Supplemental Material:

Recording earthquakes for tomographic imaging of the mantle beneath the South Pacific by autonomous MERMAID floats

Joel D. Simon[1], Frederik J. Simons[1] and Jessica C. E. Irving[2]

[1]*Department of Geosciences, Princeton University, Princeton, NJ 08544, USA, E-mail: jdsimon@alumni.princeton.edu*
[2]*School of Earth Sciences, University of Bristol, Bristol, BS8 1RJ, UK*

S1 THE EARTHSCOPE-OCEANS CONSORTIUM

The EarthScope-Oceans consortium was founded in 2016, and now counts members from the US (Princeton and Stanford University, Incorporated Research Institutions for Seismology [IRIS] Seattle, DBV Technology North Kingstown, RI), Japan (Kobe University, Japan Agency for Marine-Earth Science and Technology [JAMSTEC], Earthquake Research Institute [ERI]), France (Géoazur Sophia Antipolis, École et Observatoire des Sciences de la Terre [EOST] Strasbourg, Institut Français de Recherche pour l'Exploitation de la Mer [IFREMER] Plouzané, OSEAN SAS Le Pradet), South Korea (Korea Institute of Geoscience and Mineral Resources [KIGAM] Daejeon), New Zealand (GNS Science, Te Pū Ao, Lower Hutt), the UK (Universities of Oxford and Bristol), and China (Southern University of Science and Technology [SUSTech], Shenzhen).

EarthScope-Oceans (`http://earthscopeoceans.org/`) represents a multidisciplinary group of geoscientists who are coordinating efforts to create a global network of sensors to monitor the Earth system from within the oceanic environment. It intends to shepherd national projects into the international forum where globally relevant, applicable, and mutually agreed-upon decisions can be made on technological aspects of instrument development, science objectives and priorities on different time scales, data management, dissemination, archiving, and education and outreach efforts; much like IRIS (`https://iris.edu/`) or Observatories and Research Facilities for European Seismology (ORFEUS; `http://orfeus-eu.org/`) are doing for the land-based seismological communities today.

S2 THE DEPLOYMENT OF THE SPPIM ARRAY

A 24-hr trial run completed 12 April 2018 was led by Kobe University's Hiroko Sugioka and JAMSTEC's Masayuki Obayashi from the R/V *Fukae Maru*. During this test deployment MERMAID N0003 recorded an m_b 4.9 earthquake originating at 59.4±5.8 km depth, 56±6.6 km east of Ishinomaki, Japan (according to `https://earthquake.usgs.gov/earthquakes/eventpage/us2000dyw3/`), some 824 km distance from the instrument, which floated 469 m below the surface at the time.

The SPPIM array was deployed over several cruises led by Yann Hello, research engineer at the French Institut de Recherche pour le Développement (IRD)/Géoazur, chief designer of MERMAID in its current third generation, as commercially available from OSEAN SAS of Le Pradet, France. On the first leg (Nouméa, New Caledonia to Mata-Utu, Wallis & Futuna, 21–28 June 2018), Yann Hello deployed two Géoazur units from the IRD/Genavir vessel R/V *Alis*. On the second leg (Mata-Utu to Papeete, Tahiti, French Polynesia, 3–13 August 2018), Hello deployed five Princeton units from the R/V *Alis*. On the third leg, Frederik Simons deployed 11 Princeton units from the R/V *Alis*, which departed Papeete on 28 August 2018, returning to the same port on 16 September of the same year (`https://campagnes.flotteoceanographique.fr/campagnes/18000519/`, DOI: 10.17600/18000519). Note that no instruments named P0014 or P0015 were ever deployed. Five Japanese units were launched from the R/V *Mirai* by Masayuki Obayashi, sailing from Shimizu, Japan to Valparaiso, Chile between 11 December 2018 and 24 January 2019. The fourth leg (Papeete–Nouméa, 4–29 August 2019) was led by Hello, Obayashi, Zhen Guo, and Yong Yu (SUSTech) from the R/V *L'Atalante* (`https://campagnes.flotteoceanographique.fr/campagnes/18000882/`, DOI: 10.17600/18000882). This cruise saw the completion of the SPPIM array with the deployment 23 SUSTech MERMAIDs and an additional four from Kobe University.

S3 MATCHING MERMAID SEISMOGRAMS TO EARTHQUAKES

S3.1 *Automated Preliminary Matching*

Upon receipt of a fresh seismogram transmitted by MERMAID we immediately wish to determine whether or not the signals it contains correspond to known seismic events. To that end we developed a complete workflow executed in MATLAB to match untagged, raw seismograms to global seismic catalogs with minimal user intervention. This first step discussed next—the algorithmic querying of global catalogs, the

