Geophysical Journal International

Geophys. J. Int. (2024) **239,** 136–154
Advance Access publication 2024 July 16
GJI Seismology

https://doi.org/10.1093/gji/ggae238

Waveform modelling of hydroacoustic teleseismic earthquake records from autonomous MERMAID floats

Sirawich Pipatprathanporn and Frederik J. Simons

Department of Geosciences, Princeton University, Princeton, NJ 08544, *USA. E-mail: sirawich@princeton.edu*

Accepted 2024 July 2. Received 2024 May 23; in original form 2023 December 8

SUMMARY

We present a computational technique to model hydroacoustic waveforms from teleseismic earthquakes recorded by mid-column MERMAID floats deployed in the Pacific, taking into consideration bathymetric effects that modify seismo-acoustic conversions at the ocean bottom and acoustic wave propagation in the ocean layer, including reverberations. Our approach couples axisymmetric spectral-element simulations performed for moment-tensor earthquakes in a 1-D solid Earth to a 2-D Cartesian fluid–solid coupled spectral-element simulation that captures the conversion from displacement to acoustic pressure at an ocean-bottom interface with accurate bathymetry. We applied our workflow to 1129 seismograms for 682 earthquakes from 16 MERMAIDs (short for Mobile Earthquake Recording in Marine Areas by Independent Divers) owned by Princeton University that were deployed in the Southern Pacific as part of the South Pacific Plume Imaging and Modeling (SPPIM) project. We compare the modelled synthetic waveforms to the observed records in individually selected frequency bands aimed at reducing local noise levels while maximizing earthquake-generated signal content. The modelled waveforms match the observations very well, with a median correlation coefficient of 0.72, and some as high as 0.95. We compare our correlation-based traveltime measurements to measurements made on the same data set determined by automated arrival-time picking and ray- traced traveltime predictions, with the aim of opening up the use of MERMAID records for global seismic tomography via full-waveform inversion.

Key words: Time-series analysis; Infrasound; Computational seismology; Seismic instruments; Wave propagation.

1 INTRODUCTION

Oceans cover two thirds of Earth's surface, where seismic data, crucial for imaging mantle structure (Romanowicz 2003, 2008), remain scarce due to limitations of accessibility, logistics, technical difficulty, and cost (Hammond *et al.* 2019). Mid-column, freely floating hydrophones aboard MERMAID, short for Mobile Earthquake Recording in Marine Areas by Independent Divers, are among the alternatives touted to hold potential for improving seismic coverage over the oceans (Simons *et al.* 2006; Hello *et al.* 2011; Yu *et al.* 2023), along with ocean bottom seismometers (Collins *et al.* 2001; Stephen *et al.* 2003; Suetsugu & Shiobara 2014), cabled hydrophone arrays (Slack *et al.* 1999; Bohnenstiehl *et al.* 2002; Dziak *et al.* 2004), and distributed acoustic sensing (Marra *et al.* 2018; Williams *et al.* 2019; Sladen *et al.* 2019).

MERMAID was designed to detect, identify and report, in near real-time, first-arriving compressional *P* waves from teleseismic earthquakes (Simons *et al.* 2009; Sukhovich *et al.* 2011), autonomously transmitting brief segments of hydroacoustic seismograms suitable for improving global tomographic models (Simon *et al.* 2022; Nolet *et al.* 2024). It is a low-cost, easily deployed, long-lived and versatile instrument (Simons *et al.* 2021). Simon *et al.* (2021) and Pipatprathanporn & Simons (2022) have shown that MERMAID records provide much more information than the ability to make a traditional first-arrival 'pick': the sensor records all manner of signal and noise, including ship traffic, marine mammals, submarine volcanic eruptions and infrasonic ocean ambient noise which is a gauge for surface environmental conditions.

MERMAID seismograms are being and have been (Nolet *et al.* 2019) used for traveltime tomography. Modelling the entire finite-frequency waveform, to move beyond *P*-wave arrival-time picking and structure mapping along the ray-theoretical fastest ray path (Nolet & Dahlen 2000; Rawlinson *et al.* 2014), has remained elusive. Yet incorporating full-physics effects holds the key to imaging geodynamically important regions of strong heterogeneity (e.g. Zhang *et al.* 2023). Applying modern full-waveform inversion techniques (e.g. Tarantola 1984; Tromp 2020; Lei *et al.* 2020) to hydroacoustic seismograms requires simulating *seismic* wave propagation in a 3-D globe with a heterogeneous ocean overlying variable bathymetry in which *acoustic* waves propagate (e.g. Lecoulant *et al.* 2019;

Fernando *et al.* 2020). At the frequencies 0.1–10 Hz, where MERMAID's instrument response is flat and its signal-to-noise ratio (SNR) high (Simon *et al.* 2022), carrying out such simulations remains far too computationally expensive.

Hybrid methods have been developed to couple low- and high-complexity analytical or numerical methods between and across low- and high-complexity solution domains (for a review, see Lyu *et al.* 2022). These methods successfully bridge model classes (radially symmetric versus laterally heterogeneous models, smooth compared to strongly scattering regimes, for example Capdeville *et al.* 2003; Tong *et al.* 2014), volumetric decompositions that isolate source-side, receiver-side or interior domains of interest (e.g. Monteiller *et al.* 2013; Masson & Romanowicz 2017; Wu *et al.* 2018), or distinct media types (e.g. acoustic versus elastic, most relevant for the oceanic case, e.g. Robertsson *et al.* 1996). They match and inject the wavefield at the domain boundaries and propagate the results from one solver onward using another. Philosophically related, our approach is perhaps closest to the work of Okamoto (1994), in that we use a transfer-function approach to couple wave-propagation regimes, using a catalogue of tailored response functions that we construct ahead of time.

Simon *et al.* (2020, 2022) developed signal-processing techniques to determine multiscale earthquake traveltime anomalies from arrival-time estimates made on MERMAID records, and their uncertainties, with respect to various reference models. In this paper we present a wave-propagation simulation technique capable of modelling MERMAID waveforms in their entirety, to help us measure cross-correlation-based traveltime anomalies that will enter future tomographic inversion efforts. Our workflow begins by determining an optimal bandwidth for the analysis, and then uses a combination of fast global modelling of the teleseismic waves that travel in a 1-D solid Earth from the source to the ocean bottom below MERMAID, and a 2-D simulation in a rectangular domain encompassing the ocean-solid interface to model the wave propagation as it enters the ocean layer and arrives at the instrument. We use pre-computed Green's functions in the 1-D model and a transfer-function approach for the 2-D 'oceanic last mile'. Synthetic seismograms can be generated quickly, capturing key modelling elements including the earthquake source mechanism and the ocean layer with high-resolution ocean bathymetry. We cross-correlate synthetics with observations from 16 MERMAIDs deployed in French Polynesia to obtain traveltime anomalies. We interpret our waveform measurements statistically and compare them with anomalies between ray-theoretical values and arrival times determined via the interval-variance ratio changepoint detection method of Simon *et al.* (2020, 2022).

2 DATA

Each MERMAID seismogram, reported after triggering, is a 4–6 min long time-series (all were recorded at 40 Hz, most are reported at 20 Hz sampling rate) in digital counts that we converted to acoustic pressure by removing the instrument response. Between August 2018 and June 2021, 16 Princeton MERMAIDs reported 3887 seismograms, of which 1433 were matched to 682 catalogue earthquakes (see Fig. 1). Simon *et al.* (2022) used wavelet-decomposed time-series and an Akaike information criterion (Simon *et al.* 2020) to perform arrival-time picking, match the records to catalogue earthquakes and determine the traveltimes to the MERMAID location at the time of recording, obtained by interpolation (Joubert *et al.* 2016). They calculated traveltime residuals with respect to ray-theoretical arrivals using TauP (Crotwell *et al.* 1999) in the `ak135` model (Kennett *et al.* 1995) adjusted for the acquisition depth of MERMAID in the water column.

Fig. 2 shows an unfiltered 20 Hz record section from a M 6.5 earthquake at 529 km depth in Indonesia (IRIS ID 10936816, CMT ID C201808171535A) reported by five MERMAIDs identified by their code number. Fig. 2(a) shows the geographical situation of the receivers at the time of recording, on a map that is colour-coded for the radiation pattern, the normalized P-wave amplitude at the receiver, which we call η, of the earthquake, shown as a 'beachball' moment tensor (Dahlen & Tromp 1998). White signifies compression ($-1 \leq \eta < -1/3$), dark red stands for dilatation ($1/3 < \eta \leq 1$) and the transition region ($|\eta| \leq 1/3$) is coloured light red. The solid red line is the plate boundary separating the Pacific Plate from the Australian Plate. Solid grey grid lines emanating from the earthquake are longitudes and latitudes (with 30° and 10° spacing, respectively) of a coordinate system with the epicentre as its north pole. Fig. 2(b) shows the instrument-corrected acoustic-pressure seismograms, individually normalized for display purposes, sorted by azimuth from the earthquake, as quoted on the right vertical axis. The left vertical axis displays the epicentral distance, in degrees. The horizontal axis displays the time relative to the arrival-time pick reported by Simon *et al.* (2022). The maximum absolute amplitudes of the segments shown, which are on the order of tens of Pascal (Pa), are listed at the beginning of each trace.

Fig. 3 shows filtered records from the same MERMAID receivers corresponding to the same event as shown in Fig. 2, after applying a two-pass, four-pole Butterworth filter with a corner-frequency pair that maximizes the SNR for each record, as discussed in detail in Section 3. The filtered records in Fig. 3 show the earthquake arrivals much more clearly, compared to the raw records shown in Fig. 2(b). The records, most clearly P0011 in this example, show wave packets that reverberate in the ocean layer. The black vertical bars are spaced in intervals of the slanted round-trip traveltime for an acoustic wave that bounces between the sea surface and the ocean bottom, calculated relative to the bathymetry directly below the MERMAID, but taking into account the incidence angle.

3 PRE-PROCESSING

Earthquake arrivals in MERMAID records are often dwarfed (see Fig. 2) by ambient ocean noise (Pipatprathanporn & Simons 2022), which we wish to remove (see Fig. 3) before modelling the waveform. Noise levels vary across the frequency spectrum and depend on the location and depth of recording. Similarly, frequency-dependent signal levels are influenced by earthquake magnitude, epicentral distance and the properties of the Earth between source and the receiver. Therefore, the frequency band that maximizes the signal-to-noise ratio (SNR) will differ from one record to another. We designed a bandpass filtering method to jointly maximize the SNR and bandwidth of individual MERMAID traces. Since there is no one-frequency-band-fits-all scenario, we search for the optimal pair of corner frequencies of a 2-pass, 4-pole Butterworth filter in the range from 0.40 to 2.00 Hz with a 0.05 Hz step size. The 0.40 Hz lower bound is the frequency at which the level of secondary microseismic noise (Nakata *et al.* 2019) is significantly reduced from its peak that is consistently observed near 0.20 Hz (Pipatprathanporn & Simons 2022). The upper bound is maintained at 2.00 Hz, the Nyquist frequency of the wave propagation simulation, which we discuss in Section 4. We require a minimum bandwidth of 0.50 Hz to preserve the temporal localization of the wave packet.

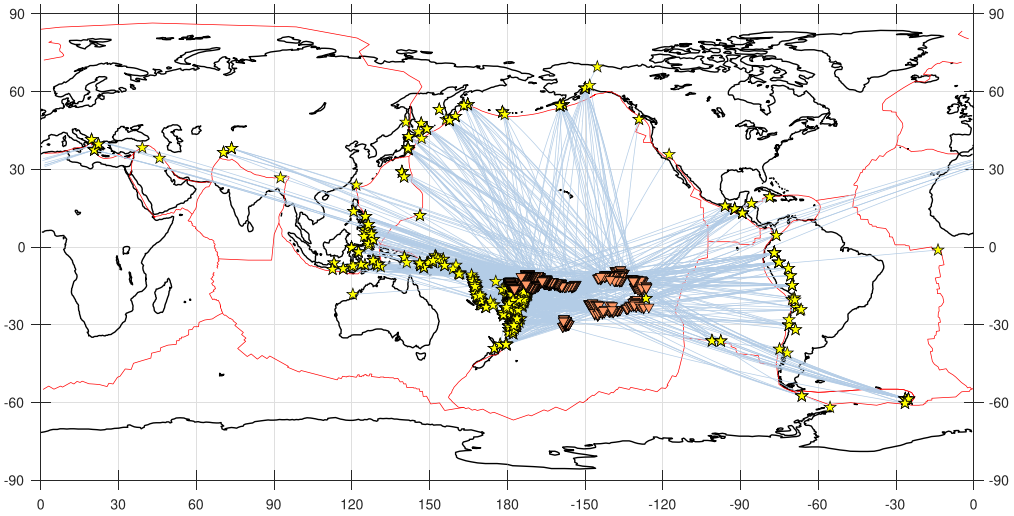

Figure 1. Epicentres of earthquakes (yellow stars) and subsurface locations of the MERMAID instruments recording them (orange triangles), connected by great circles (blue lines). Between August 2018 and June 2021, the sixteen Princeton University MERMAIDs reported 3887 seismograms from 682 earthquakes.

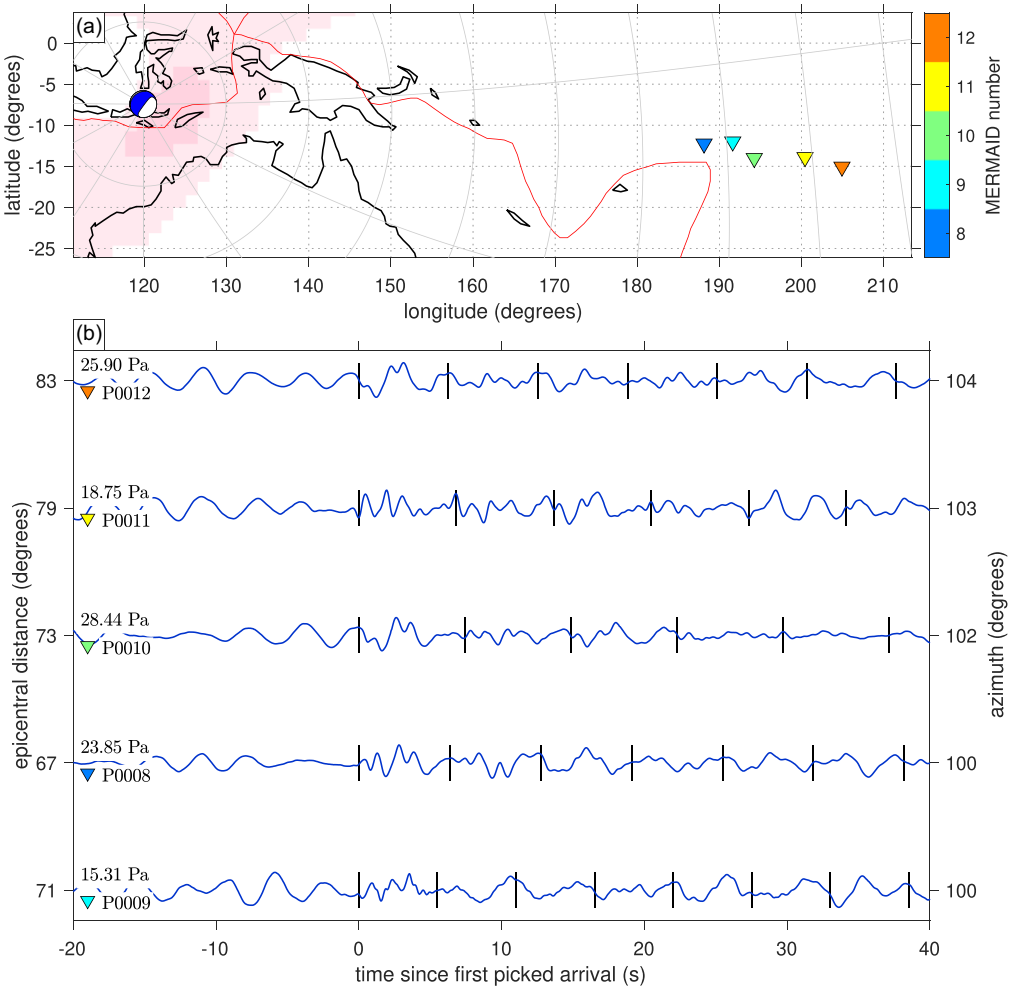

Figure 2. Unfiltered MERMAID earthquake record section. (a) Moment tensor and epicentral location of a *M* 6.5 2018 deep event in Indonesia, recorded by five MERMAID receivers (coloured triangles) with shorelines (black), plate boundaries (red) and radiation pattern (dark red for compression, white for dilatation, light red for the transition region). (b) Unfiltered acoustic pressure recorded by MERMAID, after removal of the instrument response. The traces are normalized for display by their maximum absolute amplitude quoted in Pascal (Pa). All records are aligned on the arrival times of the *P*-wave as picked by the procedure of Simon *et al.* (2022). The vertical bars count acoustic, slanted round-trip traveltimes in the oceanic water layer.

