Journal of Physical Oceanography, 44, 538 – 557, https://doi.org/10.1175/JPO-D-12-0238.1, 2014.

870 Zaron, E. D. and Elipot, S.: An assessment of global ocean barotropic tide models using geodetic mission altimetry and surface drifters, Journal of Physical Oceanography, 51, 63 – 82, https://doi.org/10.1175/JPO-D-20-0089.1, 2021.

Zhao, Z.: Mapping internal tides from satellite altimetry without blind directions, Journal of Geophysical Research: Oceans, 124, 8605–8625, https://doi.org/10.1029/2019JC015507, 2019.

Zhao, Z., Alford, M. H., Girton, J. B., Rainville, L., and Simmons, H. L.: Global observations of open-ocean mode-1 M2 internal tides,

875   Journal of Physical Oceanography, 46, 1657–1684, https://doi.org/10.1175/JPO-D-15-0105.1, 2016.