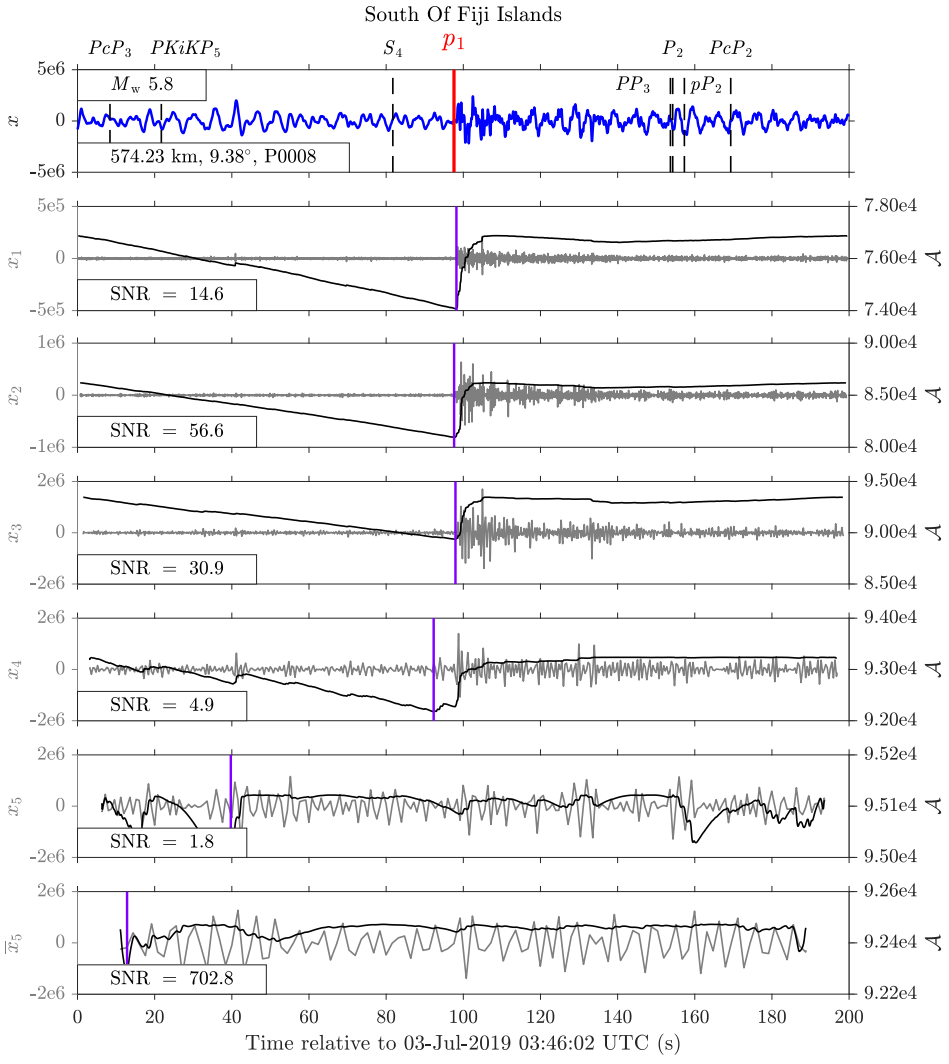

Figure S1. MERMAID seismogram after automatic preliminary matching. The blue trace in the top panel is the raw seismogram, while the gray traces below are wavelet-subspace projections at five scales, each overlain by their associated Akaike information criterion (AIC) curve (black) and AIC-based arrival-time pick (purple). The top panel is annotated with the theoretical arrival times of various phases from five distinct earthquakes, as noted in the subscripts, computed in the ak135 velocity model, and marked in time by vertical lines. These represent all the phases which have theoretical arrival times within the time window of the seismogram, associated with known global seismic events in the catalogs queried from IRIS. The time of the first-arriving phase associated with the largest earthquake in the set (p_1) is marked by a solid red vertical line. Its theoretical arrival time agrees well with the AIC-based arrival-time picks (which are independent of seismology) at the first three scales. The agreement of these two distinct arrival-time estimation methods lends itself to the confident assignment of this seismogram to the "identified" category. During manual review this figure (and a secondary, zoomed in version) is displayed to the researcher, along with the event metadata associated with all potentially matching events, and the researcher is led through a series of intuitive prompts in MATLAB for easy matching and sorting.

tagging of likely events, the annotation of seismograms with their theoretical phase arrival times, and the multiscale detection of phases against which residuals are displayed—occurs automatically and without user intervention after a MERMAID transmits a new seismogram.

The preliminary matching process begins with the querying of global seismic catalogs with `irisFetch.m` (`https://github.com/iris-edu/irisFetch-matlab/`), a software packaged and distributed by IRIS, for seismic events that occurred in the hour preceding the seismogram. Next, travel times are computed for seismic body waves that are likely to be present in the record using `taupTime.m` in the `MatTaup` package (see Data Availability and Resources in the main text) for the ak135 velocity model. Each event with one or more phase-arrivals in the time window of the seismogram is deemed a preliminary match, and all such events are sorted by magnitude (generally the single greatest factor determining phase identification) and saved together as individual structures (a MATLAB data type) in a binary (*.mat) "unreviewed" file.

Two figures display the raw seismogram at the top, upon which the theoretical phase-arrival times of the possible events are marked. The panels below plot the wavelet-subspace projections of the seismogram at five scales, each annotated with their arrival-time pick estimated using an Akaike information criterion (Simon et al. 2020). The first plot, Fig. S1, displays the complete seismogram, and the second (not shown here) shows detail about a 100 s window centered on the first arrival of the event with the largest magnitude among all potential

matches. Usually that is the true match, and thus the boxes within the top panel of Fig. S1 quote those metadata and the recording MERMAID (in this case, P0008). Its theoretical first-arriving phase is highlighted by a solid red line. All other possible phases in the time window of interest are rendered as dashed black lines. All named phases are labeled with subscripts identifying the rank of the associated event in the magnitude-sorted preliminary match list. In Fig. S1, p_1 marks the theoretical arrival time of a *p* wave generated by the first preliminary match, and, S_4 is the theoretical arrival time of an *S* wave from the fourth possible event match.

These preliminary matches are automatically generated and the algorithm only requires a SAC file (Helffrich et al. 2013) as input; i.e., the only relevant information ingested by the algorithm in this preliminary-matching stage is a (mobile) receiver location and a time window. Hence, our procedure is not specific to MERMAID data, and we may reasonably assume that it has application beyond the scope of this study. For example, for single-station or array deployments of traditional broad-band land instruments, perhaps in the context of Comprehensive Nuclear Test-Ban-Treaty (CTBT) verification, Raspberry Shakes (Bent et al. 2018; Anthony et al. 2019; Calais et al. 2019) or various other forms of crowd-sourced "citizen" seismology, e.g., recorded by mobile phones (Kong et al. 2016) or other low-cost instruments (Cochran et al. 2009; Jeddi et al. 2020), and for classroom seismic installations (Balfour et al. 2014; Subedi et al. 2020), where an experienced researcher may not be available to guide the matching process. While the code is provided with default parameters optimized for MERMAID data, these are easily tunable.

S3.2 *Manual winnowing and sorting*

The second step of the matching procedure involves manual review to sort the seismograms into two classes: "identified" and "unidentified." Those in the former class will have been assessed to contain energies consistent with phase arrivals corresponding to known earthquakes in global seismic catalogs, both by visual inspection and by considering their travel-time residuals with respect to the AIC picks. For every SAC file reviewed, the two PDFs generated in the first step are automatically opened for inspection, and the interactive program guides the user through a series of prompts to determine if the event can be identified and, if so, which event(s) and phase(s) should be saved.

The process begins with a printout of metadata for all potential events. At all times the user has quick access to all events and their corresponding residuals thanks to their MATLAB structure variables being loaded automatically with each seismogram under review. We refer again to Fig. S1, whose top panel plots the raw seismogram in blue. The arrival times marked on that top panel are the ak135 predictions. The panels below the first plot the subspace projections of the seismogram at five wavelet scales in gray, the amplitude of which corresponds to the left ordinate axis. Overlain in black in each panel is the associated AIC curve, $\mathcal{A}$, used to generate the arrival-time pick at that scale, corresponding to the right ordinate axis. This curve is essentially an inverted likelihood-curve: where it is low, an arrival is likely. The specific AIC arrival-time pick is marked at each scale by a purple vertical line. Quoted in the inset boxes are the corresponding signal-to-noise ratios (SNRs), defined to be the ratio of the maximum-likelihood estimates of the variances of the signal and noise segments, the "signal" being the segment after the AIC pick, and the "noise" the segment preceding it. This definition of the SNR is the same as that in the main text, however, the signal and noise segments are not—there, we focus on the first arrival within a single frequency band (1–5 Hz) and a short time window (30 s); here we consider the complete time series ($\sim$200–300 s) at each subspace projection (x_1–$\bar{x}_5$) resulting in five SNRs per seismogram. There, the seismograms being analyzed have already been positively matched to an identified event, which differs from the procedure here, where we wish to inspect the full bandwidth of each seismogram via a wavelet multiscale decomposition.

The AIC-based picks are ignorant of seismology. Their agreement, or lack thereof, with the theoretical arrival times of the phases from the match list informs the decision to designate a seismogram as "(un)identified." In the case of Fig. S1, the purple AIC picks at subspace projection scales one (x_1) through three (x_3) agree well with the theoretical arrival time of the first-arriving *p* wave in ak135. The AIC picks at the other scales are either low-SNR or very near an edge and may be disregarded. We also note that the picks shown here are not influenced by the edges, whose treatment we describe in Simon et al. (2020), so users need not necessarily be wary of an arrival pick near an edge. However, we have noticed that our AIC picker will on occasion report an extremely short noise or signal segment associated with a time series that has no clear arrival. An example of this behavior is the noise segment in the last panel of Fig. S1, $\bar{x}_5$, which is extremely abbreviated, consistent with its having low variance, most unlike the variance of the signal segment. Therefore, a low-SNR signal that is very near an edge does warrant a close inspection. Regardless, because of the agreement between the theoretical arrival time of the *p* wave corresponding to the largest event and the AIC picks at low scales (high frequencies) in Fig. S1, this seismogram would be counted among the identified category. This sorting is accomplished via simple prompts that guide the user through a winnowing process that ultimately results in the seismogram being classified as identified or unidentified, and the relevant event data being saved to a binary (*.mat) "reviewed" file.

Ultimately the decision to mark a seismogram as identified or unidentified comes down to experience processing MERMAID seismograms like the one presented in Fig. S1. The hope, however, is that the workflow developed here is simple enough for new researchers with some experience processing seismic data to quickly grasp and apply it to their own untagged data with minimal training. Indeed, our workflow is already being successfully applied to the 23 SUSTech instruments included in the SPPIM deployment—albeit applied to the same type of data in this case but, importantly, matched by a different researcher.