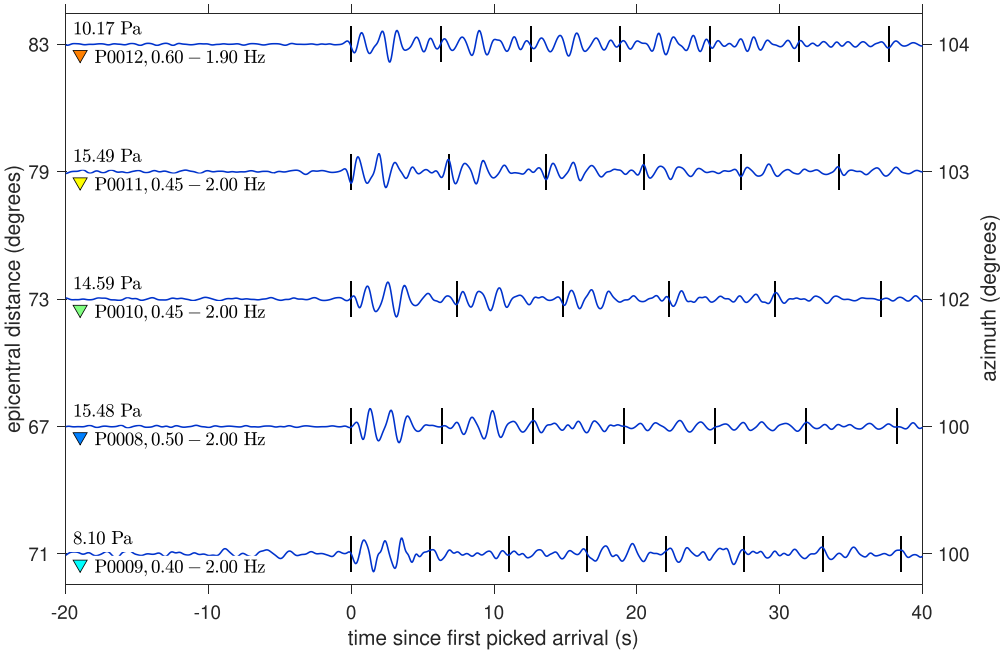

Figure 3. Filtered MERMAID earthquake record section. Layout and labelling are as in Fig. 2 but every trace has been bandpassed with the corner frequencies as labelled to maximize SNR using the method described in Section 3 and illustrated in Fig. 4. All records clearly show the arrival from the earthquake and its reverberations in the water layer, which are most distinctly noticeable for MERMAID P0011. Vertical bars mark round-trip ocean acoustic traveltimes.

As the trial separation between the 'noise' and the 'signal' windows we choose the Simon *et al.* (2022) arrival time, but we allow it to shift by half the duration of the windows in order to handle the spreading of the wave packets due to the filtering procedure. The window lengths are double the longest period in the frequency band considered, that is, twice the inverse of the lowermost corner frequency. This ensures that the signal window has at least two cycles without being so long as to include later arrivals. The SNR is then computed as the variance of the pressure within the signal window divided by that of the noise window (Simon *et al.* 2020).

Fig. 4 shows a detailed example of our bandwidth selection scheme, applied to the M 6.5 2018 Indonesia earthquake shown in Figs 2–3, recorded by MERMAID P0009 at 70.7° epicentral distance. Fig. 4(a) shows the time-dependent spectral density, the spectrogram, highpass filtered at 0.40 Hz, and computed over 20 s Hann-windowed sections with 70 per cent overlap. The horizontal black lines at 0.40 and 2.00 Hz mark the chosen corner frequencies. Fig. 4(b) shows the spectral density calculated using the Chave *et al.* (1987) method using 20 s segments with 70 per cent overlap, tapered with a discrete prolate spheroidal sequence (Simons 2010) with time-bandwidth product 4. The red curve is the median after highpass filtering at 0.40 Hz, surrounded by the 5th and 95th percentiles in grey. Also plotted for comparison, all in grey, are the equivalent three curves for the original data, before applying the highpass filter. The vertical lines at 0.40–2.00 Hz again mark the optimal bandwidth. Note that inside this frequency interval, the confidence interval is comparatively large, as the segment is composed of noise followed by signal, both spectrally distinct.

We considered four scenarios to define the 'optimal' corner frequency pairs. Simply (I) maximizing the SNR of the signal within the passband often led to narrow-band choices. Hence we modified this method to (II) maximizing the ratio of SNR in the passband to that in the complementary stopband, which we desire to be signal-free, (III) widening the bandwidth to yield a bandpass SNR at least 50 per cent of the overall maximum, and (IV) maximizing the bandwidth to yield a passband-to-stopband SNR ratio no smaller than 50 per cent of the achievable maximum. After visual inspection of the spectrograms and seismograms to appreciate the trade-offs involved, (IV) became our preferred method.

Fig. 4(c) shows the time-domain pressure records, filtered between 0.40 and 10 Hz (top), the (IV)-optimal bandwidth of 0.40–2.00 Hz (middle), and the complementary bandstopped portion (bottom). The SNR labels on the vertical axis are the variances of the signal window (orange) divided by that of the noise window (blue). Note that the vertical blue lines that divide the noise from the signal windows for each seismogram were allowed to shift by no more than half the window length as part of the optimization. The multiplicative factors indicate the scaling that was applied, for display purposes, to the middle and bottom seismograms.

Fig. 4(d) renders the ratios of the passband SNR to the stopband SNR on a grid of pairs of bandpass corner frequencies. The hidden third dimension of this panel is the time that separates the noise from the signal window. Every SNR *ratio* shown logarithmically in this panel was calculated for the optimal temporal split within both passband and stopband. In this example, the passband-to-stopband SNR ratio is highest for the range 0.60–2.00 Hz. Method (IV) picked out the range 0.40–2.00 Hz, shown by the red square, because of its wider bandwidth.

4 MODELLING AND MEASUREMENT

We must consider the ocean when modelling hydroacoustic pressure waves caused by teleseismic events. The effects of the ocean layer encompass the conversion from elastic to acoustic energy at the ocean bottom and the subsequent reverberations in the water column, which depend on ocean sound speed (whose temporal and spatial variation we may ignore) and bathymetry (which we will take

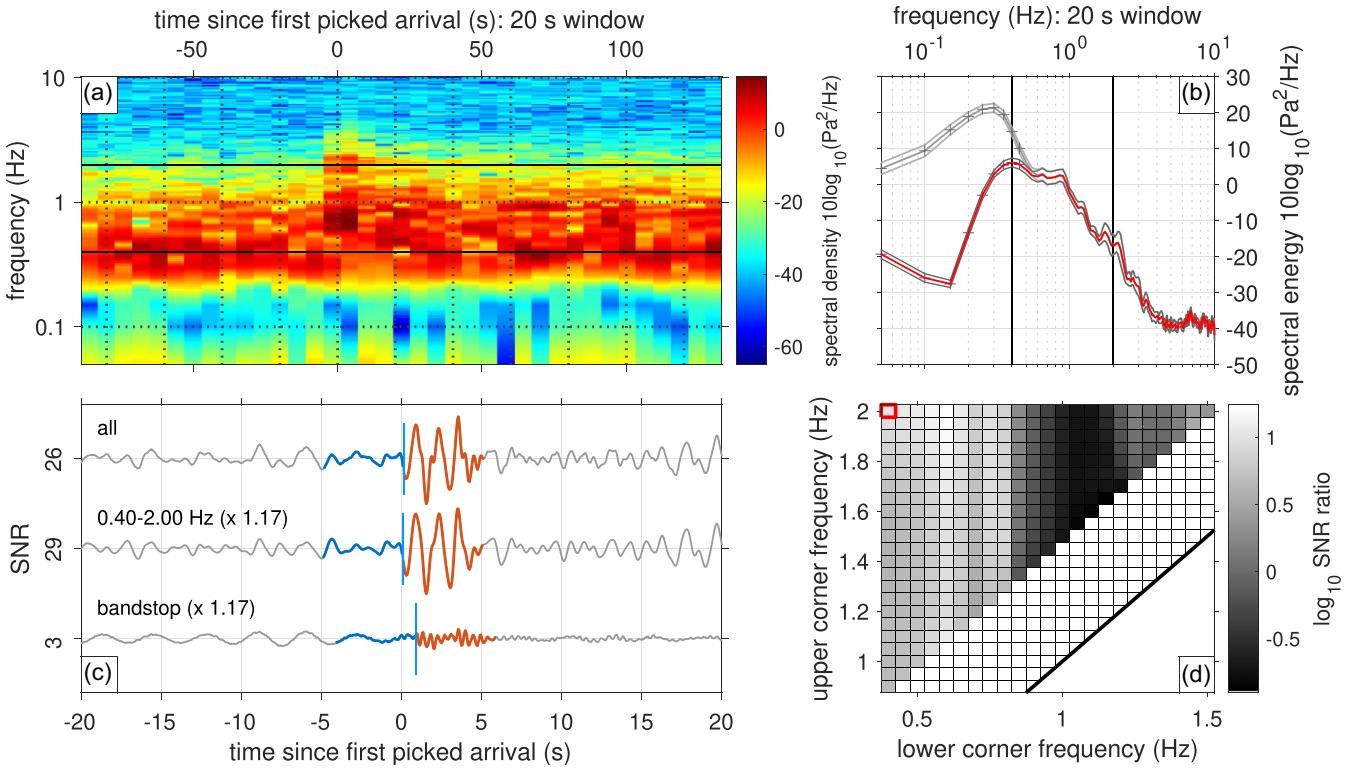

Figure 4. Bandwidth selection of hydroacoustic pressure time-series applied to the *M* 6.5 Indonesian earthquake recorded by a MERMAID float at 70.7°. (a) Spectrogram, highpass filtered at 0.40 Hz. The black horizontal lines indicate the chosen 0.40–2.00 Hz optimal corner frequencies. (b) Highpassed spectral density over the same time interval as in (a), calculated via overlapping segment analysis, showing the median (red), 5th and 95th percentiles (grey). The black vertical lines show the chosen 0.40–2.00 Hz optimal corner frequencies. Unfiltered spectral density is plotted as another set of grey curves for comparison. (c) Time-domain zooms, bandpassed between 0.40–10 Hz (top), in the optimal 0.40–2.00 Hz band (middle) and their bandstop complement (bottom). Blue and orange portions, separated by vertical blue lines, indicate the adjusted noise and signal windows. (d) The passband-to-bandstop SNR ratios of the seismograms filtered on a grid of pairs of corner frequencies. Lower corners range from 0.4–1.5 Hz, upper corners between 0.9–2.0 Hz, and the minimum bandwidth considered is 0.5 Hz. Method (IV), as described in the text, picks out the optimal range 0.40–2.00 Hz shown by the red square, by increasing the bandwidth while lowering the SNR ratio not far below its overall achievable maximum value, which is reached in the band 0.60–2.00 Hz.

into account). See also Dougherty & Stephen (1991), Komatitsch & Tromp (2002), Cristini & Komatitsch (2012), Jamet *et al.* (2013) and Bottero *et al.* (2020).

Modelling the waveforms recorded by MERMAID in our target 0.4–2.0 Hz window via self-consistent forward simulations in a 3-D earth model with a realistic ocean is very computationally expensive (Lecoulant *et al.* 2019; Fernando *et al.* 2020). Here, we present a new alternative method that treats the oceanic 'last mile' as the ultimate step in a procedure that first produces vertical-component motion on the ocean floor (in practice, via seismic wave propagation in a 1-D Earth), and implements the conversion from seismic to acoustic wave propagation as the action of a displacement-to-pressure response function which we determine as an intermediary step (using a 2-D model that captures the details of the ocean and the bathymetric interface in the direction of wave propagation).

Fig. 5 shows our modelling scheme. The first step involves the forward elastic wave propagation in a 1-D 'radial' earth model to obtain the displacement seismogram for the earthquake moment-tensor at the ocean bottom directly beneath the MERMAID, via the spectral-element package AxiSEM/Instaseis (van Driel *et al.* 2015). The second step solves for the forward propagation of an elastic plane wave that enters the solid part of the domain at the incidence angle appropriate for the incoming teleseism, converting to an acoustic pressure wave at the ocean bottom, using the 2-D spectral-element package SPECFEM2D (Komatitsch & Vilotte 1998; Komatitsch *et al.* 2000). The mesh honours the interface corresponding to the applicable bathymetric profile in the 'radial' direction, along the great-circle path between the distant earthquake source and the receiver. In this intermediary step we determine the local oceanic response by spectral division of the acoustic pressure at the MERMAID depth by the vertical displacement due to the slanted plane wave at the ocean bottom. Finally, producing the full simulated waveform for any given earthquake amounts to the convolution of the teleseismic vertical-displacement ocean-bottom synthetic from the first step with the local-ocean response obtained in the second step, which results in a synthetic hydroacoustic time-series that can be compared with observations made *in situ*.