S4 DESCRIPTION OF SUPPLEMENTAL TEXT FILES

S4.1 *First-arrival travel-time residuals in the MERMAID catalog:* `simon2021gji_supplement_residuals.txt`

We use `irisFetch.Events.m` version 2.0.10, one method of a larger software class written in MATLAB and distributed by IRIS, to query `http://service.iris.edu/fdsnws/event/1/` for event metadata archived at the IRIS Data Management Center (DMC; `http://ds.iris.edu/ds/nodes/dmc/`). In our supplementary text files we list only the "Preferred" metadata returned by `irisFetch.Events.m`, which we understand to be those metadata preferred by the National Earthquake Information Center (NEIC; `https://earthquake.usgs.gov/data/comcat/catalog/us/`, which compiles a global bulletin of earthquake metadata called the PDE for Preliminary Determination of Epicenters), at the time those metadata were reported to IRIS; i.e., they are not necessarily the most up-to-date or currently-preferred values in the NEIC PDE Bulletin. For reference, we list both the IRIS event identification number (`http://ds.iris.edu/ds/nodes/dmc/tools/event/<IRIS_ID>`), and an NEIC PDE event identification number (`https://earthquake.usgs.gov/earthquakes/eventpage/<NEIC_ID>`), and note that the latter may reference multiple origins and/or contributors whose metadata differ from what we print in our supplementary text files. Generally, the NEIC PDE event identification number we print is prefixed with "us" (DOI: 10.7914/SN/US), meaning that the preferred event origin metadata were contributed by the NEIC itself. We also include 16 prefixed with "ak" (DOI: 10.7914/SN/AK), and one prefixed with "ci" (DOI: 10.7914/SN/CI) in cases when it appears that those authors had contributed the preferred origin-metadata to the NEIC PDE at the time of archival in the DMC. Therefore, future researchers who utilize the supplemental text files compiled during this study are strongly advised to check for updates to the event metadata included here, especially once they are published in the International Seismological Centre (ISC) catalog (Bondár & Storchak 2011).

We use `MatTaup`, written in MATLAB by Qin Li at the University of Washington (dated November 2002 but without a version number), to compute theoretical travel times in the ak135 model (Kennett et al. 1995), and `LLNL-Earth3D` (`https://www-gs.llnl.gov/nuclear-threat-reduction/nuclear-explosion-monitoring/global-3d-seismic-tomography/`), to compute theoretical travel times in the LLNL-G3Dv3 model of Simmons et al. (2012). Note that `LLNL-Earth3D` provides two water corrections—one corresponding to the event, and one corresponding to the station—which are to be added to the travel times in cases when one (or both) appears to be in water due to the limits of the model resolution at that location, but one (or both) are known to be in or on solid rock. All event-side corrections are exactly 0 s in our case, meaning that the model properly located our events in solid rock. Our station-side corrections are not all zero, but they are also not included in the 3-D travel times reported in column 28 of the text file described next because MERMAID is known to have been in water at the time of recording. Ocean depths in column 12 are interpolated from GEBCO 2014 (Weatherall et al. 2015). MERMAID station latitude and longitude at the time of recording (around 1500 m depth) are interpolated using the method of Joubert et al. (2016) as codified in v3.4.0-Z of our Python package `automaid` (Simon et al. 2021). See the additional text files described in Section S4.3 for those raw GPS fixes.

The data file `simon2021gji_supplement_residuals.txt` contains travel-time residuals and other calculations relating only to first-arriving *p* and *P* phases identified in our data set. Note that all times are rounded to two decimal places. As such, columns that purport to be sums of other columns (e.g., column 24 = column 20 + column 23) may differ by 1/100 s, which is smaller than MERMAID's nominal sampling interval used in this study of 1/20 s. Note as well that we use the moniker "'travel" time to refer to the time difference between our AIC pick and the event time, and "arrival" time to refer to the time difference between the same AIC pick and the start of the seismogram.

We will use `20180808T014200.08_5B736FA6.MER.DET.WLT5.sac`, the first MERMAID SAC file quoted in the residuals text file, to explain our file-naming convention: "20180808T014200" is the UTC date of the first sample of the seismogram (truncated to integer seconds), in this case, 8 August 2018 01:42:00 UTC; "08" is the MERMAID serial number excluding the "P00" prefix; "5B736FA6.MER" names the corresponding MERMAID data file from which this SAC file was generated; "DET" means that this seismogram contains a signal detected by MERMAID's onboard algorithm, as opposed to one requested by us; and "WLT5" telegraphs that wavelet transform coefficient sets corresponding to five (out of a total of six) scales were transmitted in the .MER file (resulting in a seismogram with a nominal sampling frequency of 20 Hz after reconstruction via inverse wavelet transform because MERMAID's nominal sampling rate is 40 Hz).

The columns of `simon202gji_supplement_residuals.txt` are:

(1) SAC filename
(2) Event origin time [UTC]
(3) Event longitude [decimal degrees]
(4) Event latitude [decimal degrees]
(5) Event magnitude value (e.g., 8.1)
(6) Event magnitude type (e.g., M_{W} or m_{b})
(7) Event depth [m]
(8) Timestamp at first sample of seismogram (i.e., the reference time in the SAC header (NYZEAR, NZJDAY, ..., NZMSEC) plus any offset in the "B" header field) [UTC]
(9) Interpolated station longitude at time of recording using the method of Joubert et al. (2016) and codified by Simon et al. (2021) [decimal degrees]
(10) Interpolated station latitude at time of recording using the method of Joubert et al. (2016) and codified by Simon et al. (2021) [decimal degrees]
(11) Station depth at time of recording [m]
(12) Ocean depth at station location at time of recording according to GEBCO 2014 (Weatherall et al. 2015) [m]
(13) ak135 epicentral distance [decimal degrees]
(14) Difference between ak135 (adjusted for bathymetry and cruising depth) and ak135 epicentral distances (always 0) [decimal degrees]
(15) ak135 (adjusted for bathymetry and cruising depth) epicentral distance (always equal to column 13) [decimal degrees]
(16) Difference between LLNL-3DGv3 and ak135 epicentral distances (the former model is elliptical and the latter is spherical; = column 17 − column 13) [decimal degrees]
(17) LLNL-3DGv3 epicentral distance (= column 13 + column 16) [decimal degrees]
(18) Observed travel time using our AIC pick of Simon et al. (2020) and further described in the Appendix of the main text (think "seconds after the event"), t_{AIC} [s]
(19) Observed arrival time using our AIC pick of Simon et al. (2020) and further described in the Appendix of the main text (this is the same AIC pick of column 18, but its timing is in reference to the start time of the seismogram and not the origin time of the event; think "seconds into the seismogram") [s]
(20) Predicted travel time in ak135, t_{ak135} [s]
(21) Predicted arrival time in ak135 (= column 2 + column 20 − column 8) [s]
(22) Travel-time residual in ak135, t_{res} (eq. A1 in the main text; = column 18 − column 20) [s]
(23) Difference between ak135 (adjusted for bathymetry and cruising depth) and ak135 predicted travel times, t_{adj} (eq. A5 in the main text; = column 24 − column 20) [s]
(24) Predicted travel time in ak135 (adjusted for bathymetry and cruising) depth, $t^{\star}_{\mathrm{ak135}}$ (= column 20 + column 23) [s]
(25) Predicted arrival time in ak135 (adjusted for bathymetry and cruising) depth [s]
(26) Travel time residual in ak135 (adjusted for bathymetry and cruising) depth, $t^{\star}_{\mathrm{res}}$ (eq. A2 in the main text; = column 18 − column 24) [s]
(27) Difference between LLNL-G3Dv3 and ak135 predicted travel times (= column 28 − column 20) [s]
(28) Predicted travel time in LLNL-G3Dv3, t_{LLNL} (no event-side correction; station-side correction not included because MERMAID is in water; = column 20 + column 27) [s]
(29) Predicted arrival time in LLNL-G3Dv3 (= column 2 + column 28 − column 8) [s]
(30) Travel time residual in LLNL-G3Dv3, $t^{\oplus}_{\mathrm{res}}$ (eq. A3 in the main text; = column 18 − column 28) [s]
(31) Two-standard deviation estimate of the uncertainty, $2\mathrm{SD}_{\mathrm{err}}$, on our AIC pick (i.e., on the observed travel and arrival times, columns 18 and 19, both of which tag the same UTC time, but are given as different elapsed times) using the M1 Method of Simon et al. (2020) [s]
(32) Signal-to-noise ratio using eq. 1 in the main text and the definitions of the "signal" and "noise" in described in the Appendix [rounded, dimensionless]
(33) Maximum (±)amplitude of signal within 1.75 s of arrival [rounded counts]
(34) Time difference between the maximum amplitude and the onset of the signal (think "seconds after the AIC pick") [s]
(35) NEIC PDE event identification number
(36) IRIS event identification number