4.1 The role of bathymetry

Bathymetry plays a role in wave conversion and transmission from the crust into the ocean layer, and vice versa, and in the multiple reflections of the pressure wave inside the ocean column. Beyond the first-order effects on waveform *timing* from the elevation of MERMAID above the seafloor (which can be obtained from the ocean depth directly below the float, noting that MERMAID maintains its *depth* with respect to the sea surface), the regional slope of the ocean bottom influences the kinematic relations between the incoming and

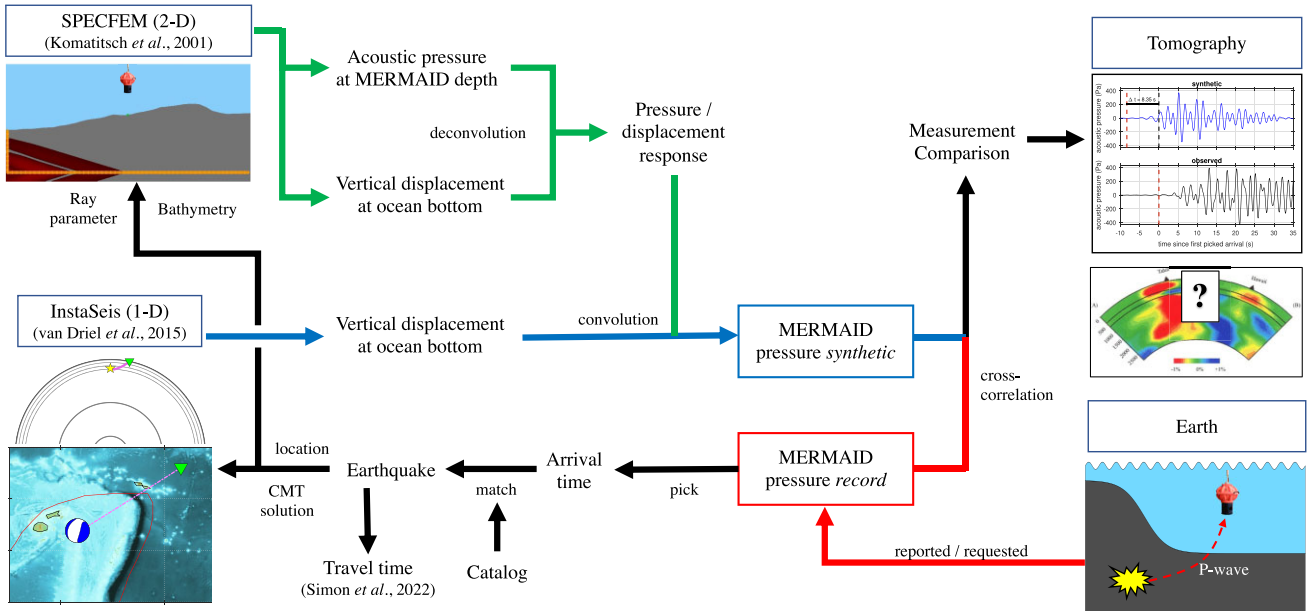

Figure 5. Waveform modelling methodology of MERMAID records. (Bottom right) MERMAID seismograms may be automatically reported after triggering or obtained upon request. An arrival time is picked and turned into a traveltime by matching the record with a catalogue earthquake (bottom left). The corresponding location and moment tensor are used for generating vertical displacement at the ocean bottom using a 1-D earth model, using Instaseis. The ray parameter is used to obtain the pressure/displacement response, using `SPECFEM2D`, taking into account the bathymetry, water depth and MERMAID depth (top left). The vertical displacement at the ocean bottom is convolved with the pressure/displacement response to obtain the synthetic pressure. The synthetic is cross-correlated in the bandwidth of interest with the observation to determine a traveltime anomaly, which can be used for tomographic inversion (top right).

the transmitted waves. Higher-order variations in bathymetry distort the wave front, introducing further complexity to the pressure records. Depth variations between the points of elastic-wave conversion into the acoustic system, as well as points of subsequent scattering and reflection, altogether influence the shapes and relative timings of the different wave packets making up the MERMAID wave train.

In striving to incorporate 3-D bathymetry into the setup of the forward simulations as best we can manage in our 1-D-to-2-D coupled modelling scheme, we construct bathymetric profiles for each event epicentre-to-MERMAID path along the great circles that connect them. We assume that the Earth is spherical both in shape and in wave speed distribution, hence the great-circle paths trace the surface trajectories of seismic waves propagating at depth inside the Earth. We linearly interpolate bathymetry available on a 15 arcsec grid (GEBCO Bathymetric Compilation Group 2019) along 20-km-long segments of the geodesics centred on individual MERMAIDs.

Fig. 6 illustrates this procedure. Fig. 6(a) shows a bathymetric map surrounding the M 6.5 Indonesian event, recorded by five numbered MERMAID receivers (coloured triangles). A great-circle segment connects the epicentre to MERMAID P0009. Fig. 6(b) shows bathymetric profiles in the radial directions centred on the MERMAIDs as identified by the triangles positioned off to the right. The profile for MERMAID P0009 is rendered bold. The point on the seafloor located directly below each MERMAID is marked by a cross, accompanied by the ocean depth at that location. To the left, epicentral distances, Δ, are listed in decimal degrees. Fig. 6(c) shows the situation and the ray geometry that applies to the `SPECFEM2D` simulation for MERMAID P0009. The teleseismic plane wave enters the rectangular domain from the bottom left with an incidence angle of 10.4° and, due to the wave speed contrast between crust (wave speed c_p) and water (speed c_w), refracts, towards the normal to the interface, into the water layer at the ocean bottom. The colour-filled triangle denotes the MERMAID float, whereas the unfilled inverted triangle represents a hypothetical ocean bottom sensor positioned directly below it. A choice of ray paths (green for upgoing and red for downgoing legs) and wave fronts (white lines) spaced 0.5 s apart are drawn.

4.2 Elastic-acoustic coupling via `SPECFEM2D`

To produce the oceanic response function that converts vertical displacement on the sea floor to the acoustic pressure recorded by MERMAID in the water column, we consider the elastic-acoustic system of a plane wave entering a 2-D modelling domain consisting of a homogeneous crust with variable topography overlaid by a homogeneous water layer. We respect the bathymetric profile along the great-circle path that connects the distant event to the floating hydrophone receiver. The plane-wave approximation, appropriate for teleseismic events, only requires specifying a ray parameter, which we predict from geometric ray theory within the 1-D reference model `ak135`. The homogeneity of crust and ocean wave speeds are non-limiting simplifications that can be abandoned provided the availability of more detailed information.

We use the spectral-element package `SPECFEM2D` (Komatitsch & Vilotte 1998; Komatitsch *et al.* 2000) to build the setup illustrated in Fig. 6(c). Our simulation domain is 20 km wide and 9.6 km tall, sufficient to cover the full-ocean depth except around the deepest trenches. The bottom and top layers consist of solid crust and fluid ocean water, respectively, with the parameter values listed in Table 1. The domain is meshed into 500 × 240 spectral elements (with five Gauss–Lobatto–Legendre nodes each) using the built-in `SPECFEM2D` mesher. The average size of the elements is 40 × 40 m^2 and their base shape is square, gradually

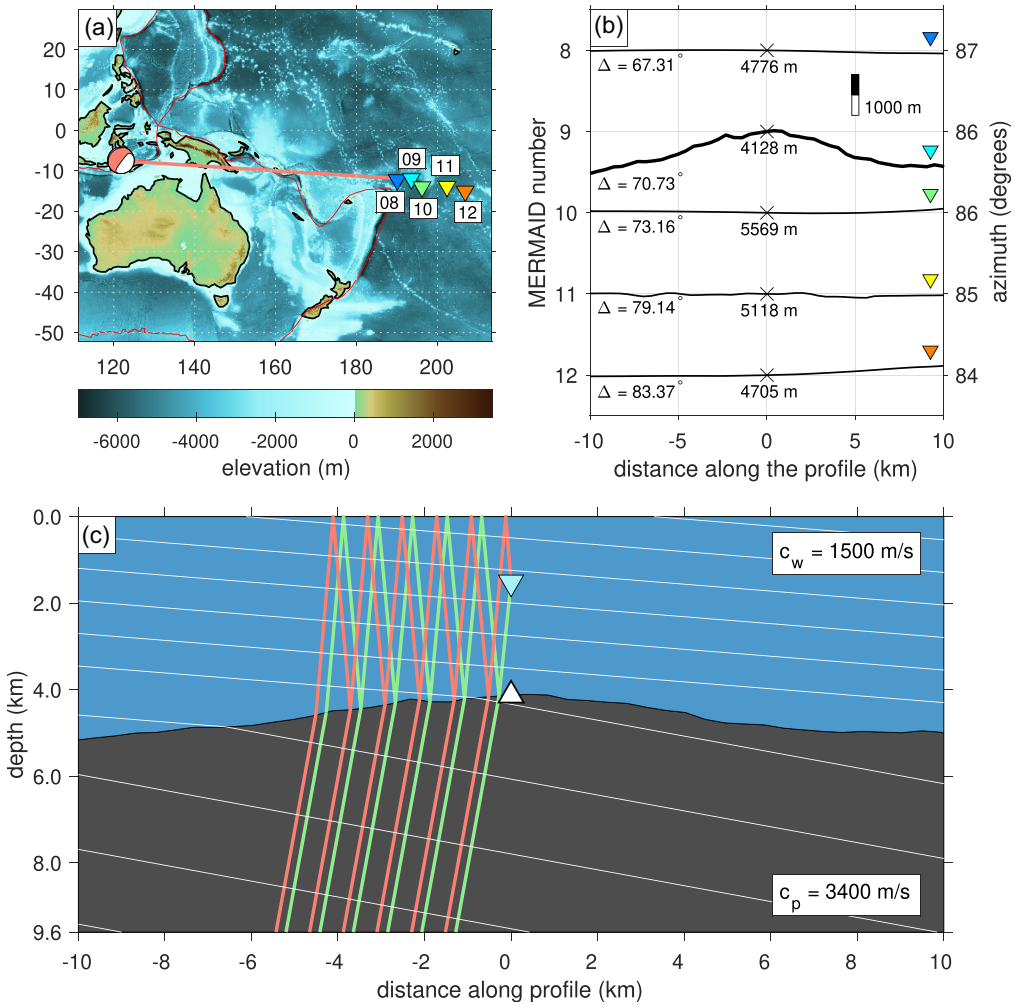

Figure 6. Determining the response function to convert displacement from teleseismic earthquakes at the seafloor into hydroacoustic pressure recorded by MERMAID in the water column via spectral-element modelling: a real data example highlighting the role of bathymetry. (a) Moment tensor and epicentral location of the 2018 Indonesian event, recorded by five numbered MERMAID receivers (coloured triangles), and the great circle connecting the source to MERMAID P0009 (red line), with GEBCO bathymetry. (b) Bathymetric profiles in the radial direction centred on every MERMAID, with shared scale bar. The seafloor depths directly below each MERMAID are marked by crosses. Epicentral distances Δ are in degrees. (c) Rays and wave fronts illustrating the setup for the SPECFEM2D simulation. A plane wave enters the crust (wave speed c_p) at an angle and refracts into the water layer (wave speed c_w) at the ocean bottom. The blue triangle denotes MERMAID and the white triangle a hypothetical ocean bottom sensor directly below it. Wave fronts, white, are spaced 0.5 s apart.

Table 1. Properties of the fluid ocean and the isotropic elastic crust used in the SPECFEM2D simulations for the displacement-to-pressure response calculations.

Medium	Density ($\mathrm{kg\,m^{-3}}$)	Compressional wave speed ($\mathrm{m\,s^{-1}}$)	Shear wave speed ($\mathrm{m\,s^{-1}}$)
Ocean	1020	1500	-
Crust	2500	3400	1963

deformed to exactly accommodate the interior domain boundary dictated by the bathymetry. In just 9 out of 1138 cases the layer thickness variation was too large to be handled by the SPECFEM2D internal mesher and the simulation failed. Stacey (1988) absorbing conditions are applied at all boundaries but the top ocean surface.

We create plane waves computationally by aligning point sources with ruptures timed to constructively interfere into producing a planar wave front with the desired incidence angle. The downgoing wave front is discarded once it leaves the domain. As we do not include the effects of bathymetry outside of the model domain, only sources within the crust are included, and none are placed within the water layer lest we should have to fine-tune their timing to match the shape of the acoustic pressure wave produced by a plane wave originating in the crustal portion outside of the domain. The size of the model domain is chosen to model about 30 s of wave reverberation, and to ensure that the paths of all bouncing arrivals at the MERMAID float within the simulation time originate from the bottom left side of the solid layer. Moreover, as the lines of point sources create non-planar, circular, wave fronts at their end points, due to the lack of cancellation from sources beyond the tips of the array, we make the domain and the source sequence large enough for the first arrival at the ocean-bottom receiver to be planar and sufficiently separated from subsequent arrivals generated by any circular wave fronts.

Again referring to Fig. 6, two receivers are placed within the domain. The first one records hydroacoustic pressure in the water column at the depth of MERMAID and is centred along the horizontal

axis. The second one records vertical displacement in the crust directly beneath the floating receiver. As the response function that we seek is the deconvolution of the vertical displacement record from the acoustic pressure time-series, it will be independent of the precise form of the source–time function. To drive the simulation we use a Ricker (1940) wavelet, the second derivative of a Gaussian (Wang 2015), with a 1 Hz dominant frequency. The simulations run for 32.5 s, that is, 65 000 steps with a 0.5 ms step size, using a 2nd order Newmark (1959) scheme, each run consuming 30 s on a 80 GB A100 NVIDIA multi-instance GPU.

Fig. 7 shows snapshots of the velocity wavefield from the `SPECFEM2D` simulation conducted as in Fig. 6. The vertical velocity is rendered red and blue with brighter colours signifying greater magnitude. Near the left and bottom boundaries of the domain, individual point sources are drawn as orange crosses. The two receivers (a MERMAID hydrophone in the water column and a hypothetical ocean-bottom sensor directly below it) are marked by green filled squares. Fig. 7(a) shows the wavefield at 1.6 s when the plane wave, manufactured by the constructive interference from a series of carefully placed and timed point sources, has entered the domain from the bottom left, propagating to the top right of the domain. At 3.1 s, Fig. 7(b) shows the entrance of the plane wave into the oceanic portion, distorted by the uneven ocean bottom, which reflects back some of its energy into the solid crust. At 3.9 s, Fig. 7(c), the pressure wave first arrives at the MERMAID float. Although this example is one that displays significant oceanic bathymetry, the shape of the wave front remains relatively planar in the water. Note the scattered waves that follow the leading wave front, which complicate the pressure record received by the hydrophone. After 5.9 s, Fig. 7(d) shows a snapshot of the wavefield as it starts to be reflected back from the ocean surface.

4.3 Response-function determination via deconvolution

The deconvolution of the vertical displacement at the ocean bottom from the acoustic pressure at the MERMAID depth is performed in the frequency domain (Bendat & Piersol 2010). All of the pressure time-series but only the first-arriving pulse of the ocean-bottom displacement is utilized, the latter gracefully tapered down using a Hanning window to discard any trailing coda signals. Using Nyquist-Shannon (1949) interpolation (Gubbins 2004; Scherbaum 2001) both 2 kHz synthetic time-series were resampled to the 20 Hz sampling rate reported by MERMAID from the field. We demeaned and detrended displacement and pressure and applied a two-pole, two-pass Butterworth low-pass filter with a 5 Hz corner frequency to filter out high-frequency content falling outside the target modelling band of 0.4–2 Hz. Both time-series, p and s, are Hanning-tapered before conversion to the frequency domain via Fourier transformation, $\tilde{p} = \mathcal{F}\{p\}$ and $\tilde{s} = \mathcal{F}\{s\}$, and spectrally divided with a regularizing 10^{-4} per cent water-level damping to obtain $r = \mathcal{F}^{-1}(\tilde{p}/\tilde{s})$, the pressure response in the ocean to vertical displacement on the seafloor. The convolution $\hat{p} = s * r$ is compared against the original waveform p to check the quality of the deconvolution step.