Computer code to read and parse `simon2021gji_supplement_residuals.txt` into a MATLAB structure is available at `https://github.com/joelsimon/omnia/` in a routine called `read_simon2021gji_supplement_residuals.m`

S4.2 *All events in the MERMAID catalog:* `simon2021gji_supplement_events.txt`

We also include `simon2021gji_supplement_events.txt`, a file similar to `simon2021gji_supplement_residuals.txt`, except that it includes a line for each of the 1363 SAC files processed in the main text. This means that it includes lines for unidentified SAC files (for which we fill event metadata columns with NaNs), as well as identified seismograms in which the first-arriving phase is not a *p* or *P* wave.

The columns of each `simon2021gji_supplement_events.txt` are:

(1)–(13) Identical to columns 1–13 in `simon2021gji_supplement_residuals.txt`, or NaN in columns corresponding to event data for unidentified seismograms
(14)–(15) Identical to columns (35)–(36) `simon2021gji_supplement_residuals.txt`, or NaN for unidentified seismograms

Computer code to read and parse `simon2021gji_supplement_events.txt` into a MATLAB structure is available at `https://github.com/joelsimon/omnia/` in a routine called `read_simon2021gji_supplement_events.m`

S4.3 *Every GPS fix from the SPPIM deployment:* `simon2021gji_supplement_P00??_gps.txt`

Should other researchers be interested in re-locating our stations at the time of recording we also provide individual text files of every GPS fix recorded by MERMAID while at the surface through the first dive cycle ending in 2020. MERMAID records data at depth and then ascends to the surface to transmit those data at a later date with the location at the time of recording being interpolated (Joubert et al. 2016; Simon et al. 2021). We name these files `simon2021gji_supplement_P00??_gps.txt`, where the "??" is understood to mean the unique two-digit serial number of the recording MERMAID; one from the inclusive list $[08, \ldots, 25]$, excluding 14 and 15, which never existed. Note that MERMAID transmits these data in so-called "MER" and "LOG" files. Very commonly, two GPS acquisitions that are similar in time and space (with differences on the order of seconds and meters) will be reported first in a "LOG" file and next in a "MER" file. In these cases it is advisable to use the location data in the "LOG" file and the timing data in the "MER" file.

The columns of each `simon2021gji_supplement_P00??_gps.txt` are:

(01) Time of GPS acquisition
(02) Latitude in decimal degrees (computed from the raw string of column 9)
(03) Longitude in decimal degrees (computed from the raw string of column 10)
(04) Horizontal dilution of precision (not contained in MER files)
(05) Vertical dilution of precision (not contained in MER files)
(06) Clock drift since last synchronization (GPS time − MERMAID time) [s]
(07) Clock frequency of internal MERMAID clock [Hz]
(08) Source filename (LOG or MER file)
(09) Latitude in raw string as quoted in LOG or MER file (some concatenated form of degrees and decimal minutes)
(10) Longitude in raw string as quoted in LOG or MER file (some concatenated form of degrees and decimal minutes)
(NB: undefined values are given as nan; large clock drifts signal an internal reset; clock frequencies between 3000000 and 4000000 signal valid GPS acquisition)

Computer code to read and parse `simon2021gji_supplement_P00??_gps.txt` into a MATLAB structure is available at `https://github.com/joelsimon/omnia/` in a routine called `read_simon2021gji_supplement_gps.m`

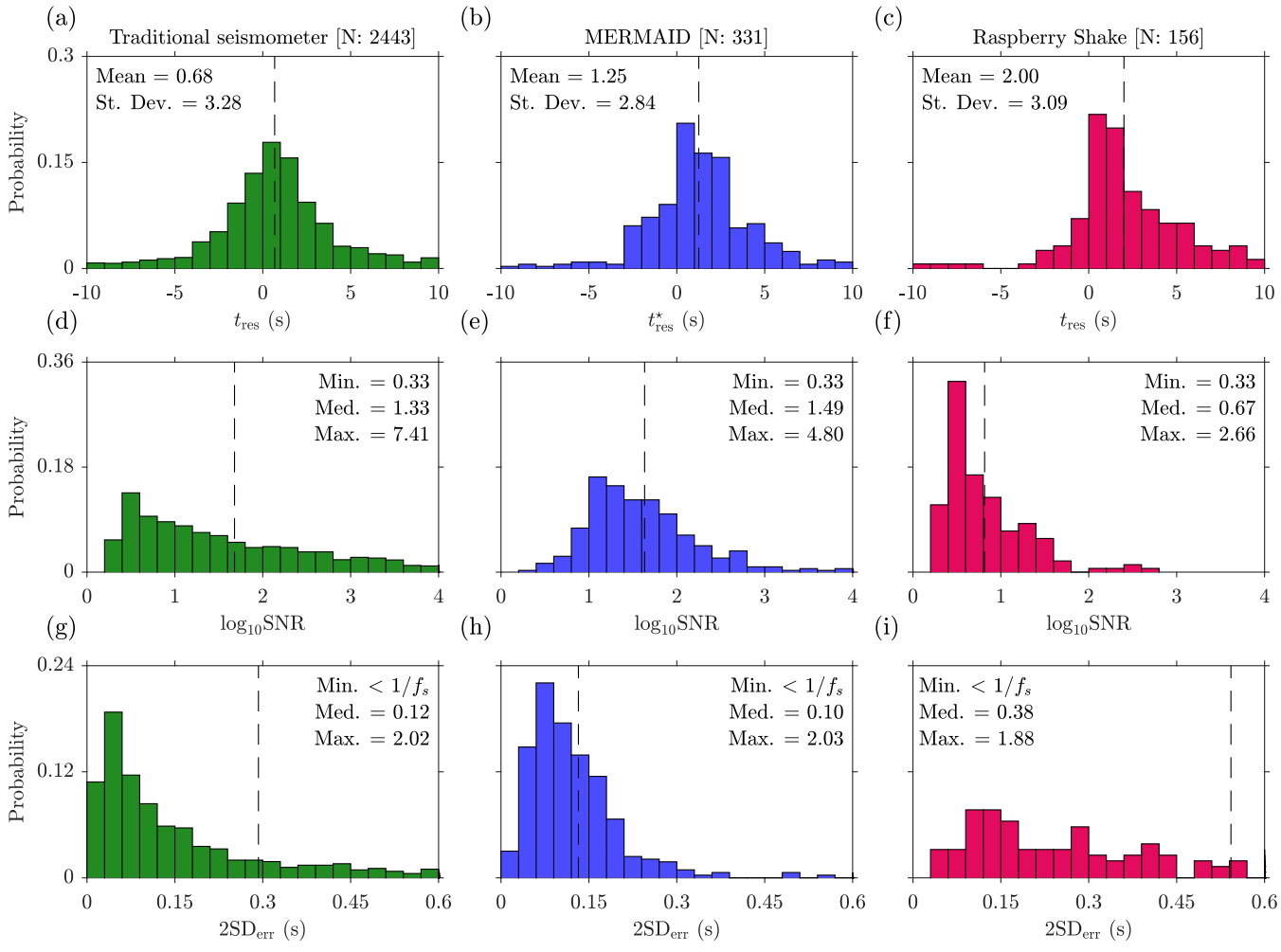

Figure S2. Fig. 10 of the main text, remade considering only the subset of events for which data existed for at least one station within each instrument class.

S5 COMPARING THE HIGHEST-SNR SEISMOGRAMS ACROSS THE INSTRUMENT CLASSES

Fig. S2 redraws Fig. 10 of the main text but considering only the subset of data representing events in the MERMAID catalog that occurred while at least one traditional and one Raspberry Shake instrument were also installed—i.e., the catalog of events common to all instrument classes. This comparison is made here because some of the larger events present in the data included in Figs 9 and 10 of the main text are not in the Raspberry Shake catalog because those stations had not yet been installed.

Figs S3–S5 each plot the 12 highest-SNR signals in this catalog of common events for traditional island stations, MERMAID and Raspberry Shake island stations, respectively. It is readily apparent that Raspberry Shake instruments are generally noisier than either of the other two instrument classes because the variance of the gray noise segment that precedes the colored signal segment is often visible, whereas for the other two instrument classes this is not the case (the noise looks flat at this range of ordinate values). Also, the uncertainties associated with Raspberry Shake seismograms are generally higher than those of the other two instrument classes. Differences in epicentral distance are not considered here.

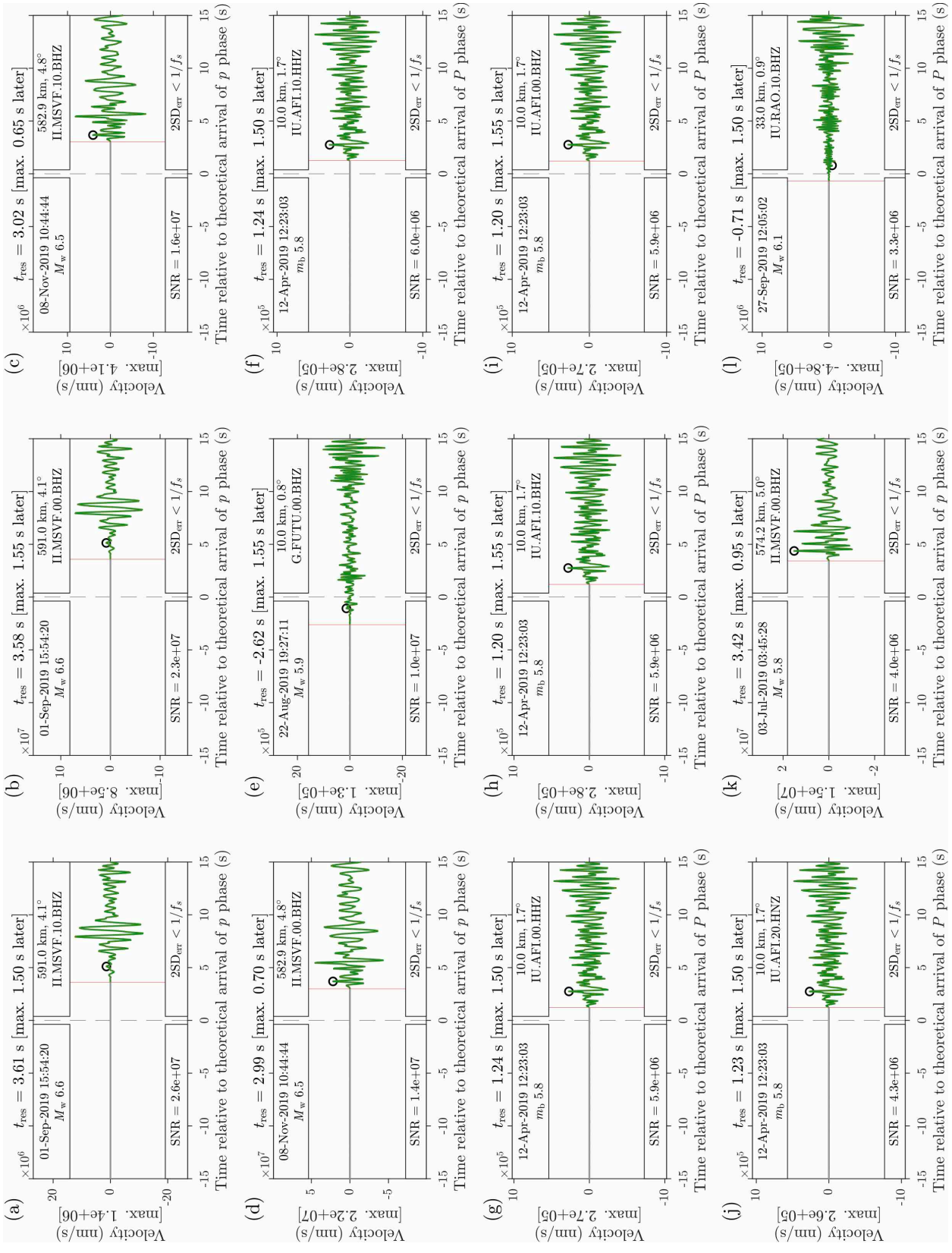

Figure S3. The 12 highest-SNR signals recorded by traditional island stations considering the catalog of events common to all three instrument classes. They are presented in the same format as Fig. 9 of the main text, except that the residuals are in reference to the standard ak135 model (eq. A1 of the main text). The seismograms are plotted in units of velocity (nm/s), and the signals are colored green.