For the ray parameter of an earthquake at $\Delta = 70.7^\circ$, Fig. 8 shows the pressure, p, and its estimate, $\hat{p}$, obtained by (de)convolution of the vertical displacement, s, and the response function r. This procedure is repeated for every event-MERMAID pair in our catalogue. In order to assess the consequences of neglecting 3-D effects and bathymetric uncertainty, Fig. 8 includes, in grey, the pressure and displacement records and response functions for bathymetric profiles taken $\pm 45^\circ$ azimuthally from the radial direction of seismic wave propagation. Only secondary arrivals in the pressure traces are affected in this particular case, which validates our approach to first order.

4.4 Axisymmetric spectral-element modelling via Instaseis

The waveform modelling package Instaseis (van Driel *et al.* 2015) uses seismic Green's functions precomputed by the axisymmetric spectral-element method AxiSEM (Nissen-Meyer *et al.* 2014) to rapidly generate seismograms at any station worldwide. For our simulations we used earth model `ak135f_1s` to suit our target frequency range of 0.4–2 Hz. This particular model assumes the ocean to be a 3-km-thick fluid layer. We specify the earthquake moment tensor in order to take the radiation pattern into account. The moment-rate source–time function is a Gaussian with a scale given by the half-duration which scales as the cubed root of the scalar moment (Ekström *et al.* 2012). Instaseis does not compute or store Green's functions for buried receivers, hence the nominal receiver is a station at the sea surface, at the same latitude and longitude as MERMAID, recording vertical displacement. To obtain the seismogram at the depth of the ocean bottom, we shift the time of the seismograms by the expected traveltime of the first-arriving phase (usually *P* or *PKIKP*) from the ocean bottom to the surface, assuming the *P*-wave speed of the crust in model `ak135` and calculating the incidence angle of the phase from the corresponding ray parameter. Since we focus on modelling waveforms in a few seconds surrounding the arrival time of the first phase, we ignore incorrect adjustments of the later phases. Convolution with the response as described in Section 4.3 produces synthetic pressure waveforms for comparison with actual MERMAID records.

Fig. 9 shows the application of the response function obtained as illustrated in Fig. 8 to matching the waveforms from the Indonesian earthquake. Fig. 9(a) shows the pressure observed by MERMAID P0009 (blue) alongside the synthetic pressure obtained via our procedure, before (grey), and after (red) alignment with the observed pressure time-series. All records shown were filtered between 0.40 and 2.00 Hz, and presented at the 20 Hz sampling rate of the observations. Fig. 9(b) shows the synthetic vertical displacement at the ocean bottom below MERMAID, calculated by Instaseis. Fig. 9(c) shows the displacement-to-pressure response function applied to it, repeated from Fig. 8(c).

Fig. 10 shows a record section with synthetic displacement seismograms modelled by Instaseis for the M 6.5 2018 Indonesian earthquake. Shown are synthetic vertical displacements computed by Instaseis at the ocean bottom below the five MERMAID receivers, timed relative to 1-D ray theory. The traces are filtered by two-pass, four-pole bandpass Butterworth filters, with the labelled corner frequency pairs established via the procedure laid out in Section 3. The arrival times picked on the MERMAID records by Simon *et al.* (2022) are marked by black filled circles.

4.5 Traveltime residual measurements by cross-correlation

Simon *et al.* (2022) determined traveltime residuals, of arrival-time picks with respect to predictions using ray theory in a 1-D reference model, adjusted for the presence of the ocean. Here, we measure traveltime residuals by maximizing the cross-correlation of the newly obtained synthetic waveforms with the observations (Luo & Schuster 1991; Dahlen *et al.* 2000; Mercerat & Nolet 2013; Yuan *et al.* 2019), after filtering both records using identical four-pole,

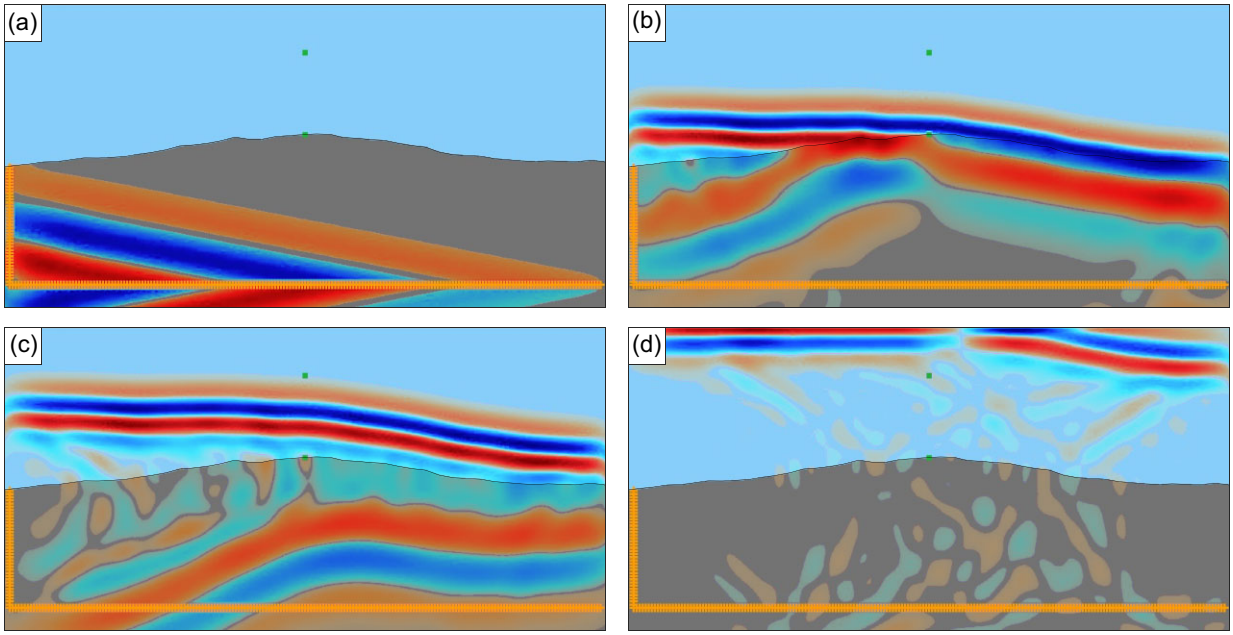

Figure 7. Snapshots from the SPECFEM2D simulation set up as in Fig. 6, for the M 6.5 2018 Indonesia earthquake. The vertical velocity wavefield is rendered in red for positive and blue for negative of varying intensity. Lines of point sources shown as orange crosses generate the plane wave. Wavefield snapshots at (a) 1.6 s, (b) 3.1 s, (c) 3.9 s and (d) 5.9 s. The green filled squares represent a mid-column floating MERMAID pressure-recording hydrophone and a fictitious ocean-bottom displacement sensor, between which our procedure seeks to characterize the response function.

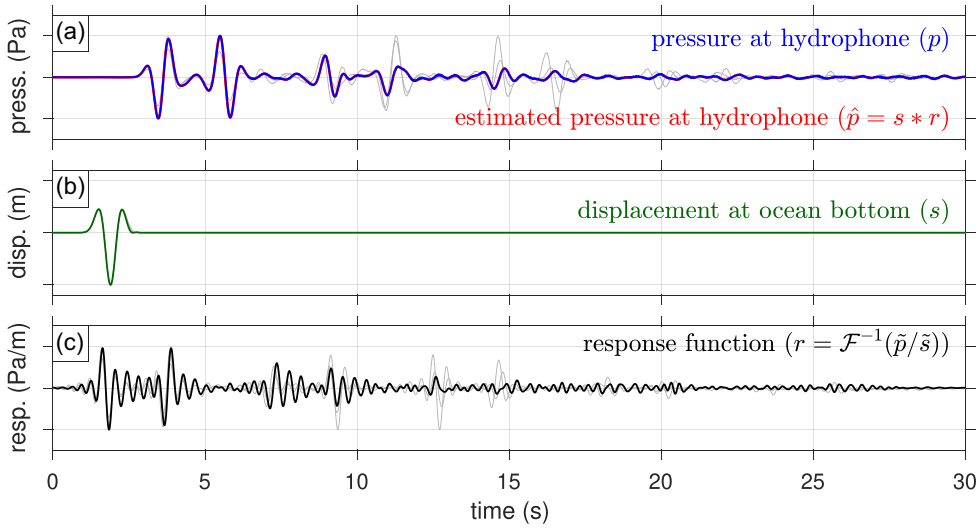

Figure 8. Oceanic response function determination for teleseismic data, from the deconvolution of vertical ocean-floor displacement from hydroacoustic pressure in the water column, obtained by a coupled elastic-acoustic bathymetric spectral-element simulation of an incoming plane wave. (a) Pressure synthetic at a MERMAID parking depth (in blue). (b) Vertical displacement (green) due to a plane-wave Ricker wavelet at the ocean bottom. (c) The displacement-to-pressure response function (black). Applied to the trace in (b), convolution with the response results in a perfectly estimated pressure, shown in red in (a). The grey traces in (a)–(c) are the pressure synthetics, vertical displacements, and displacement-to-pressure response functions, respectively, when the bathymetry cross-sections are 45° away from great-circle path alignment in either direction.

two-pass Butterworth filters with the corner frequencies chosen as described in Section 3.

We must take several precautions to guarantee robustness and stability of our measurements. MERMAID pressure records consist of series of wave packets caused by oceanic reverberation, which makes cycle-skipping in the comparison between short waveform segments likely, while using extended analysis windows to capture the entire wave train introduces unwanted complexity. More than one phase may arrive within the selected window. Especially for shallow earthquakes (above 100 km depth), our wave-propagation simulations revealed numerous instances where the pP depth phase dominates in amplitude over the P wave arrival. Separations between those two arrivals are in the 10 s range, which is large for a traveltime residual but not improbable for distant earthquakes. To ensure that all traveltime residuals result from earth structure and not picking error, we limit the waveform matching by centring the

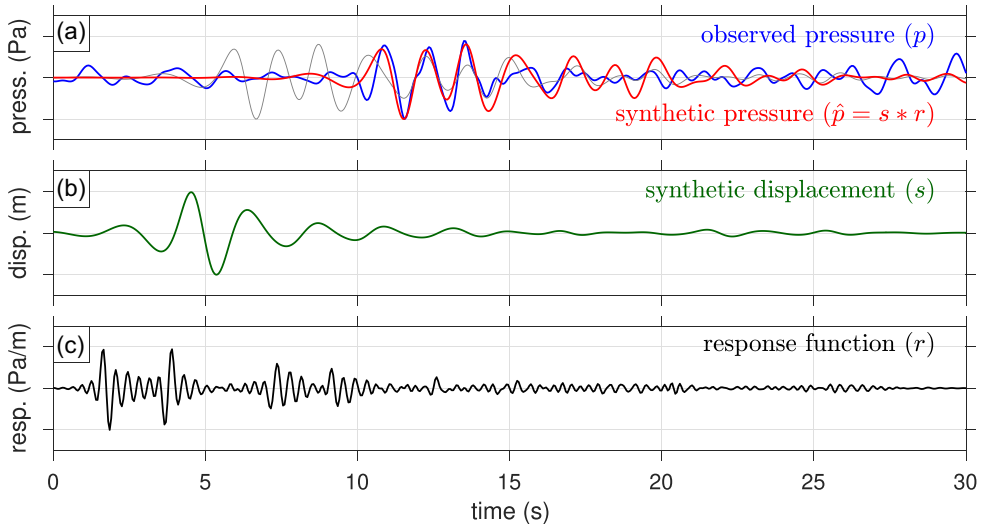

Figure 9. Application of the response function determined in Fig. 8 to the *M* 6.5 2018 Indonesia earthquake. (a) Pressure observation (blue) at MERMAID P0009 at a parking depth of 1500 m floating at 2610 m above the ocean floor. Synthetic pressure (grey) obtained by convolving the synthetic displacement at the ocean floor with the oceanic response function. In red, the pressure synthetic after shifting by the 4.85 s lag that maximizes the cross-correlation between the observed and the synthetic traces. (b) Synthetic vertical displacement (green) at the ocean bottom below MERMAID, calculated by Instaseis. (c) The displacement-to-pressure response function (black), as in Fig. 8(c).

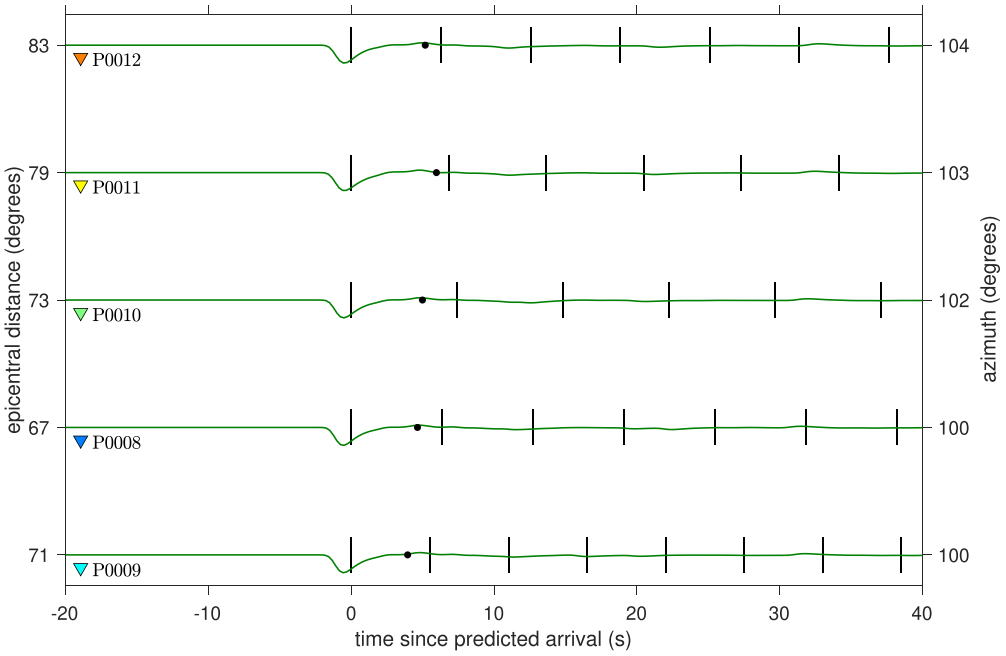

Figure 10. Synthetic record section. Vertical displacement computed by Instaseis at the ocean bottom below MERMAID, filtered by a 2-pass, 4-pole Butterworth bandpass with corner frequencies as individually labelled. All traces are aligned on the *P*-wave arrival time predicted by ray theory within the 1-D model `ak135`. The traveltime anomalies identified by Simon *et al.* (2022) on the MERMAID records are marked by black filled circles.

search window on the ray-theoretical arrival-time prediction (calculated at the ocean bottom by TauP, then time-adjusted for the ocean transit). Adding in the time to the first peak in the response function (see Fig. 9c) we check whether that time falls within the greater of 2 per cent of the predicted traveltime or 5 s counted from the picked arrival time, and if does not, we first correlate the envelope of the waveforms, computed using the Hilbert transform. Only then do we allow for small adjustments in phase alignment by another round of waveform correlation.