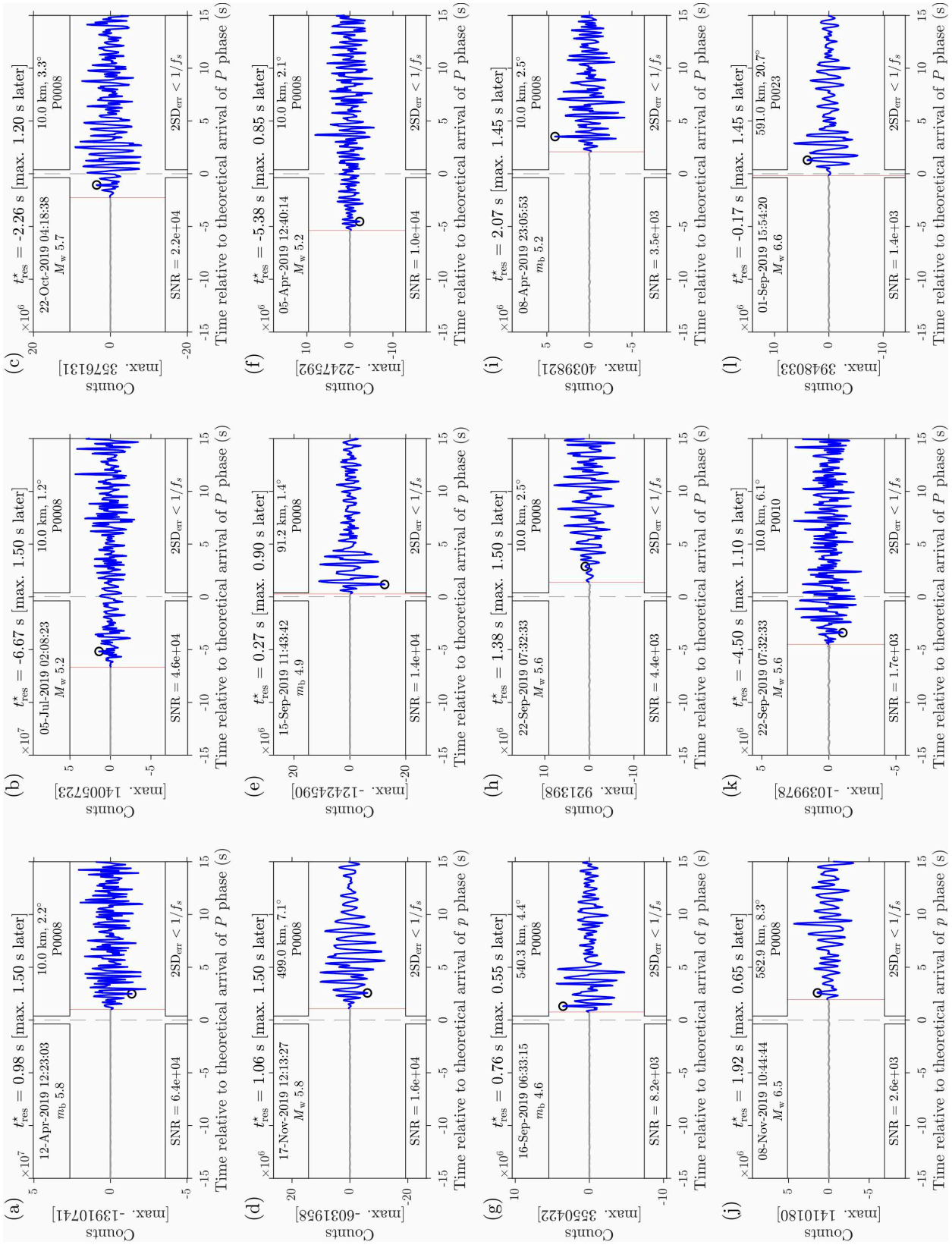

Figure S4. The 12 highest-SNR signals recorded by MERMAID considering the catalog of events common to all three instrument classes, presented in the same format as Fig. 9 of the main text.

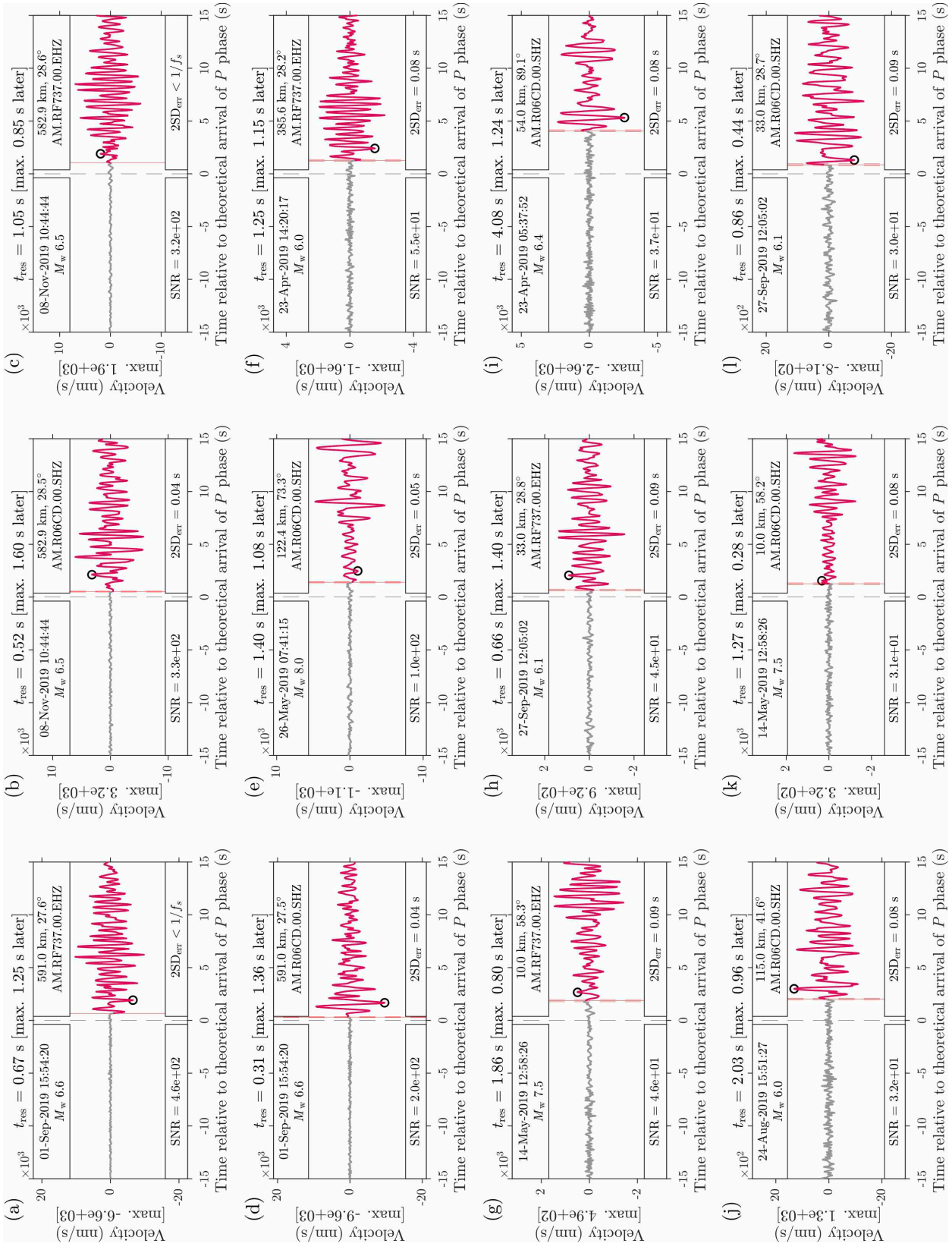

Figure S5. The 12 highest-SNR signals recorded by Raspberry Shake island stations considering the catalog of events common to all three instrument classes. They are presented in the same format as Fig. 9 of the main text, except that the residuals are in reference to the standard ak135 model (eq. A1 of the main text). The seismograms are plotted in units of velocity (nm/s), and the signals are colored raspberry.

S6 WRITING SAC POLE-ZERO FILES FOR RÉSEAU SISMIQUE POLYNÉSIEN STATIONS

S6.1 *The SACPZ file*

Seismic Analysis Code (SAC) pole-zero (SACPZ) files specify the frequency response of a digital seismic instrument. They describe how a seismometer converts ground motion to digital counts. The output in digital counts of a seismometer is a record of true ground motion multiplied by the response of the instrument in the frequency domain. With a SACPZ file, one may recover an accurate record of ground motion via deconvolution (division in the frequency domain) of the seismogram with the frequency response. This process is referred to as "removing the instrument response," and it can be accomplished in the SAC program with the `TRANSFER` command (Helffrich et al. 2013). See also Burky et al. (2021).