Regarding the *envelope* correlation, we trim the two waveforms between 20 s before and 20 s after the expected arrival of the first wave, in order to keep both a pre-arrival section consisting only of ambient noise and a post-arrival section containing a few reverberations. We cross-correlate the envelopes of the waveform segments, limiting the time shifts by tapering to within a maximum of 15 s relative delay. The lag between the envelopes maximizes the windowed cross-correlation function thus obtained. The waveform correlation that follows begins with that time shift. As to the *waveform*

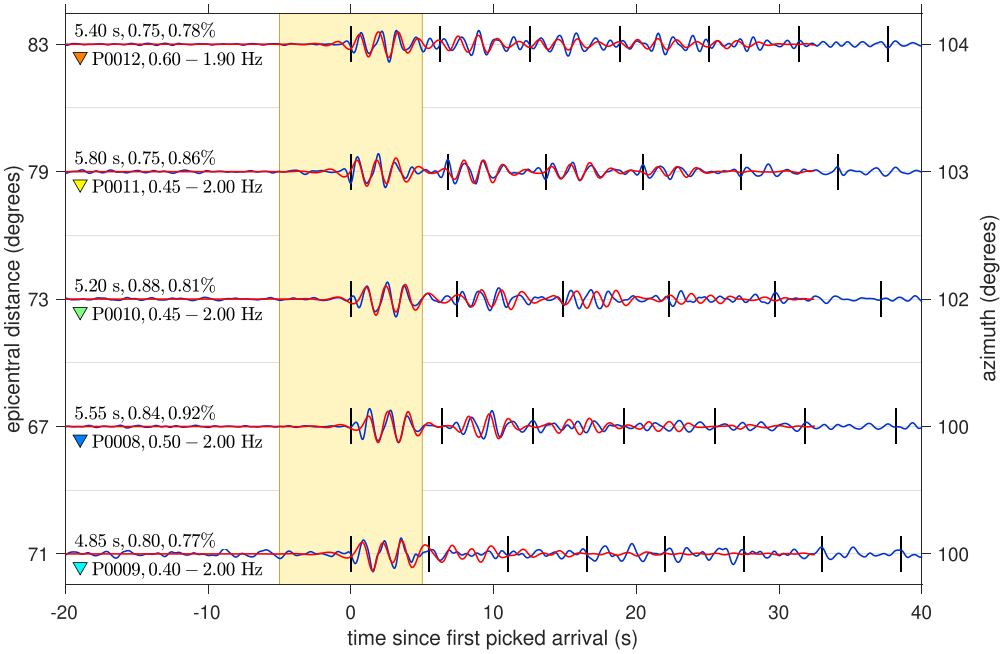

Figure 11. Determining cross-correlation traveltime delays between synthetics (red) and observations (blue) for the 2018 Indonesian event. The synthetic waveforms were amplitude-scaled and time-shifted to maximize the correlation with the observations. Time is relative to the arrival time picked by Simon *et al.* (2022) on the observations and the vertical bars count oceanic water layer reverberations. To the left below each trace we list optimal corner-frequency pairs. Above the traces we write the traveltime residuals (the time added to the synthetic pressure waveform to align with the observed pressure waveform), the correlation coefficient (between the synthetic and observed waveforms calculated within the yellow window 5 s before and after the picked arrival time), and the relative traveltime residual expressed as a percentage of the traveltime calculated in the 1-D reference model `ak135`.

correlation, we cut the synthetic from 5 s before to 5 s after the selected starting point and correlate with a 10 s window of the observed pressure waveform centred at the picked arrival time. Hence we are able to focus on aligning the waves within the first wave packet and exclude the subsequent arrivals which could be distorted by uneven bathymetry. The quoted traveltime residual adds the time shift from envelope correlation to that obtained from waveform correlation.

Fig. 11 shows the cross-correlation of the synthetic pressure waveforms (in red) and the corresponding observed pressure waveforms (in blue) from five MERMAID receivers reporting the 2018 M 6.5 earthquake in Indonesia. Shown are the time-shifted, scaled synthetic (red) pressure waveforms, and the observed (blue) pressure waveforms, timed relative to the arrival time picked by Simon *et al.* (2022). Traveltime residuals, maximum correlation coefficients and the traveltime residuals relative to the prediction within 1-D reference model `ak135` are marked on the left above each trace. The cross-correlation window, from 5 s before to 5 s after the picked arrival time, is highlighted. For many traces, the synthetic waveforms match the observed waveforms even outside the highlighted window, demonstrating the overall robustness and reliability of the waveform modelling.

To aid the subsequent discussion, Fig. 12 summarizes concepts and nomenclature used so far in this paper and in the *opus citatum* by Simon *et al.* (2022). The seismograms are from the same record by MERMAID M0009 of the M 6.5 event in Indonesia in 2018 that we have used throughout this paper. The MERMAID pressure record (top, in blue) is bandpassed between the optimally determined corner frequencies of 0.40–2.00 Hz. The synthetic trace (middle, in red) is the predicted pressure at the MERMAID depth bandpassed using the same filter. The bottom traces are the synthetic vertical displacement at the ocean floor right below the MERMAID, both in the raw original (dark green, thick line) and filtered in the same way as the other two traces (light green, thin line).

A generic traveltime anomaly is the difference between the timing of a phase arrival observed in a seismogram and its model prediction. Simon *et al.* (2022) defined the traveltime anomaly in MERMAID records as their picked arrival minus the ray-theoretical prediction by TauP in the `ak135` model, adjusted for the effect of the water column (bathymetry and MERMAID cruising depth). This quantity, which they called t^*_{res}, is depicted as the blue arrow on the seismogram at the top of Fig. 12. In our present study, however, the traveltime anomaly, $\Delta\tau$, is not based on any picked arrival, rather, it is determined by cross-correlation as the amount of the time shift required for the synthetic waveform to optimally align with the pressure record observed by MERMAID (as was shown in Fig. 11). It is represented by the red arrow above the middle seismogram in Fig. 12. The slanted lines guide the eye towards the ultimate alignment of both waveforms.

The green arrow underneath the Instaseis seismogram at the ocean bottom (which we recognize from Fig. 10) shown at the bottom of Fig. 12 corresponds to the difference between the arrival that we would pick on a displacement seismogram (marked as an upside-down triangle) and its arrival predicted by TauP (marked as an upright triangle). We measure this discrepancy on every seismogram in our data base and apply it as a correction term to address the difference between (infinite-frequency) ray tracing predictions and arrival picks made on actual (finite-frequency) records, that is, whenever we shall next compare our cross-correlation traveltime anomalies $\Delta\tau$ to the t^*_{res} measured and tabulated by Simon *et al.* (2022). See also Appendix A.

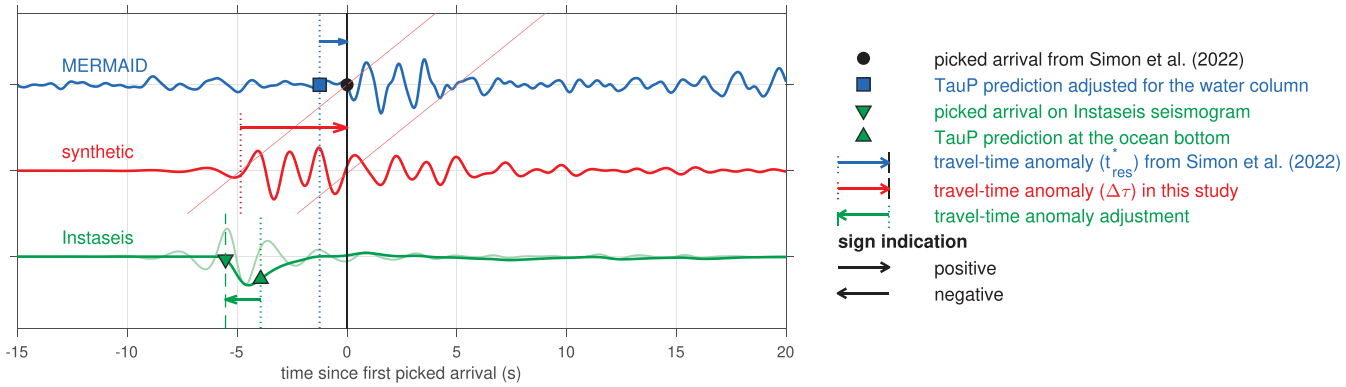

Figure 12. Terms and concepts to describe 'traveltime anomalies' measured on MERMAID records, in the ray-theoretical sense used by Simon *et al.* (2022) and in the finite-frequency cross-correlation framework described by the present paper. The observed trace (top, blue) is the pressure record from the 2018 Indonesian earthquake at 70.7°, bandpassed between 0.40 and 2.00 Hz. The synthetic trace (middle, red) is the identically filtered prediction obtained via the methods introduced in this paper. The bottom traces are the vertical displacement at the ocean bottom directly underneath MERMAID, both unfiltered (dark green), and filtered identically to the MERMAID traces (light green). All traces are normalized by their absolute maximum. The time axis is relative to the arrival time picked by Simon *et al.* (2022), whose traveltime anomaly t^*_{res} (blue arrow) is defined as the picked arrival minus the TauP prediction adjusted for the water column. The traveltime anomaly in this study $\Delta\tau$ (red arrow) is the time shift to apply to the synthetic waveform in order to align it with the observed record via cross-correlation. Arrival times picked on synthetic displacement seismograms are not identical to ray-theoretical predictions made at infinite frequency. The green arrow in the bottom seismograms shows their difference, which we apply as a correction to the broad-band traveltime anomalies obtained in this study, to compare with the traveltime anomalies of Simon *et al.* (2022). Note that if an arrow points right, the corresponding value is positive.

5 APPLICATION TO THE MERMAID SPPIM CATALOGUE

We are finally in the position to apply the methods for frequency-band selection (Section 3) and waveform prediction and measurement (Section 4) developed in this paper, on the 1129 records (Section 2) collected by sixteen Princeton MERMAID floats between August 2018 and June 2021 as part of the ongoing South Pacific Plume Imaging and Modeling (SPPIM) project led by the French institute for ocean science Ifremer under the flag of the international academic EarthScope-Oceans consortium. The full catalogue of picked and earthquake-matched records from 61 different floats owned by the collective EarthScope-Oceans institutions continuously operating between August 2018 and December 2023 contains 11 138 records from 3763 earthquakes and will be updated online and uploaded to international data centres.

Figs 13–14 are graphical representations of where our data 'live' with respect to their chosen bandwidths and SNRs, and in terms of the maximum cross-correlation values obtainable with the modelled traces after scaled time-shifting with respect to the observations. The occupied bandwidth of each trace is represented by a coloured vertical strip. The sorting is in order of increasing occupied bandwidth from left to right in all panels. Next in the sort order is the frequency midpoint of the chosen band, and then, either, the SNR of the bandpassed traces (for Fig. 13), or the maximum correlation coefficients at the optimal time shift (in Fig. 14). Each of those vertical bars is coloured either by the SNR (in the top panels, Figs 13a and 14a), or by the maximum correlation coefficient (in the bottom panels, Figs 13b and 14b). The occupied bandwidth is generally wide, with the bandpass window of the bulk of the observations terminating at the maximum modelable upper corner frequency of 2 Hz. Many of the lower corner frequencies are around 0.6 Hz. The median bandwidth is 1.45 Hz. SNRs range between 3 and 10 000 with a median of 39. The maximum correlation coefficients range from 0.05 to 0.97 with a median of 0.72. No strong relation was detected between bandwidths, SNRs, and correlation coefficients. In other words, the goodness of the waveform match is not easily predicted by the quality of the earthquake signals in the record.

Fig. 15 sums up all of our measurements. Fig. 15(a) shows the distribution of the traveltime residuals in our data set, measured as described in Section 4.5. The distributions of the SNRs and maximum cross-correlation coefficients that correspond to these measurements are shown in Figs 15(b) and (c), respectively. Fig. 15(d) shows the traveltime anomalies after normalization by the arrival-time obtained by ray tracing in 1-D model `ak135`, where they are roughly equivalent to a relative traveltime anomaly, in per cent. Fig. 15(e) shows the distributions of the occupied bandwidth, and Fig. 15(f) shows the effective coverage of the frequency spectrum, which amounts to the zeroth norm (row sum, tracking whether occupied or not) in the row dimension of any of the panels Figs 13 and 14. The corresponding great-circle epicentral distances are summarized in Fig. 15(g), event depths in Fig. 15(h), and backazimuths in Fig. 15(i).

Excluding 28 outliers, the median of the traveltime residuals shown in Fig. 15(a) is a positive 3.5 s, or a median of 0.91 per cent for the relative values shown in Fig. 15(d). Hence, according to our measurements, the *P*-wave speed of the actual Earth along the various trajectories sampled (see Fig. 1) is smaller that the 1-D `ak135` model average, as is consistent with the presence of pervasive low-velocity anomalies in the broad mantle domain around the Pacific Superswell (see also Simon *et al.* 2022, and references therein), which our waveform measurements will help constrain tomographically. The SNRs shown in Fig. 15(b) range from 3 to 10 000 with a median of 39. As to the correlation coefficients shown in Fig. 15(c), their median is 0.72, and more than 70 per cent of all measurements have a correlation coefficient of 0.60 or greater. Per Fig. 15(e) the median of the occupied bandwidth is 1.45 Hz, and Fig. 15(f) shows the favourable frequency coverage of the entirety of our measurements made on our complete data set. Fig. 15(g) shows that the majority of our records correspond to earthquakes closer than 100°, about half of which are nearer than 20°. The handful that appears beyond the core shadow are from core-phase arrivals such as *PKIKP*

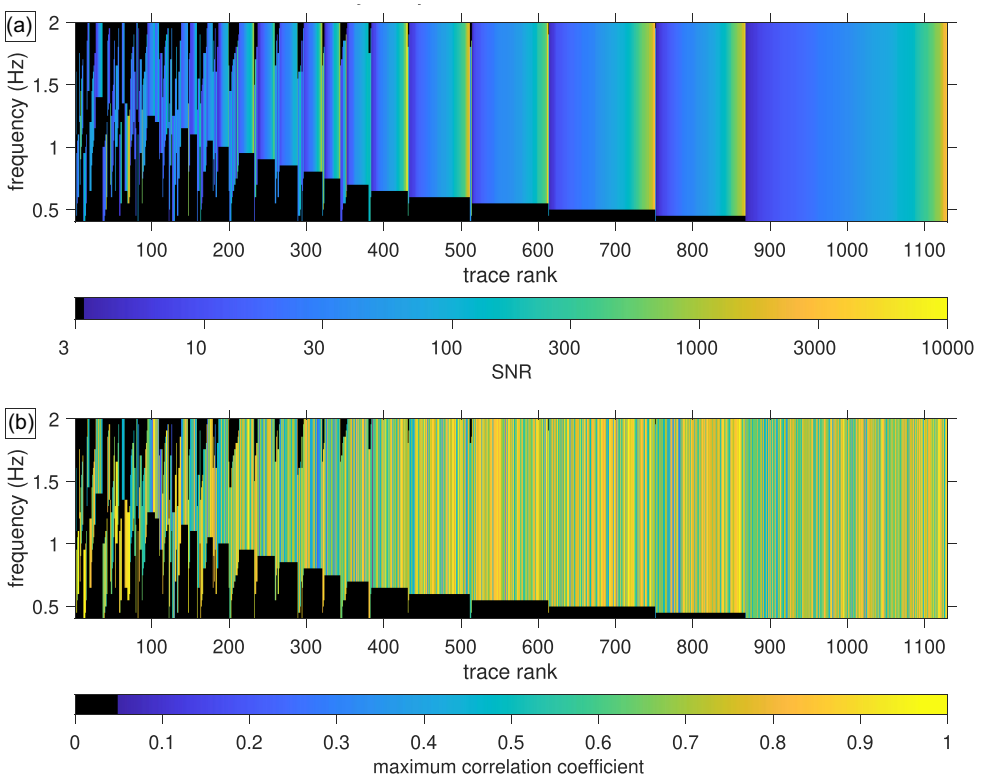

Figure 13. Occupied bandwidth and signal metrics of our data set. The sort order of the vertical strips representing every MERMAID seismogram is by occupied bandwidth (i.e. by their length, increasing from left to right) followed by the midpoint of the band, followed by the signal-to-noise ratio of the observations, for both panels. In the top panel (a), the colour maps the signal-to-noise ratio, whereas in the bottom panel (b), the colour renders the correlation coefficient.

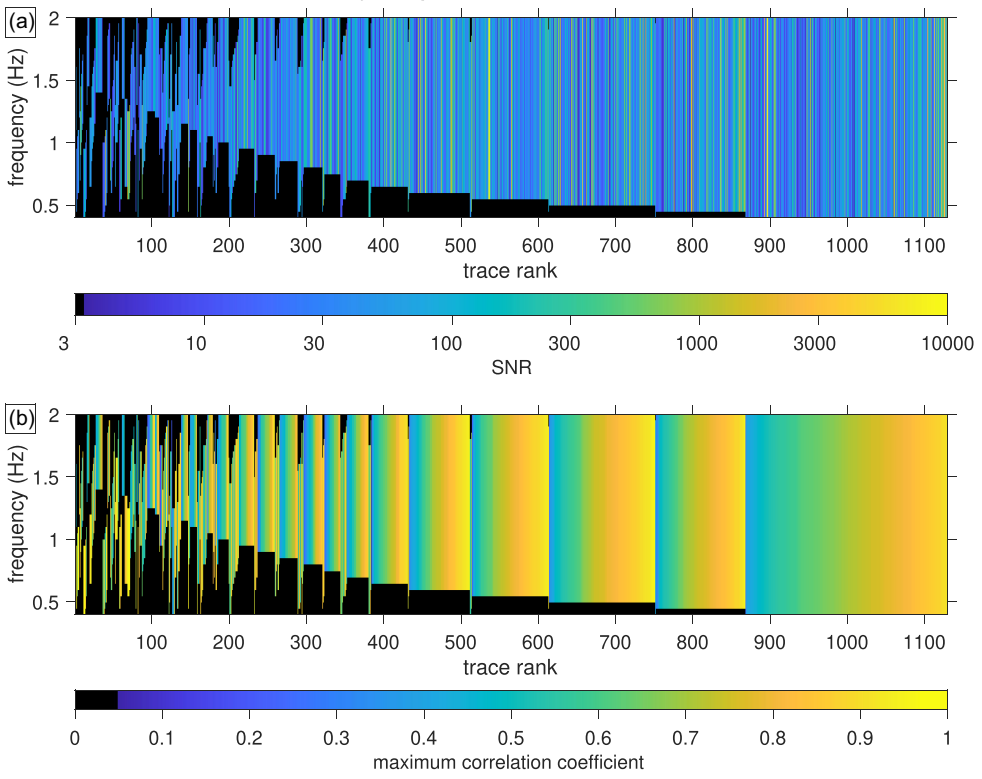

Figure 14. Occupied bandwidth and signal metrics of our data set represented as in Fig. 13, and as there, sorted by occupied bandwidth first, and by midband second, but in contrast, the final sort order is the correlation coefficient between the observations and the synthetics. The colour maps to the signal-to-noise ratio in (a) and to the correlation coefficient in (b). As in Fig. 13, the smooth colour gradations reveal the property that is last in the sort order.

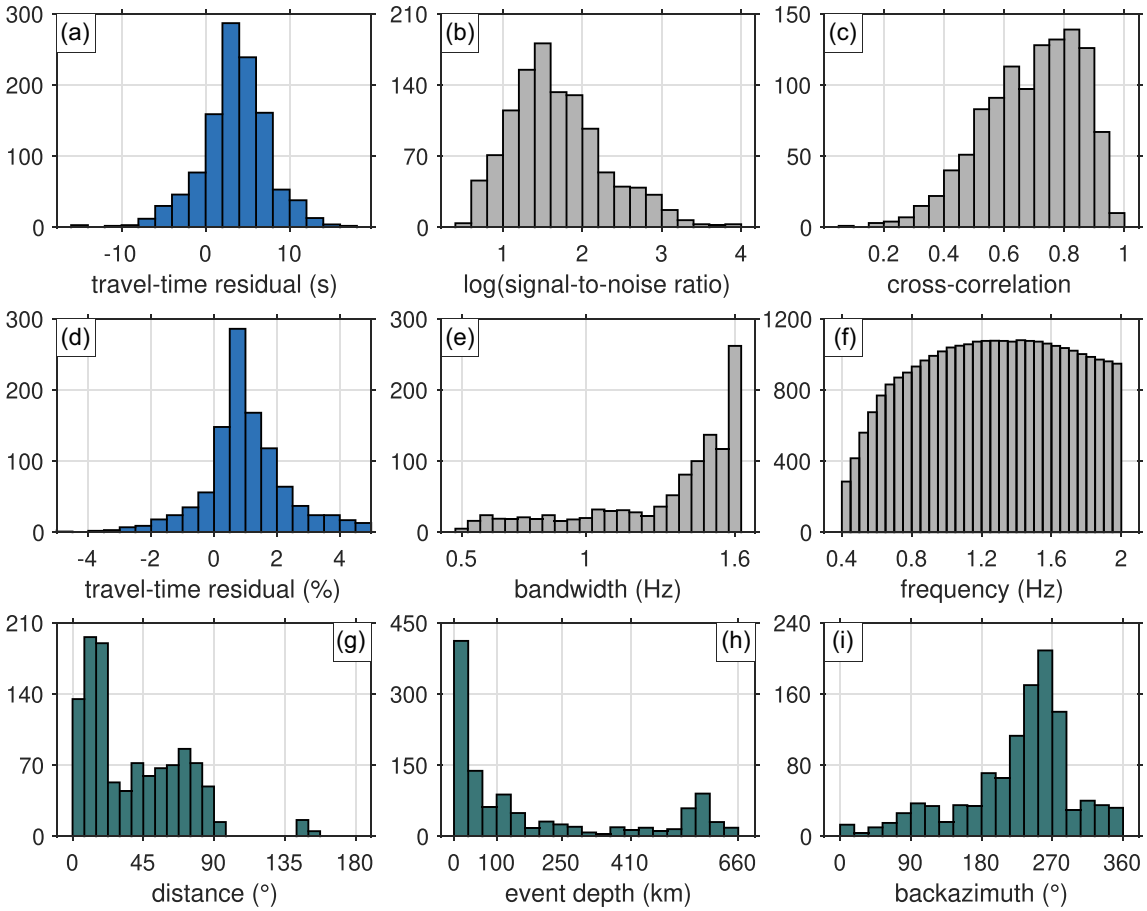

Figure 15. Distribution of data and measurement metrics in our data set. Histograms showing the (a) traveltime residual, (b) data signal-to-noise ratio, (c) maximum correlation coefficient between observed and modelled data, the (d) traveltime residual relative to the wave speed in `ak135`, (e) the occupied bandwidth, (f) the overall coverage of the available spectrum, (g) epicentral distance, (h) depth and (i) backazimuths.

(see also Simon *et al.* 2021). The earthquake depth histogram in Fig. 15(h) shows that the deepest events are in the transition zone while the majority are crustal events shallower than 25 km. The backazimuthal distribution in Fig. 15(i) is reflective of the central location of the MERMAID array, surrounded by the Pacific 'Ring of Fire' (see again Fig 1). The majority of the records are from earthquakes located southwest or west from the floats, with a great concentration in the nearby Tonga trench.

6 DISCUSSION

Cross-correlation waveform-synthetics-based traveltime measurements sense the Earth differently compared to ray-theoretical traveltime anomalies made from high-frequency 'picks' (e.g. Dahlen *et al.* 2000; Hung *et al.* 2000; Nolet & Dahlen 2000; Hung *et al.* 2001; Tromp *et al.* 2005; Mercerat & Nolet 2012). Since we developed our methodology and performed all of our measurements on the data set presented by Simon *et al.* (2022), we can now compare our (variably bandpassed, cross-correlated with synthetics) finite-frequency (limited to 2 Hz) results with their catalogue, which measures the difference between an 'event' or 'changepoint' arrival-time pick (Simon *et al.* 2020), made on a (typically 5 Hz) pressure waveform, with a ray-theoretical phase arrival-time prediction made by ray tracing. In addition to providing partial validation for our methodology, and theirs, the comparison will give us the chance to reappreciate the role of frequency-dependent effects on traveltime measurements (Nolet & Moser 1993; Baig *et al.* 2003; Baig & Dahlen 2004), which will help formulate a roadmap towards their use in tomographic inversions. We refer to Section 4.5, the discussion of Fig. 12, and Appendix A, as a reminder that we updated our measurements by adding a traveltime adjustment to our raw traveltime anomalies to maximize the fairness of the comparison.

Fig. 16 shows the comparison of our adjusted traveltime anomalies with those from Simon *et al.* (2022), both in unnormalized (Fig. 16a) and in relative (Fig. 16b) terms. Positive residuals signify observations that are slow compared to their prediction in the reference 1-D model. Compared to the complete data set whose statistics we revealed in Section 5, Fig. 15, we impose three cut-off criteria for the plotting. We only show traveltime anomaly measurement pairs with cross-correlation coefficients 0.6 or greater, SNRs above 15, and great-circle distances exceeding 20°. High correlation coefficients and large SNRs ensure that we maintain focus on high-quality, well-matched earthquake signals. The minimum epicentral distance is maintained to make sure that the *P*-to-acoustic conversion points at the ocean bottom are not unduly sensitive to hypocentre location errors and the incidence angles are high, such that the predicted arrival times have low uncertainty. Using these selection criteria leaves 423 measurements out of 1129 available for analysis. Shallow events with hypocentral depths above 33 km are most prone to phase misidentification by the automated method of Simon *et al.* (2022) and are rendered in grey to mark them as outliers which do not form part of the subsequent statistical analysis.

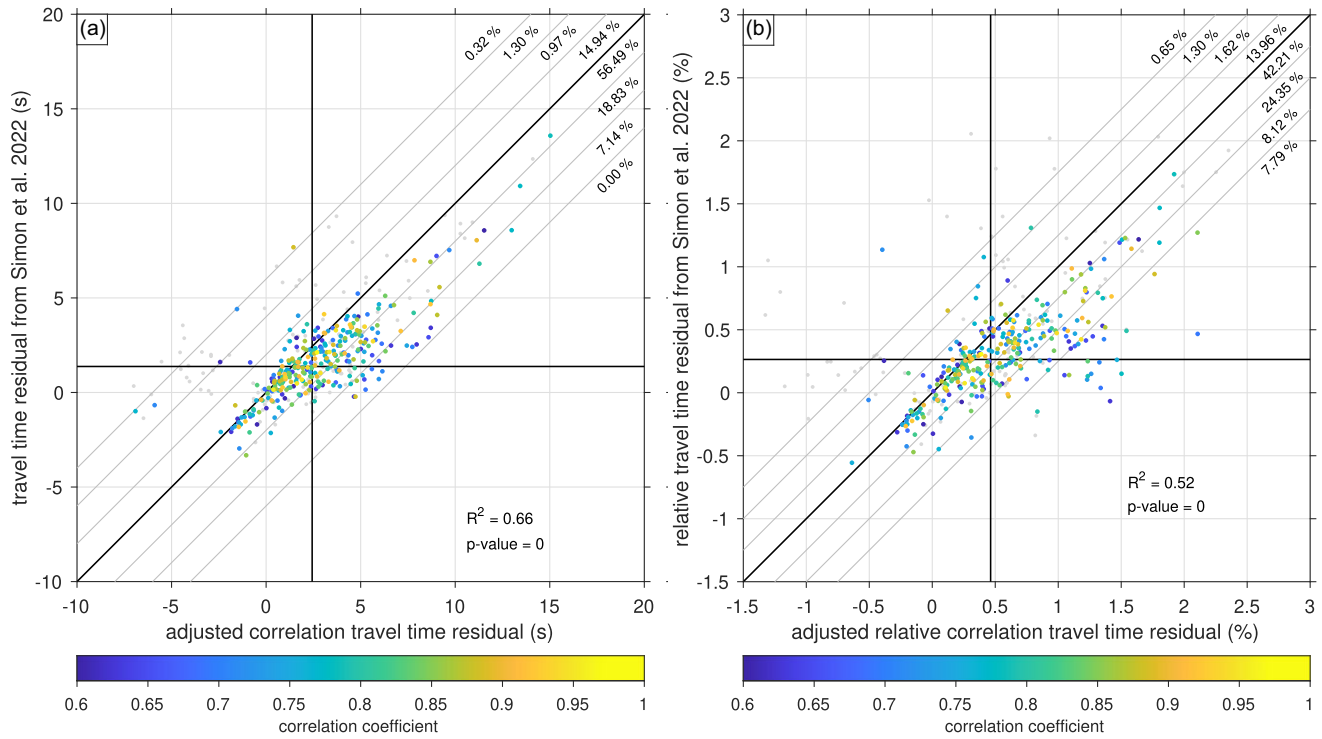

Figure 16. Cross-correlation-based traveltime measurements as made by our method compared to the arrival-time estimates obtained by Simon *et al.* (2022) via the method of Simon *et al.* (2020), for unnormalized (*a*) and relative (*b*) measurements. The solid black vertical and horizontal lines are drawn at the medians. Colours express the correlation coefficient (>0.60) between filtered observations and synthetic waveforms obtained using the methods introduced in this paper. Only data with a signal-to-noise ratio >15 and an epicentral distance > 20° are plotted here, which leaves 423 measurements out of 1129. Grey filled circles correspond to shallow earthquakes and are excluded from all statistical analysis. Diagonals mark the line of parity and define intervals where the two sets of values are within increments of 2 s, or 0.25 per cent of one another, with the proportion of values falling into these bins indicated by the numbers in the top right corner of the plots, in the quadrant where the actual 3-D Earth is slow compared to the 1-D reference.