SACPZ files contain poles, zeros, and a constant. The first two are the complex roots (poles: denominator; zeros: numerator) of the transfer function of the analog instrument, and the last is a constant that describes the gain of the entire system. By analog we mean the physical seismic instrument—for example, an inertial mass held in place by a varying electric current—that intakes ground motion (e.g., m/s) and outputs voltage (V). Following the analog stage, seismometers pass their data through multiple stages of digitization where the voltage is converted to counts. Ignoring any frequency effects during digitization (which SACPZ files do not include), the poles and zeros are sufficient to describe the phase response of the system—i.e., with no constant, the ungained seismogram after deconvolution will have the proper shape but incorrect amplitude. Phase shifts acquired during the digital stages are usually negligible and can be ignored, as noted (in bold) on p. 152 of the Standards for the Exchange of Earthquake Data (SEED) Reference Manual Version 2.4 (2012, `https://fdsn.org/pdf/SEEDManual_V2.4.pdf`; hereafter referred to as the SEED manual), p. 409 of the Seismic Analysis Code Users Manual Version 101.6a (2014, `https://ds.iris.edu/files/sac-manual/sac_manual.pdf`; hereafter referred to as the SAC manual), and as has been independently verified by the authors by comparing waveforms deconvolved with SACPZ and RESP files (the latter of which take into account all digitization stages). We include this comment to make the point that our method of removing the instrument response using the SACPZ file (and not other file standards like RESP or StationXML, which encode information concerning the full cascade of digital filters) is sufficient to recover an accurate record of ground motion for the Réseau Sismique Polynésien (RSP) instruments used in this study.

A SACPZ file may be specified in terms of displacement, velocity, or acceleration. Very commonly a seismometer will physically measure ground velocity, in which case the poles and zeros will likely be reported for the velocity transfer function of the analog stage, and the gain constant (also called the "sensitivity") will be in units like counts/(m/s). SACPZ files were not available for the six stations used in this study from the RSP. However, we were provided the poles and zeros, and a gain constant at a specific frequency, which is enough to write our own SACPZ files. It is important to note that a gain constant at a single frequency is not the same thing as the constant of a SACPZ file. To be unambiguous we will hereafter refer to the former as the sensitivity. Indeed, the sensitivity describes a gain factor at a single frequency, while the SACPZ constant describes the gain factor at all frequencies. Using the notation of the SEED manual, and the pole-zero representation of the transfer function, the frequency response at any stage of the system is (eq. 4 p. 158),

$$G(f) = S_d R(f), \tag{S1}$$

where $R(f)$ is some function of frequency and the S_d is the sensitivity. For the analog stage,

$$R(f) = A_0 H_p(s), \tag{S2}$$

where A_0 is a normalization factor at frequency f_s in Hz (note that the normalization factor may be derived at a frequency, f_n, different from f_s, but this is goes against the SEED convention [p. 157], and is not done here), and $H_p(s)$ is the transfer function at $s = 2\pi i f$ rad/s. Note that we assumed that the poles and zeros of $H_p(s)$ are in rad/s (SEED type "A"), and not in Hz (SEED type "B"), as is the convention used in the SEED manual, and which has been our experience when retrieving RESP and StationXML files from both IRIS and International Federation of Digital Seismograph Networks (FDSN) Web Services (`https://www.fdsn.org/webservices/`). At all stages $R(f)$ is defined such that its modulus is unity at the specified frequency of the sensitivity, $f = f_s$,

$$\left|R(f_s)\right| = 1, \tag{S3}$$

leading to the relationship at the analog stage,

$$A_0 = 1/H_p(s_s). \tag{S4}$$

Therefore, ignoring frequency effects beyond the analog stage, and defining S_D to be the multiplicative combination of sensitivities at all stages, the complete frequency response of the entire system at any frequency in Hz is

$$G(f) = S_D A_0 H_p(s) \tag{S5}$$

$$= \mathcal{C} H_p(s), \tag{S6}$$

where $\mathcal{C}$ is the constant included in the SACPZ file.

S6.2 *The SACPZ constant*

With the delivery of seismic data from the Réseau Sismique Polynésien (RSP), we were also provided poles, zeros, and a sensitivity corresponding to each station. Equal for all six stations were their two zeros (0;0)(0;0) and two poles $(-4.44;-4.44)(-4.44;+4.44)$. The sensitivity of stations PAE and TVO was given as 0.5236 (nm/s)/LSB at 1 Hz, and for stations PMOR, VAH, TBI, and RKT as 0.212 (nm/s)/LSB at 1 Hz. Here, LSB stands for "least significant bit," and in this case refers to digital counts. Therefore, for all six stations, the poles, zeros, and sensitivity frequency of $f_s = 1$ Hz are identical, but the sensitivities differ.

As provided, these data required slight transformation before computation of the the constants of eq. (S6). First, the sensitivities were given in terms of velocity per counts, whereas the convention used in the SEED manual (p. 12) and the IRIS and FDSN Web Services specifies those data in terms of counts per unit of ground motion. Therefore, the sensitivities were inverted to convert them to counts/(nm/s). Next, they were converted from nm to m by multiplication with 10^9 to conform to the SEED convention that the transfer function be given in SI units.

Finally, we converted the pole and zero data from velocity (describing the transformation from digital counts to m/s) to displacement (counts to m). This was done to conform to the SAC standard that a `TRANSFER` to `NONE` (deconvolution in the SAC program) results in a displacement seismogram. Otherwise, if left as is, `TRANSFER` to `NONE` produces a velocity seismogram—when using SACPZ files the SAC `TRANSFER` function does not automatically convert ground motion units to displacement, if required, as it does with RESP files (SAC Manual, p. 406). To that end, the sensitivities were multiplied by $2\pi f_s$ (recalling that the sensitivities hold at a specific frequency in Hz, but were computed from a transfer function in rad/s), and a zero was added to the set of poles and zeros (resulting from the integration of the complex transfer function). Note that SAC does not use SI units, but rather assumes (and populates the relevant header variables accordingly) that a `TRANSFER` to `NONE` results in a displacement seismogram in units of nm/s. However, we chose to prioritize SEED standards over SAC standards (true for the IRIS and FDSN Web Services in our experience), and thus we were careful to apply the proper multiplication factor of 10^9 in the SAC program after deconvolution such that the resultant waveforms were in nm (or nm/s for a `TRANSFER` to `VEL`, as was done with all data in the main text from nearby island stations), to properly match the units written to the SAC header variable "IDEP."

We therefore report the following displacement SACPZ files in SI units for RSP stations PAE and TVO written using our `sacpzconstant.m` and `cpptsacpzconstant.m` routines,

```
ZEROS 3
+0.000000e+00 +0.000000e+00
+0.000000e+00 +0.000000e+00
+0.000000e+00 +0.000000e+00
POLES 2
-4.440000e+00 -4.440000e+00
-4.440000e+00 +4.440000e+00
CONSTANT  2.699191e+09,
```

and for stations PMOR, VAH, TBI, RKT,

```
ZEROS 3
+0.000000e+00 +0.000000e+00
+0.000000e+00 +0.000000e+00
+0.000000e+00 +0.000000e+00
POLES 2
-4.440000e+00 -4.440000e+00
-4.440000e+00 +4.440000e+00
CONSTANT  6.666493e+09.
```

The functions relevant to this section that compute the SACPZ constants, $\mathcal{C}$, and normalization factors, A_0 are accessible at `github.com/joelsimon/omnia/`. Included there as well is `transfunc.m`, a function which may be of use to those interested in the conversion between SACPZ, RESP, and StationXML files, as well as their transformation between displacement, velocity, and acceleration responses.