The colours in Fig. 16 scale with the correlation coefficient of the waveform fits. Away from the main diagonal (solid black), the grey diagonals demarcate every subsequent ±2 s (in Fig. 16a) or 0.25 per cent (in Fig. 16b) offset increment in (relative) traveltime anomaly. The percentages of the data that fall between and beyond those markers are written as numbers in the top right corner of each panel. The majority of the data cluster within ±2 s from the 1:1 line, with a statistically significant ($p \approx 0 \ll 0.05$) linear correlation coefficient of 0.66 and 0.52, respectively. Hence our adjusted, variably bandpassed, cross-correlation-based traveltime measurements can be deemed close to those reported by Simon *et al.* (2022). Most of the pairs in the cluster lie in the upper right quadrant, suggesting that both methods suggest the *P*-wave speed in the actual 3-D Earth is slower than predicted from 1-D radial velocity variations alone. While this is broadly in line with expectations for the Pacific region under consideration (see Fig. 1) based on prior 3-D models (e.g. Cottaar & Lekić 2016), both our new measurements and those of Simon *et al.* (2022) are part of an ongoing effort to conduct a full-fledged tomographic inversion (as Nolet *et al.* 2019) for the South-Pacific mantle, a geologically and geodynamically intriguing target for geophysical study (Wamba *et al.* 2024).

As to why the correlation-based traveltime anomalies are generally greater than the ray-theoretical ones, forthcoming explanations will need to consider full-frequency wave propagation in a heterogeneous, anelastic, Earth. There are hints in our data that measurements made on records from the most distant earthquakes (> 100°) are skewed somewhat differently, towards smaller cross-correlation traveltime anomalies compared to ray-theoretical ones. No clear-cut straightforward relation with frequency content has emerged from our analysis so far.

Phase misidentification in the automated method of Simon *et al.* (2022) is a concern that is only partially alleviated by the depth selection as a means of outlier rejection. Depending on focal mechanism and radiation pattern, high-amplitude *pP* phases could follow low-amplitude *P* phases that could trigger variance-ratio based arrival-time algorithms, arriving within a few seconds of the *P* phase. On the other hand, our own waveform correlation measurements are not immune to cycle skips due to the presence of secondary maxima in the correlation function. Efforts to fully automate any measurement strategy will continue to be confronted with the need for visual inspection and cross-checking across record sections, and, where possible, comparison with nearby stations. This effort is underway.

Future possible improvements to our workflow might include making multifrequency, multiscale, envelope-based waveform correlation measurements (Yuan & Simons 2014; Simon *et al.* 2020; Liu *et al.* 2024), denoising of the hydrophone records, expanding the frequency range for forward modelling (now limited to 2 Hz), incorporating more detailed crustal models into the simulation, and embracing fully 3-D hybrid methods. As a final note, while the chosen example of Fig. 11 shows great consistency, as was the case with many of the records that we examined (see Supporting Information), all of our measurements were made on individual earthquake–MERMAID pairs, without any consideration of the collection of MERMAID instruments as an *array*, which will form the basis of future work.

7 CONCLUSIONS

We have proposed a new computational workflow for the analysis and modelling of hydroacoustic pressure waveforms due to distant earthquakes recorded, and autonomously reported, by MERMAID instruments at (typically) 1500 m depth in the open ocean. Simulating the full response of the Earth at the relatively high frequencies (0.5–5 Hz) that MERMAID is sensitive to requires the ability to consider elastic wave propagation inside of the three-dimensionally heterogeneous solid Earth and its coupling to the seismo-acoustic wavefield within the heterogeneous fluid oceanic envelope that overlies its variable bathymetry. As this goal remains out of reach for all current computational routines known today, our alternative, hybrid, method relies on a combination of approximate 1-D wave propagation (using AxiSEM/Instaseis) for the teleseismic, elastic, mantle part and uses highly accurate spectral-element modelling (via `SPECFEM2D`) in order to implement the displacement-to-pressure response of a realistic near-surface crust and ocean layer in the vicinity of the mid-column floating recorder.

At the time MERMAID was designed (Simons *et al.* 2006, 2009), it was not yet clear whether such waveforms would be faithfully recorded, let alone whether they could be computationally modelled. We devised an optimal frequency-band selection method that relatively amplifies the signal of impulsive earthquake arrivals to enhance the SNR. For a variably bandpassed test data set of 1129 high-fidelity MERMAID records from 235 unique earthquakes collected by 16 Princeton MERMAID floats between August 2018 and June 2021, we created a corresponding set of synthetic pressure waveforms with very high similarity to the observations.

Determining traveltime anomalies via cross-correlation yielded a new catalogue of finite-frequency measurements that we compared to the Simon *et al.* (2022) catalogue, which was based on high-frequency phase arrival-time estimation in conjunction with ray-theoretical traveltime modelling. Both types of measurements are in relatively good agreement, an important validation of both our modelling procedure and that of Simon *et al.* (2020) which formed the basis of the Simon *et al.* (2022) data sets. On the other hand, fully understanding the details of the comparison, the nature of the remaining discrepancy, and any possible systematic causes remain the subjects of ongoing work. Both types of traveltime anomaly modelling and determination will be cross-validated with more traditional methods of seismological analysis.

On average, each of the ∼65 MERMAIDs deployed worldwide surfaces weekly to report new earthquake waveforms within hours of recording. Having demonstrated that their waveforms can be modelled using a combination of elastic-wave forward simulations in a 1-D Earth and a 2-D spectral-element solver to implement the effect of the ocean layer for the last few kilometres of seismo-acoustic wave propagation, our new workflow will allow us to fully realize the promise of MERMAID filling in the coverage gap in global seismic tomography, including via full-waveform inversion.

ACKNOWLEDGMENTS

This work was sponsored by the Royal Thai Government, by Princeton University and by the U.S. National Science Foundation under grant EAR-2341811 to FJS. We thank Jessica C.E. Irving, Guust Nolet, Joel D. Simon and Mathurin Dongmo Wamba for input and advice, and William Eaton for comments on an early version of the paper. We acknowledge the constructive feedback from the Associate Editor, Huajian Yao and appreciate the comments from James Hammond and a second, anonymous, reviewer.

SUPPORTING INFORMATION

Supplementary data are available at *GJIRAS* online.

Pipatprathanporn+2024_SM1.pdf

Pipatprathanporn+2024_SM2.pdf

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

DATA AVAILABILITY

MERMAID seismograms, instrument responses and metadata are deposited with the EarthScope Data Management Center (use FDSN station code MH). Additional information can be found on http://www.earthscopeoceans.org. Green's functions and Earth models are available from the EarthScope (IRIS) Syngine service. Earthquake moment tensors are available from the Global CMT project, which relies on the Global Seismographic Network (GSN). Code used for the waveform modelling workflow and analysis is available from the first author's version-controlled GitHub repository, on the Web at https://github.com/sirpipat/MERMAID_waveform.

REFERENCES

Baig, A.M. & Dahlen, F.A., 2004. Traveltime biases in random media and the *S*-wave discrepancy, *Geophys. J. Int.,* **158**(3), 922–938.

Baig, A.M., Dahlen, F.A. & Hung, S.-H., 2003. Traveltimes of waves in three-dimensional random media, *Geophys. J. Int.,* **153**(2), 467–482.

Bendat, J.S. & Piersol, A.G., 2010. *Random Data: Analysis and Measurement Procedures,* 4th edn, John Wiley.

Bohnenstiehl, D.R., Tolstoy, M., Dziak, R.P., Fox, C.G. & Smith, D.K., 2002. Aftershock sequences in the mid-ocean ridge environment: an analysis using hydroacoustic data, *Tectonophysics,* **354,** 49–70.

Bottero, A., Cristini, P. & Komatitsch, D., 2020. On the influence of slopes, source, seabed and water column properties on T waves: Generation at shore, *Pure appl.Geophys.,* **177,** 5695–5711.

Capdeville, Y., Chaljub, E. & Montagner, J.-P., 2003. Coupling the spectral element method with a modal solution for elastic wave propagation in global earth models, *Geophys. J. Int.,* **152**(1), 34–67.

Chave, A.D., Thomson, D.J. & Ander, M.E., 1987. On the robust estimation of power spectra, coherences, and transfer functions, *J. geophys. Res.,* **92**(B1), 633–648.

Collins, J.A., Vernon, F.L., Orcutt, J.A., Stephen, R.A., Peal, K.R., Wooding, F.B., Spiess, F.N. & Hildebrand, J.A., 2001. Broadband seismology in the oceans: lessons from the Ocean Seismic Network Pilot Experiment, *Geophys. Res. Lett.,* **28**(1), 49–52.

Cottaar, S. & Lekić, V., 2016. Morphology of seismically slow lower-mantle structures, *Geophys. J. Int.,* **207**(2), 1122–1136.

Cristini, P. & Komatitsch, D., 2012. Some illustrative examples of the use of a spectral-element method in ocean acoustics, *J. acoust. Soc. Am.,* **131**(3), EL229–EL235.

Crotwell, H.P., Owens, T.J. & Ritsema, J., 1999. The TauP Toolkit: flexible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.,* **70**(2), 154–160.

Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology,* Princeton Univ. Press.

Dahlen, F.A., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes — I. Theory, *Geophys. J. Int.,* **141**(1), 157–174.

Dougherty, M.E. & Stephen, R.A., 1991. Seismo/acoustic propagation through rough seafloors, *J. acoust. Soc. Am.,* **90**(5), 2637–2651.

Dziak, R.P. *et al.*, 2004. *P*- and *T*-wave detection thresholds, *Pn* velocity estimate, and detection of lower mantle and core *P*-waves on ocean sound-channel hydrophones at the Mid-Atlantic Ridge, *Bull. seism. Soc. Am.,* **94**(2), 665–677.

Ekström, G., Nettles, M. & Dziewonski, A.M., 2012. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes, *Phys. Earth planet. Inter.,* **200–201,** 1–9.

Fernando, B., Leng, K. & Nissen-Meyer, T., 2020. Oceanic high-frequency global seismic wave propagation with realistic bathymetry, *Geophys. J. Int.,* **222**(2), 1178–1194.

GEBCO Bathymetric Compilation Group, 2019. The GEBCO_2019 grid—a continuous terrain model of the global oceans and land, Tech. rep., British Oceanographic Data Centre, National Oceanography Centre, NERC.

Gubbins, D., 2004. *Time Series Analysis and Inverse Theory for Geophysicists,* Cambridge Univ. Press.

Hammond, J. O.S., England, R., Rawlinson, N., Curtis, A., Sigloch, K., Harmon, N. & Baptie, B., 2019. The future of passive seismic acquisition, *Astron. Geophys.,* **60**(2), 2.37–2.42.

Hello, Y., Ogé, A., Sukhovich, A. & Nolet, G., 2011. Modern mermaids: new floats image the deep Earth, *EOS, Trans. Am. geophys. Un.,* **92**(40), 337–338.

Hung, S.-H., Dahlen, F.A. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes — II. Examples, *Geophys. J. Int.,* **141**(1), 175–203.

Hung, S.-H., Dahlen, F.A. & Nolet, G., 2001. Wavefront healing: a banana-doughnut perspective, *Geophys. J. Int.,* **146,** 289–312.

Jamet, G., Guennou, C., Guillon, L., Mazoyer, C. & Royer, J.-Y., 2013. *T*-wave generation and propagation: a comparison between data and spectral element modeling, *J. acoust. Soc. Am.,* **134**(4), 3376–3385.

Joubert, C., Nolet, G., Bonnieux, S., Deschamps, A., Dessa, J.-X. & Hello, Y., 2016. *P*-delays from floating seismometers (MERMAID), part I: data processing, *Seismol. Res. Lett.,* **87**(1), 73–80.

Kennett, B. L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from travel-times, *Geophys. J. Int.,* **122**(1), 108–124.

Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation — II. Three-dimensional models, oceans, rotation and self-gravitation, *Geophys. J. Int.,* **150**(1), 303–318.

Komatitsch, D. & Vilotte, J.P., 1998. The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. seism. Soc. Am.,* **88**(2), 368–392.

Komatitsch, D., Barnes, C. & Tromp, J., 2000. Wave propagation near a fluid-solid interface: a spectral-element approach, *Geophysics,* **65**(2), 623–631.

Lecoulant, J., Guennou, C., Guillon, L. & Royer, J.-Y., 2019. Three-dimensional modeling of earthquake generated acoustic waves in the ocean in simplified configurations, *J. Acoust. Soc. Am.,* **146**(3), 2113–2123.

Lei, W. *et al.*, 2020. Global adjoint tomography—model GLAD-M25, *Geophys. J. Int.,* **223**(1), 1–21.

Liu, Z., Hoffmann, J., Bachmann, E., Cui, C., Simons, F.J. & Tromp, J., 2024. Laplace-domain crosstalk-free source-encoded elastic full waveform inversion using time-domain solvers, *Geophysics,* **89**(4), 1–21.

Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics,* **56**(5), 654–663.

Lyu, C., Zhao, L. & Capdeville, Y., 2022. Novel hybrid numerical simulation of the wave equation by combining physical and numerical representation theorems and a review of hybrid methodologies, *J. geophys. Res.,* **127**(5), e2021JB022368.

Marra, G. *et al.*, 2018. Ultrastable laser interferometry for earthquake detection with terrestrial and submarine cables, *Science,* **361**(6401), 486–490.

Masson, Y. & Romanowicz, B., 2017. Box tomography: localized imaging of remote targets buried in an unknown medium, a step forward for understanding key structures in the deep earth, *Geophys. J. Int.,* **211**(1), 141–163.

Mercerat, E.D. & Nolet, G., 2012. Comparison of ray- and adjoint-based sensitivity kernels for body-wave seismic tomography, *Geophys. Res. Lett.,* **39**(12), L12301.

Mercerat, E.D. & Nolet, G., 2013. On the linearity of cross-correlation delay times in finite-frequency tomography, *Geophys. J. Int.,* **192,** 681–687.

Monteiller, V., Chevrot, S., Komatitsch, D. & Fuji, N., 2013. A hybrid method to compute short-period synthetic seismograms of teleseismic body waves in a 3-D regional model, *Geophys. J. Int.,* **192**(1), 230–247.

Nakata, N., Gualtieri, L. & Fichtner, A., 2019. *Seismic Ambient Noise,* Cambridge Univ. Press.

Newmark, N.M., 1959. A method of computation for structural dynamics, *J. Eng. Mech. Div.–ASCE,* **85**(3), 67–94.

Nissen-Meyer, T., van Driel, M., Stähler, S.C., Hosseini, K., Hempel, S., Auer, L., Colombi, A. & Fournier, A., 2014. AxiSEM: broadband 3-D seismic wavefields in axisymmetric media, *Solid Earth,* **5**(1), 425–445.

Nolet, G. & Dahlen, F.A., 2000. Wave front healing and the evolution of seismic delay times, *J. geophys. Res.,* **105**(B8), 19043–19054.

Nolet, G. & Moser, T.J., 1993. Teleseismic delay times in a 3-D earth and a new look at the *S* discrepancy, *Geophys. J. Int.,* **114**(1), 185–195.

Nolet, G. *et al.*, 2019. Imaging the Galápagos mantle plume with an unconventional application of floating seismometers, *Sci. Rep.,* **9,** 1326.

Nolet, G., Simon, J.D. & Bonnieux, S., 2024. How accurately are MERMAID seismograms located?, *Seismol. Res. Lett.,* **95**(4), 2368–2374.

Okamoto, T., 1994. Teleseismic synthetics obtained from 3-D calculations in 2-D media, *Geophys. J. Int.,* **118**(3), 613–622.