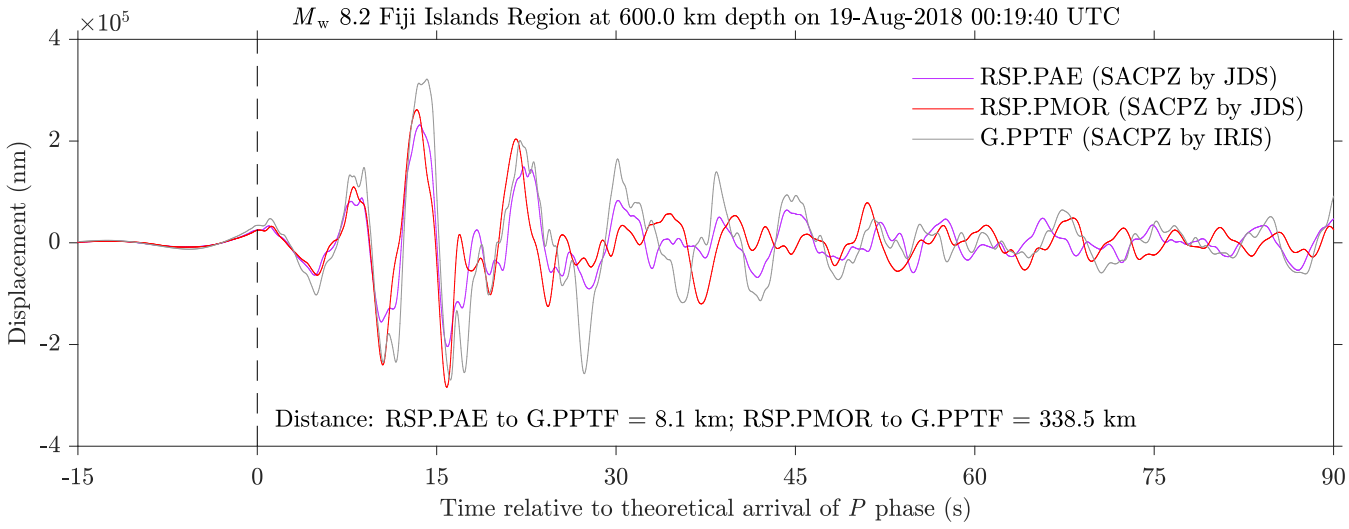

Figure S6. Unfiltered seismograms from RSP.PAE (purple), RSP.PMOR (red), G.PPTF (gray) of a nearby great earthquake. The SACPZ files corresponding to the two RSP stations were written by the authors, and that corresponding to G.PPTF was provided by IRIS. The similarity of the waveforms, both in phase and amplitude, proves that our SACPZ files are correct.

S6.3 *Verification*

Fig. S6 proves that the displacement SACPZ files we wrote for RSP stations are correct. It compares the unfiltered (apart from those corner frequencies specified during deconvolution, see the Appendix of the main text) seismograms, plotted in displacement (nm), corresponding to a great earthquake in the Fiji Islands region that was recorded by three nearby stations. The traces are each aligned on the theoretical arrival time of the first-arriving *P* wave computed in the ak135 velocity model. Two of the seismograms were recorded by stations in the RSP (PAE and PMOR, in purple and red seismograms), each serving as the archetypal station for their respective group's SACPZ file written by the authors, and the other by station G.PPTF (in gray), for which the displacement SACPZ file was available from IRIS. The distance between each RSP station and G.PPTF is listed inside the axis (8.1 km for PAE and 338.5 km for PMOR), and they are near enough to one another that we would expect the ground motion at the three stations to be very similar, given the magnitude of earthquake. We see that the waveforms agree very well, both in amplitude and phase, both before and after the first-arrival, but especially for the first fifteen seconds after the first arrival. Therefore, we conclude that the two SACPZ files we wrote in Section S6.2 corresponding to six stations in the RSP are correct.

We include as further verification an example in the header of `sacpzconstant.m` that rederives the values printed on an IRIS help page that explains how to convert a velocity RESP file to a displacement SACPZ file (`ds.iris.edu/ds/support/faq/24/what-are-the-fields-in-a-resp-file/`). That example shows that our SACPZ constant agrees with the one provided by IRIS to within 0.003%, well within acceptable error.

REFERENCES

Anthony, R. E., Ringler, A. T., Wilson, D. C. & Wolin, E., 2019. Do low-cost seismographs perform well enough for your network? An overview of laboratory tests and field observations of the OSOP Raspberry Shake 4D, *Seismol. Res. Lett.*, **90**(1), 219–228, doi: 10.1785/0220180251.

Balfour, N. J., Salmon, M. & Sambridge, M., 2014. The Australian Seismometers in Schools Network: Education, outreach, research, and monitoring, *Seismol. Res. Lett.*, **85**(5), 1063–1068, doi: 10.1785/0220140025.

Bent, A. L., Cassidy, J., Prépetit, C., Lamontagne, M. & Ulysse, S., 2018. Real-time seismic monitoring in Haiti and some applications, *Seismol. Res. Lett.*, **89**(2A), 407–415, doi: 10.1785/0220170176.

Bondár, I. & Storchak, D., 2011. Improved location procedures at the International Seismological Centre, *Geophys. J. Int.*, **186**(3), 1220–1244 doi: 10.1111/j.1365-246X.2011.05107.x.

Burky, A., Irving, J. C. E. & Simons, F. J., 2021. Instrument response removal and the 2020 M3.1 Marlboro, New Jersey, earthquake, *Seismol. Res. Lett.*. doi: 10.1785/0220210118.

Calais, E., Boisson, D., Symithe, S., Momplaisir, R., Prépetit, C., Ulysse, S., Etienne, G. P., Courboulex, F., Deschamps, A., Monfret, T., Ampuero, J.-P., de Lépinay, B. M., Clouard, V., Bossu, R., Fallou, L. & Bertrand, E., 2019. Monitoring Haiti's quakes with Raspberry Shake, *Eos Trans. AGU*, **100**(11), 17–21, doi: 10.1029/2019EO123345.

Cochran, E. S., Lawrence, J. F., Christensen, C. & Jakka, R. S., 2009. The Quake-Catcher Network: Citizen science expanding seismic horizons, *Seismol. Res. Lett.*, **80**(1), 26–30, doi: 10.1785/gssrl.80.1.26.

Helffrich, G., Wookey, J. & Bastow, I., 2013. *The Seismic Analysis Code. A Primer and User's Guide*, Cambridge Univ. Press, Cambridge, UK.

International Seismological Centre, 2015, *On-line Bulletin*, http://www.isc.ac.uk. doi: 10.31905/D808B830.

Jeddi, Z., Voss, P. H., Sørensen, M. B., Danielsen, F., Dahl-Jensen, T., Larsen, T. B., Nielsen, G., Hansen, A., Jakobsen, P. & Frederiksen, P. O., 2020. Citizen seismology in the Arctic, *Front. Earth Sci.*, **8**, 139, doi: 10.3389/feart.2020.00139.

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, doi: 10.1785/0220150111.

Kennett, B. L. N., Engdahl, E. R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, **122**(1), 108–124, doi: 10.1111/j.1365-246X.1995.tb03540.x.

Kong, Q., Allen, R. M. & Schreier, L., 2016. MyShake: Initial observations from a global smartphone seimic network, *Geophys. Res. Lett.*, **43**(18), 9588–9594, doi: 10.1002/2016GL070955.

Simmons, N. A., Myers, S. C., Johannesson, G. & Matzel, E., 2012. LLNL-G3Dv3: Global P wave tomography model for improved regional and teleseismic travel time prediction, *J. Geophys. Res.*, **117**, B10302, doi: 10.1029/2012JB009525.

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, *B. Seismol. Soc. Am.*, **110**(3), 970–997, doi: 10.1785/0120190173.

Simon, J. D., Bonnieux, S., Simons, F. J. & The EarthScope-Oceans Consortium, 2021, *automaid*, https://github.com/earthscopeoceans/automaid. doi: 10.5281/zenodo.5057096.

Subedi, S., Hetényi, G., Denton, P. & Sauron, A., 2020. Seismology at school in Nepal: A program for educational and citizen seismology through a low-cost seismic network, *Front. Earth Sci.*, **8**, 73, doi: 10.3389/feart.2020.00073.

Weatherall, P., Marks, K. M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J. E., Rovere, M., Chayes, D., Ferrini, V. & Wigley, R., 2015. A new digital bathymetric model of the world's oceans, *Earth Space Sci.*, **2**(8), 331–345, doi: 10.1002/2015EA000107.