Pipatprathanporn, S. & Simons, F.J., 2022. One year of sound recorded by a MERMAID float in the Pacific: hydroacoustic earthquake signals and infrasonic ambient noise, *Geophys. J. Int.,* **228**(1), 193–212.

Rawlinson, N., Fichtner, A., Sambridge, M. & Young, M.K., 2014. Seismic tomography and the assessment of uncertainty, *Adv. Geophys.,* **55,** 1–76.

Ricker, N., 1940. The form and nature of seismic waves and the structure of seismograms, *Geophysics,* **5**(4), 348–366.

Robertsson, J. O.A., Levander, A. & Holliger, K., 1996. A hybrid wave propogation simulation technique for ocean acoustic problems, *J. geophys. Res.,* **101**(B5), 11225–11241.

Romanowicz, B., 2003. Global mantle tomography: progress status in the last 10 years, *Annu. Rev. geophys. Space Phys.,* **31,** 303–328.

Romanowicz, B., 2008. Using seismic waves to image Earth's structure, *Nature,* **451,** 266–268.

Scherbaum, F., 2001. *Of Poles and Zeros: Fundamentals of Digital Seismology,* 2nd edn, Kluwer.

Shannon, C.E., 1949. Communication in the presence of noise, *Proc. IRE,* **37**(1), 10–21.

Simon, J.D., Simons, F.J. & Nolet, G., 2020. Multiscale estimation of event arrival times and their uncertainties in hydroacoustic records from autonomous oceanic floats, *Bull. seism. Soc. Am.,* **110**(3), 970–997.

Simon, J.D., Simons, F.J. & Irving, J. C.E., 2021. A MERMAID miscellany: seismoacoustic signals beyond the *P* wave, *Seismol. Res. Lett.,* **92**(6), 3657–3667.

Simon, J.D., Simons, F.J. & Irving, J. C.E., 2022. Recording earthquakes for tomographic imaging of the mantle beneath the South Pacific by autonomous MERMAID floats, *Geophys. J. Int.,* **228,** 147–170.

Simons, F.J., 2010. Slepian functions and their use in signal estimation and spectral analysis, in *Handbook of Geomathematics,* Chapter 30, pp. 891–923, eds Freeden, W., Nashed, M.Z. & Sonar, T., Springer.

Simons, F.J., Nolet, G., Babcock, J.M., Davis, R.E. & Orcutt, J.A., 2006. A future for drifting seismic networks, *EOS, Trans. Am. geophys. Un.,* **87**(31), 305–307.

Simons, F.J., Nolet, G., Georgief, P., Babcock, J.M., Regier, L.A. & Davis, R.E., 2009. On the potential of recording earthquakes for global seismic tomography by low-cost autonomous instruments in the oceans, *J. geophys. Res.,* **114**(B5), B05307.

Simons, F.J., Simon, J.D. & Pipatprathanporn, S., 2021. Twenty-thousand leagues under the sea: recording earthquakes with autonomous floats, *Acoust. Today,* **17**(2), 42–51.

Slack, P.D., Fox, C.G. & Dziak, R.P., 1999. *P* wave detection thresholds, P_n velocity estimates, and *T* wave location uncertainty from oceanic hydrophones, *J. geophys. Res.,* **104**(B6), 13061–13072.

Sladen, A., Rivet, D., Ampuero, J.P., Barros, L.D., Hello, Y., Calbris, G. & Lamare, P., 2019. Distributed sensing of earthquakes and ocean-solid Earth interactions on seafloor telecom cables, *Nat. Commun.,* **10,** 5777.

Stacey, R., 1988. Improved transparent boundary formulations for the elastic-wave equation, *Bull. seism. Soc. Am.,* **78**(6), 2089–2097.

Stephen, R.A., Spiess, F.N., Collins, J.A., Hildebrand, J.A., Orcutt, J.A., Peal, K.R., Vernon, F.L. & Wooding, F.B., 2003. Ocean seismic network pilot experiment, *Geochem. Geophys. Geosys.,* **4**(10), 1092.

Suetsugu, D. & Shiobara, H., 2014. Broadband ocean-bottom seismology, *Annu. Rev. Earth. planet. Sci.,* **42,** 27–43.

Sukhovich, A., Irisson, J.-O., Simons, F.J., Ogé, A., Hello, Y.M., Deschamps, A. & Nolet, G., 2011. Automatic discrimination of underwater acoustic signals generated by teleseismic *P*-waves: a probabilistic approach, *Geophys. Res. Lett.,* **38**(18), L18605.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.

Tong, P., Chen, C., Komatitsch, D., Basini, P. & Liu, Q., 2014. High-resolution seismic array imaging based on an SEM-FK hybrid method, *Geophys. J. Int.,* **197**(1), 369–395.

Tromp, J., 2020. Seismic wavefield imaging of Earth's interior across scales, *Nat. Rev. Earth Environ.,* **1,** 40–53.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160**(1), 195–216.

van Driel, M., Krischer, L., Stähler, S.C., Hosseini, K. & Nissen-Meyer, T., 2015. Instaseis: instant global seismograms based on a broadband waveform database, *Solid Earth,* **6**(2), 701–717.

Wamba, M.D., Simons, F.J. & Irving, J.C.E., 2024. Data-space cross-validation of mantle structure in global tomographic models underneath the Pacific Ocean, *Geophys. J. Int.,* (in process)

Wang, Y., 2015. Frequencies of the Ricker wavelet, *Geophysics,* **80**(2), A31–A37.

Williams, E.F., Fernández-Ruiz, M.R., Magalhaes, R., Vanthillo, R., Zhan, Z., González-Herráez, M. & Martins, H.F., 2019. Distributed sensing of microseisms and teleseisms with submarine dark fibers, *Nat. Commun.,* **10**(1), 5778.

Wu, W., Ni, S., Zhan, Z. & Wei, S., 2018. An SEM-DSM three-dimensional hybrid method for modelling teleseismic waves with complicated source-side structures, *Geophys. J. Int.,* **215**(1), 133–154.

Yu, Y., Chen, Y.J., Guo, Z. & Ge, Z., 2023. Long-term seismic network in South China Sea by floating MERMAIDs, *Sci. China Earth Sci.,* **66,** 1979–1993.

Yuan, Y.O. & Simons, F.J., 2014. Multiscale adjoint waveform-difference tomography using wavelets, *Geophysics,* **79**(3), WA79–WA95.

Yuan, Y.O., Bozdağ, E., Ciardelli, C., Gao, F. & Simons, F.J., 2019. The exponentiated phase measurement, and objective-function hybridization for adjoint waveform tomography, *Geophys. J. Int.,* **221**(2), 1145–1164.

Zhang, Z., Irving, J. C.E., Simons, F.J. & Alkhalifah, T., 2023. Seismic evidence for a 100 km mantle discontinuity under the Pacific, *Nat. Commun.,* **14,** 1714.

APPENDIX A: TauP – INSTASEIS CORRECTION

As mentioned in Section 4.5 there is a difference (green arrow in Fig. 12) between the point an analyst would pick as an arrival on a seismogram (inverted green triangle in Fig. 12) calculated by Instaseis (van Driel *et al.* 2015), and the ray-theoretical prediction (green triangle in Fig. 12) obtained by TauP (Crotwell *et al.* 1999), even when both correspond to the same 1-D earth model (e.g., `ak135` Kennett *et al.* 1995). When comparing the high-frequency picks and ray-theory based traveltime anomalies from Simon *et al.* (2022) with the cross-correlation and finite-frequency modelling based traveltime anomalies that we obtain in this paper, we need to adjust for that difference.

To pick the arrival time on an Instaseis seismogram we first identify the peak of the main *P*-wave arrival. This can be achieved by finding the absolute maximum of the seismogram in the 30 s window centred at the ray-theoretical arrival-time prediction. Since the simulation is noise-free, the displacement is numerically zero prior to the first arrival. Therefore, we define the first-arrival time as the first instance when the absolute displacement exceeds 2 per cent of the absolute maximum determined in that time window.

Fig. A1 shows an example. Fig A1(a) shows the vertical displacement seismogram produced by Instaseis in model `ak135` for the *M* 6.5 2018 event in Indonesia used throughout this paper, at a hypothetical ocean-bottom station beneath a MERMAID float located at 70.7°. Red lines with phase names mark the TauP ray traced arrival times. Figs A1(b) and (c) show 30 s windows centred at the *P*- and *S*-wave arrivals. The picked *P*-arrival time is marked as a blue vertical line in Fig. A1(b). The displacement at the time pick is 2 per cent of the absolute maximum displacement in this window, which occurs just right of 0 s. No picks are made on the *S*-wave arrival. Since the time axis is expressed relative to the ray-traced arrival time in this plot, the picked arrival time equals the correction time by which we adjust our measurements for comparison with the ray-theoretical results.

Fig. A2 shows a summary of the traveltime corrections made on all of our 1129 measurements. Fig. A2(a) shows the distribution of traveltime corrections. Excluding 8 outliers, almost all, that is, 1121 of them, required traveltime corrections between −5 and 1 s, with a median −1.96 s. The relation of these corrections with event depth, epicentral distance and event magnitude is explored in Figs A2(b)–(d). Traveltime corrections generally fall between −3 and −1 s for most cases, except for some shallow, nearby or small-magnitude events where the correction can be as negative as −5 s.

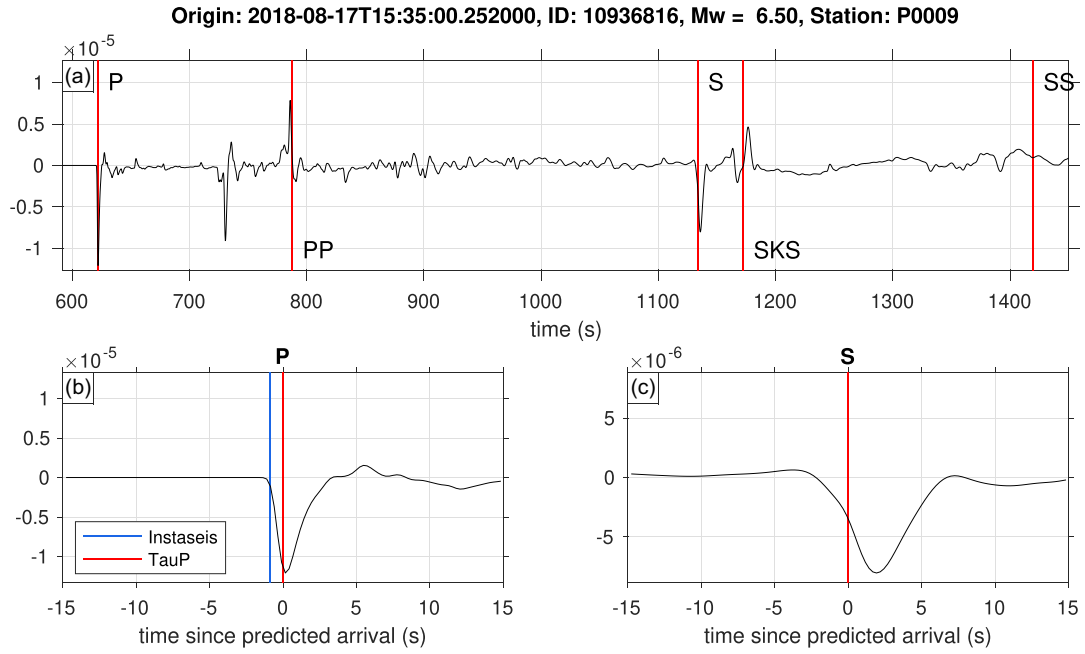

Figure A1. A synthetic vertical displacement seismogram produced by Instaseis for the M 6.5 event in Indonesia arriving at a station beneath a MERMAID float at 70.7°, computed within 1-D earth model `ak135`. (a) Seismogram containing the first P-wave arrival through the later arriving S waves. Red lines with phase names mark the ray-traced arrival times calculated by TauP using the same earth model. (b) Zoom centred on the P-wave arrival. The blue vertical line is the picked arrival time. Since the seismogram is plotted relative to the TauP arrival time, the timing of the blue line coincides with the traveltime adjustment, which is the adjustment made to our measurements to facilitate the comparison in Fig. 16 with the ray-theoretical results obtained by Simon *et al.* (2022). (c) Zoom centred on the S-wave arrival, where no measurements are made.

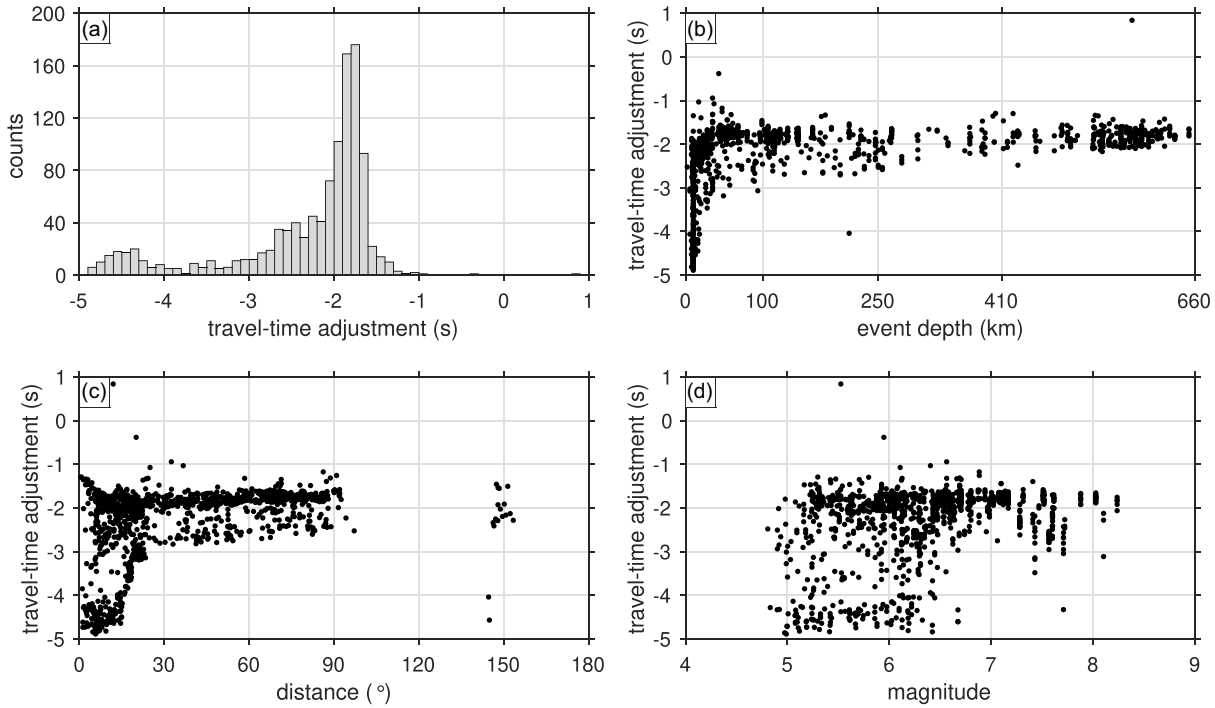

Figure A2. Summary of the applied traveltime adjustment for all 1129 measurements discussed in this paper. (a) Histogram, and scatter plots of the traveltime adjustment against (b) event depth, (c) epicentral distance, and (d) event magnitude. Shallow and nearby events show the largest spread, and most extremely negative, traveltime adjustments.